



Ca' Foscari
University
of Venice

Master's Degree
in Language Sciences

Final Thesis

Interaction, Design, and Assessment: An Exploratory Study on ChatGPT in Language Education

Supervisor

Ch. Prof. Marcella Menegale

Assistant supervisor

Ch. Prof. Fabiana Fazzi

Graduand

Corinne Baradel

Matriculation Number 856384

Academic Year

2022 / 2023

*To G. and all young people who
found this world too heavy to bear.
Your light continues to shine in the
hearts you have left behind.*

*Forse occorre far loro capire che
noi ci siamo davvero, coi fatti, e
non solo con le parole. Forse è ora,
perché forse è mai più.*

- Enrico Galiano

TABLE OF CONTENTS

ABSTRACT (English)	8
ABSTRACT (Italian)	9
ACKNOWLEDGEMENTS	10
INTRODUCTION	11
CHAPTER I. ARTIFICIAL INTELLIGENCE.....	14
1.1. Definition and development of Artificial Intelligence	14
1.2. Large Language Models	21
1.3. Artificial Intelligence in Educational Contexts	22
1.4. AI in foreign or second language education.....	25
1.5. Conclusion and gaps in the literature on AI in education.....	27
CHAPTER II. CHATGPT.....	29
2.1. Definition and Development of ChatGPT	29
2.2. ChatGPT in teaching and learning.....	34
2.3. ChatGPT on the assessment practices	36
2.4. Limits and potential of ChatGPT.....	41
2.5. Conclusions and future research directions	43
CHAPTER III. METHODOLOGY	45
3.1. The research context.....	45
3.2. Research questions, hypotheses and aims	46
3.3. The participants	47
3.4. Data collection instruments	50
3.4.1. Student teacher online questionnaire.....	50
3.4.2. Student teacher online interviews.....	51
3.4.3. Student teachers' interactions with ChatGPT.....	52

3.4.4. Student teachers' final TBLL activities.....	54
3.4.5. Professor online interview	55
3.4.6. ChatGPT's feedback.....	56
3.4.7. Author's feedback	56
3.4.8. Author's prompt engineering	56
3.5. The research methodology.....	57
3.6. Activity description	57
3.7. Data collection procedure	58
3.7. Ethics	59
3.8. Data analysis procedure.....	59
CHAPTER IV. DATA ANALYSIS.....	61
4.1. Premises.....	61
4.2. Data analysis of RQ1	64
4.3. Data analysis of RQ2.....	73
4.4. Data analysis of RQ3	81
4.5. Data analysis of RQ4.....	101
4.6. Conclusion of the data analysis	106
CHAPTER V. DISCUSSION	108
5.1. RQ1: Usefulness of ChatGPT (time, quality of activities and critical thinking).....	108
5.2. RQ2: Strategies used by student teachers when interacting with ChatGPT.....	111
5.3. RQ3: Ways to assess the student teachers' activities with ChatGPT	112
5.4. RQ4: Usefulness of ChatGPT in assessing the activities in terms of planning time..	114
5.5. Conclusion of the discussion of the results	115
5.6. Limitations to the research study.....	116
CONCLUSION	118
Proposals for future research	120

REFERENCES.....	121
APPENDIX A. INTERACTION WITH CHATGPT ON ASSESSMENT PRACTICES.....	135

ABSTRACT (English)

The launch of the revolutionary Artificial Intelligence (AI) tool named ChatGPT has raised several concerns in various fields, including language education. Despite critical perspectives and even prohibitions within some countries, regions and communities, ChatGPT is expected to remain a prominent presence. Many researchers have attempted to utilise ChatGPT to support teaching and learning and discover its potentialities as well as limits. However, due to its recent release, little is known about how ChatGPT can be utilised, to what extent ChatGPT is useful for educators and students and affects students' critical thinking skills. The present exploratory study examines the usefulness of ChatGPT in creating and assessing Task-Based Language Learning (TBLL) activities designed by the student teachers, the extent to which ChatGPT affects student teachers' critical thinking skills and the ways student teachers interact with ChatGPT. To investigate these research areas, a mixed-method approach was applied. After having analysed the data gathered from an online questionnaire, online interviews, interactions with ChatGPT, TBLL activities, a self-study, ChatGPT-generated feedback and a self-made feedback text, various considerable insights emerged from the results. First, evidence confirmed that student teachers could design the same or a better quality TBLL activity and save time with the help of ChatGPT. Second, findings revealed that about half of the participants copied the outputs given by the AI tool to perform the TBLL activity. Third, findings disclosed that ChatGPT was not able to create a detailed rubric for assessing student teachers' TBLL activities, but it could generate good-quality feedback by giving it a rubric in the prompt. Fourth, the results showed that ChatGPT allowed educators to save time in elaborating feedback. Based on the findings obtained, some possible solutions as well as proposals for future research in the realm of ChatGPT applied to language education have been formulated.

Keywords: artificial intelligence, ChatGPT, language education, task-based language learning, assessment.

ABSTRACT (Italian)

Il lancio del rivoluzionario strumento di Intelligenza Artificiale (IA) denominato ChatGPT ha destato molte preoccupazioni in diversi ambiti tra cui quello dell'apprendimento linguistico. Nonostante le critiche e i vari divieti all'interno di alcuni Paesi e comunità, ChatGPT sembra conservare un ruolo di rilevante importanza. Molti ricercatori hanno provato ad utilizzare ChatGPT al fine di scoprirne i limiti e le potenzialità e supportare l'insegnamento e l'apprendimento. Data la sua recente introduzione, poche sono le informazioni in merito all'utilizzo di ChatGPT, alla sua utilità per educatori e studenti e alla sua influenza sulla capacità di elaborare il pensiero critico. Il presente studio esplorativo si propone di indagare l'utilità di ChatGPT nel creare delle attività di *Task-Based Language Learning* (TBLL), i vari modi con cui gli studenti aspiranti insegnanti interagiscono con ChatGPT e gli eventuali metodi di valutazione con l'aiuto di ChatGPT. Al fine di indagare queste aree di ricerca è stato adottato un approccio misto. Mediante la raccolta e l'analisi dei dati con questionari e interviste online, conversazioni con ChatGPT, attività di TBLL, uno studio autonomo, feedback generati da ChatGPT e un feedback elaborato dall'autore della tesi, è stato possibile ricavare dei risultati e fare delle considerazioni. Per prima cosa è emerso che gli studenti aspiranti insegnanti sono stati capaci di realizzare delle attività di TBLL migliori o della stessa qualità con l'uso di ChatGPT. Inoltre, la metà dei partecipanti ha copiato gli output di ChatGPT nella creazione delle attività. In aggiunta, si è evidenziato che ChatGPT non è stato in grado di realizzare una rubrica valutativa dettagliata, ma è riuscito a fornire un feedback di buona qualità usando la rubrica come *prompt*. Infine, ChatGPT ha permesso agli educatori di risparmiare del tempo nell'elaborazione dei feedback. In base ai risultati ottenuti, sono state formulate alcune soluzioni e proposte per eventuali ricerche future relative all'uso di ChatGPT nell'ambito linguistico-educativo.

Parole chiave: intelligenza artificiale, ChatGPT, insegnamento delle lingue, task-based language learning, valutazione.

ACKNOWLEDGEMENTS

The present Master's thesis would not have been possible without the support and help that I received from many important people, to whom I would like to communicate my most heartfelt gratitude.

First, I would like to thank my supervisor and assistant supervisor for having given me the opportunity to carry out the experiment in their class, for their helpful insights and guidance, and for their deep interest in my research field. My sincere thanks to all the professors and colleagues (especially to Ilaria, Beatrice and Mariangela) of the Master who gave me the chance to grow in the field of language education. I would also like to give a special thanks to the participants of the surveys and interviews who helped me during the process of data collection.

Furthermore, I would like to express my special gratitude to Mike: in these years, he always stayed closed to me by giving me some advice, supporting and helping me. He always instilled strengths in me to fight for reaching my dreams and not giving up, even in the most difficult situations. He is a glimmer of light in my life.

Moreover, I heartily thank my family and friends for their constant support and for transmit energy and positivity to me. I would also like to thank all the support received by the people who are not physically present here, but they are near to me with their souls: Ennio, Gian Luigi, my grandparents and my aunt.

Additionally, I would like to deeply thank teacher Silvia, who was able to teach me the secrets of children, for believing in me from the first moment, for her exceptional positivity, patience and some significant advice. She is very important to me.

Finally, a heartfelt thanks gives to Professor Pianca, the first English teacher who transmitted me the passion of teaching foreign languages, Professor Anna Corredera and Enrico Galiano for being my role models to follow.

INTRODUCTION

Over the last few years, the introduction of Artificial Intelligence (henceforth AI) technologies into our everyday life has dramatically changed various significant fields in today's society, including education. The present AI tools have great potentialities to enrich learning experiences, make education globally accessible, and personalise learning paths according to students' needs as well as learning styles. Amidst the numerous AI instruments, ChatGPT has stood out since November 2022 and caught notable attention from research, academia as well as industries thanks to its capabilities. In fact, this Large Language Model (hereafter LLM) chatbot can be used as a virtual assistant due to its ability to generate quick human-like conversations, respond to questions, explain complicated concepts, and many more (Kasneji et al., 2023).

Despite its vast potential in the realm of education, AI tools including ChatGPT can present possible security risks as well as ethical challenges. In fact, these types of technologies often collect sensitive data from students as well as teachers in order to personalise learning experiences. Hence, guaranteeing to handle the data gathered in a responsible way is fundamental. Moreover, these tools can make errors of various nature due to the presence of bias in the data that they were trained with. Further concerns are connected to the loss of young generations' critical thinking skills as well as creativity, plagiarism and cheating. In order to tackle these issues, the European Parliament (2023) made some amendments to the document named *Artificial Intelligence Act* to regulate AI.

Given ChatGPT recent release and its impact on education, the present thesis aims at investigating the utility of ChatGPT in creating as well as assessing Task-Based Language Learning (henceforth TBLL) activities addressed to language students, the ways in which student teachers interact with ChatGPT, and the impact of ChatGPT on student teachers' critical thinking skills. In light of the research gaps found in the literature, four research questions (RQs) have been elaborated:

RQ1: *To what extent is the use of ChatGPT useful for student teachers in order to design classroom activities for language learners, in terms of planning time, quality of the final products, and development of their critical thinking?*

RQ2: *What are the strategies used by student teachers when interacting with ChatGPT to create a task-based language learning activity?*

RQ3: *How can university professors assess the activities created by their students through the use of ChatGPT?*

RQ4: *To what extent is the use of ChatGPT useful for student teacher educators in order to assess the activities created by their students in terms of planning time?*

In order to try to address the aforementioned research questions, an exploratory study was conducted.

The present Master's thesis consists of five chapters. Chapters I and II encompass the literature review, which covers AI and ChatGPT. Chapters III, IV and V present a detailed description of the research methodology applied, an analysis of the data gathered, a discussion of the results and limitations to the study. The last section named *Conclusion* provides a summary of the study as well as further research proposals.

Chapter I presents an overview of AI and LLMs with a special reference to Artificial Intelligence in Education (henceforth AIED). Some related examples have been provided from the literature aiming at illustrating both the benefits (Chen et al., 2020; Cardona et al., 2023; Holmes et al., 2019; Luckin et al., 2016; Pedró et al., 2019; UNESCO, 2019) and challenges (Cardona et al., 2023; Holmes et al., 2019; Kasneci et al., 2023) when using AI in the realms of teaching and learning. In addition, a specific section concerning AI in foreign or second language education is described (Kannan & Munday, 2018; Pokrivcakova, 2019). Finally, the literature review includes some missing investigations into AI applied to language education.

Chapter II describes ChatGPT (OpenAI, 2022), its potentialities as well as limits (Jalil et al., 2023; Rudolph et al., 2023; Su et al., 2023; Zhou et al., 2023). The review also focuses on the application of ChatGPT to teaching (Kasneci et al., 2023), learning (Ali et al., 2023; Bitzenbauer, 2023; Muñoz et al., 2023) and assessment (Mizumoto & Eguchi, 2023; Rudolph et al., 2023). In the end, the chapter identifies limited research on examining ChatGPT potentialities in language education, how students interact with and teachers use ChatGPT (Rudolph et al., 2023).

Chapter III contains a description of the research methodology adopted and the participants involved in the study. Furthermore, the chapter explains the TBLL activity carried out by the student teachers, instruments used to gather the data, data collection and analysis procedures.

Chapter IV analyses the data collected from an online questionnaire, three online interviews, conversations with ChatGPT, TBLL activities, a self-study, ChatGPT-generated feedback and a self-made feedback text.

Chapter V discusses the results obtained in the context of the prior research literature and based on the research questions set at the beginning of the study. Moreover, the chapter specifies several limitations to the research. In conclusion, the final section summarises the most important aspects of the present study, pointing out some proposals for future research.

The present exploratory study attempts to contribute to the cutting-edge research field of AI in language education. In fact, the investigation may provide some considerable insights into ChatGPT to the AIED research community.

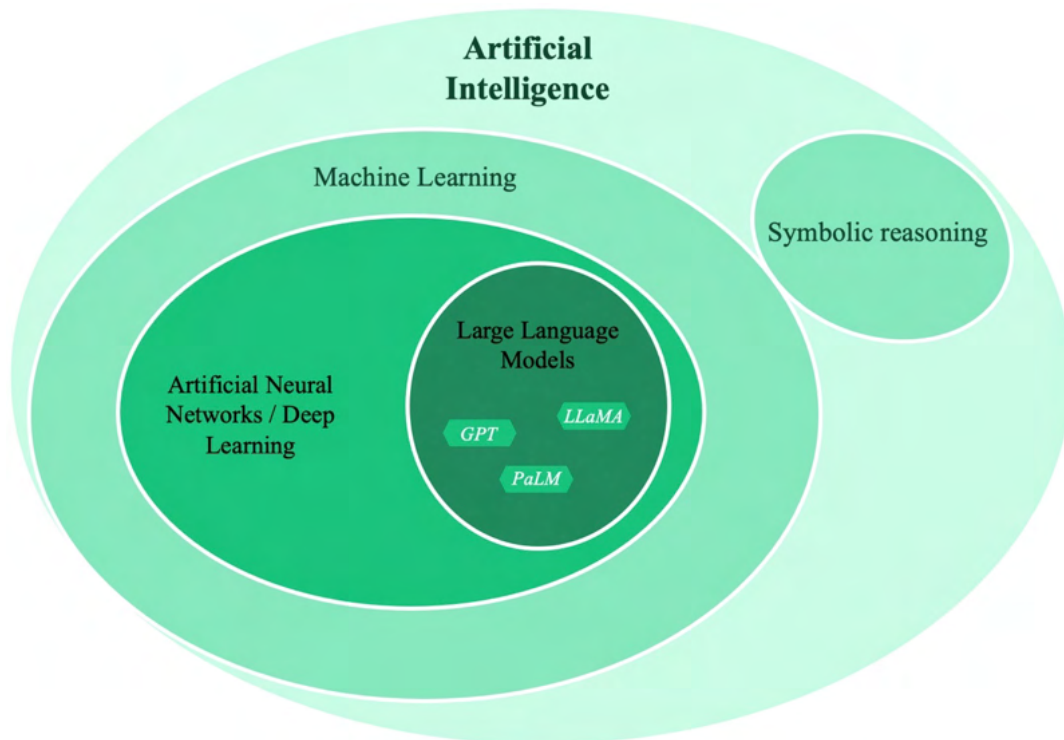
Concerning the reasons behind this research thesis, the author decided to tackle the topic of ChatGPT in language education as she was very interested in AI applied to language education. Specifically, the author wanted to better understand how AI, which is gradually becoming part of everyday life, is changing the field of education. Moreover, the author believes that in the next few years, AI in education will be a really important research area, capable of revolutionising language learning and teaching for all individuals around the world. With respect to this, several investigations reported the benefits of AI and ChatGPT employed in the field of language learning. First, the usage of AI applications can enhance students' motivation as well as engagement (Huang et al., 2023). Second, the utilisation of ChatGPT provides students with assistance in better comprehending difficult concepts and in learning a foreign language (Kasneci et al., 2023). Third, AI and ChatGPT can also aid teachers in creating lesson plans as well as activities (Kasneci et al., 2023). Consequently, the author considered her Master's thesis as an opportunity to acquire more scientific knowledge about AI and ChatGPT in the field of language education.

CHAPTER I. ARTIFICIAL INTELLIGENCE

The current chapter will provide a theoretical overview of Artificial Intelligence (AI), its history and advancement in today's society. Moreover, a deep explanation of the machine learning approach as well as the importance of artificial neural networks will be supplied. Subsequently, large language models, which represent particular artificial neural networks, will be described since they act as a basis for ChatGPT (Figure 1). Furthermore, the literature on the application of AI in education will be examined with particular reference to foreign and second language education. In conclusion, the chapter will include some research gaps and further directions in intelligent computer-assisted language learning.

Figure 1

A Venn diagram embracing the main topics tackled in this chapter



1.1. Definition and development of Artificial Intelligence

For many years, scientists have attempted to comprehend how human beings think and are able to predict, perceive, learn, understand and manipulate the whole world. The area of AI tries to go beyond people's functions. In fact, AI attempts not only to comprehend but also to create intelligent machines based on human behaviours (Russell & Norvig, 2010). AI started to

develop after the Second World War when Alan Turing elaborated the Turing Test (previously called the *imitation game*) aiming at exploring if machines were able to think and finding a definition of what can be intelligent (Kannan & Munday, 2018). In 1950 the computer scientist Turing published one of the first papers about AI titled *Computing Machinery and Intelligence* (Turing, 1950). In the above-mentioned test, humans had to distinguish whether a conversation was produced by another person or a machine. In case individuals were not able to draw a distinction of the conversation, it could be concluded that the machine was an intelligent system. The test was elaborated in the following fashion (Coppin, 2004): a questioner was authorised to interact with a human and a computer located in two separate rooms. The questioner could interrogate both the individual and computer without interacting with them directly. Likely, the queries were written on a computer and the answers were shown on the screen. The person could answer by saying that s/he was the person, and so could the machine. The correct mode to understand whether there was a human behind the screen was to write complicated responses that a machine was not able to give. Obviously, the creators of the intelligent machine provided their programmes with capabilities of predicting such intricacies. Subsequently, the Turing test was applied to many computer programmes elaborated for conversations, for instance, the chatterbot ELIZA designed by Weizenbaum in 1965 (Coppin, 2004).

Only in 1956, the term AI was officially coined by John McCarthy during the Dartmouth research workshop (McCarthy et al., 2006). The professor defined AI as the science of creating programmes capable of performing tasks that require processes of learning, reasoning, comprehension, as well as perception (McCarthy, 2007). Still today, however, it is extremely difficult to find a perfect statement which embraces the whole complexity of AI. In fact, there are several definitions of AI in the literature which have changed over time as this science has been influenced by mathematics, biology, linguistics, neuroscience, philosophy, and so on and so forth (Popenici & Kerr, 2017). For instance, Coppin (2004) described AI as the study of models which use several methods akin to human behaviours. Similarly, Whitby (2003) explained that AI represents the analysis of human, animal and machine behaviours and the effort to engineer these conducts into artefacts. A further definition is given by Chassignol together with other researchers (2018) who consider AI as a field of computer science capable of performing tasks which need human intelligence, such as speech recognition or translations

into different languages. Based on the aforementioned definitions, it can be observed that AI embodies the approximate imitations of human abilities to be performed by machines.

Currently, AI represents one of the most innovative fields of engineering and science and is applied to further research areas such as information and communication technology, medical science, finance (Zhang & Lu, 2021) and education (Chen et al., 2020). Additionally, AI is part and parcel of our daily life: from applications and features present in mobile phones or computers to domestic devices or public transport. Some typical instances of AI are virtual assistants, such as *Siri* created by Apple or *Alexa* developed by Amazon (Perez Garcia et al., 2018), or *Face ID* elaborated by Apple, which represents a machine learning system able to recognize faces (Baqeel & Saeed, 2019). Sometimes, AI systems are closer than we can imagine. In fact, AI tools also work behind notorious social media such as Facebook, Instagram, TikTok and Twitter (Augustine & Xavier, 2021). Numerous famous technological companies like Google, Apple, Mercedes, Tesla, Facebook, OpenAI, and Facebook are investing in research and innovative software to compete in the sector of AI (Popenici & Kerr, 2017). The aforementioned tools are changing the ways people are searching for information, how individuals communicate and even behave (Chassignol et al., 2018). Therefore, the whole society should pay attention to and be educated on AI instruments (UNESCO, 2022).

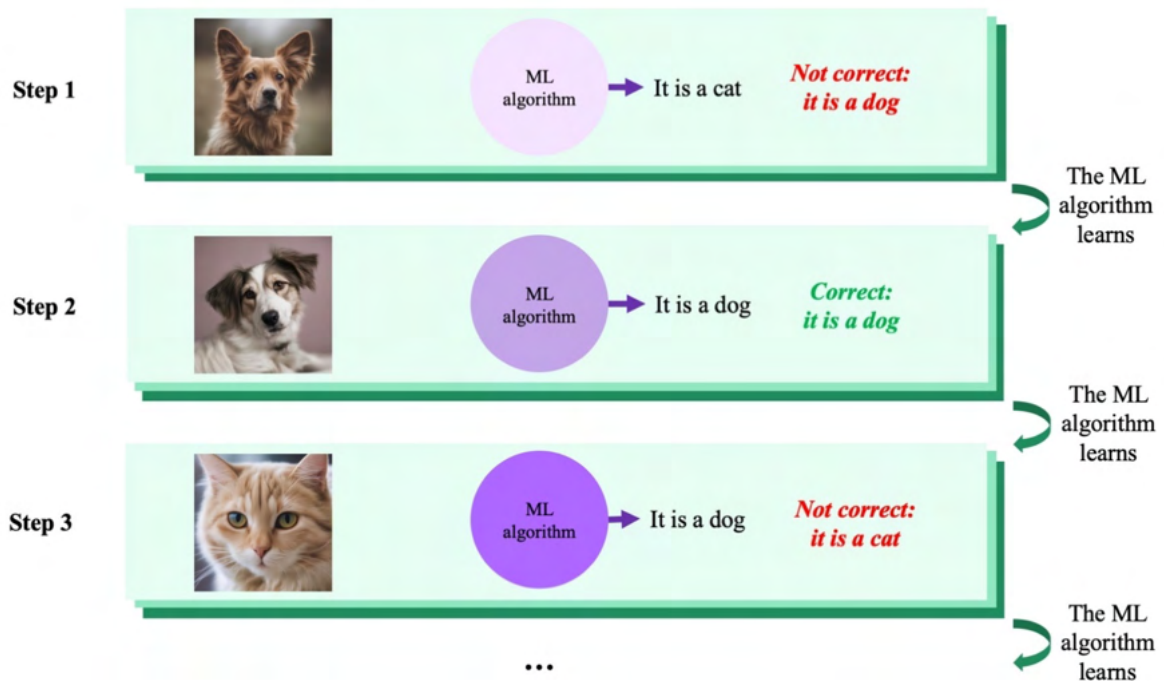
There are two main approaches used by humans to develop AI (Herrmann, 2022). The first one is the *symbolic paradigm* referring to traditional algorithms that rely on predefined rules or heuristics designed manually by human experts. An example is *Deep Blue* that was the first algorithm to beat a human champion of chess in 1997 thanks to its superhuman capabilities (Campbell et al., 2002). This was evidence that the machine could perform some tasks better than human beings. In order to reach this aim, the algorithm efficiently explored the several possibilities of game for each movement and then, *Deep Blue* selected the best option amidst all the possibilities according to numerous heuristic evaluative functions. These functions were manually and carefully developed by human specialists (programmers and chess experts) (Campbell et al., 2002). The current approach, however, displays some limitations (Garnelo & Shanahan, 2019). It is complicated to implement a good set of rules and teach the AI the keywords included in the rules. For instance, if scientists want to build an AI capable of distinguishing between cat and dog images, the researchers will need to use their expert knowledge to define sets of classifications, such as “If the animal has a long snout, it is a dog”

or “If the animal is higher than 25 cm, it is a dog”. Furthermore, scientists have to teach the machine what a snout or height is.

The second approach is named *Machine Learning* (ML) (Coppin, 2004). The algorithms of ML are more generic as they are able to solve issues by learning from the data. Algorithms of ML do not require an expert or programmer to define explicit rules or heuristic functions. Indeed, algorithms will learn the set of rules and heuristic evaluations on their own from the given data. For instance, training an algorithm of ML on two different groups of images (such as cats and dogs) signifies that the algorithm will learn on its own the necessary rules to differentiate the presence of a cat or dog in a picture. A modern case of ML is *AlphaGo* corresponding to the equivalent of *Deep Blue*. *AlphaGo* was the first algorithm, developed by Google DeepMind, able to defeat a world champion of Go (a challenging board game) in 2016 (Holmes et al., 2019). *AlphaGo* works similarly to *Deep Blue* by exploring a range of possible future movements, but the way *AlphaGo* learnt how to predict the correct option in order to win is different. In fact, *AlphaGo* learnt to evaluate each movement by studying a huge amount of data concerning not only movements and famous matches of Go, but also matches played against a copy of itself (Silver et al., 2016; Haenlein & Kaplan, 2019). Basically, *AlphaGo* was not given a set of rules by humans, but the algorithm learnt by experience. However, it is obvious that the main challenge concerning ML is to find and collect a rich set of data to provide to the model (Breck et al., 2019). Nevertheless, this approach has produced better results and has shown to be easier to apply. For this reason, starting from the 1980s ML has become the main approach to AI (Goksel & Bozkurt, 2019).

Figure 2

A representation of how a ML algorithm learns



Note. The images of the dogs and cat were created with *Clipdrop* (<https://clipdrop.co/>)

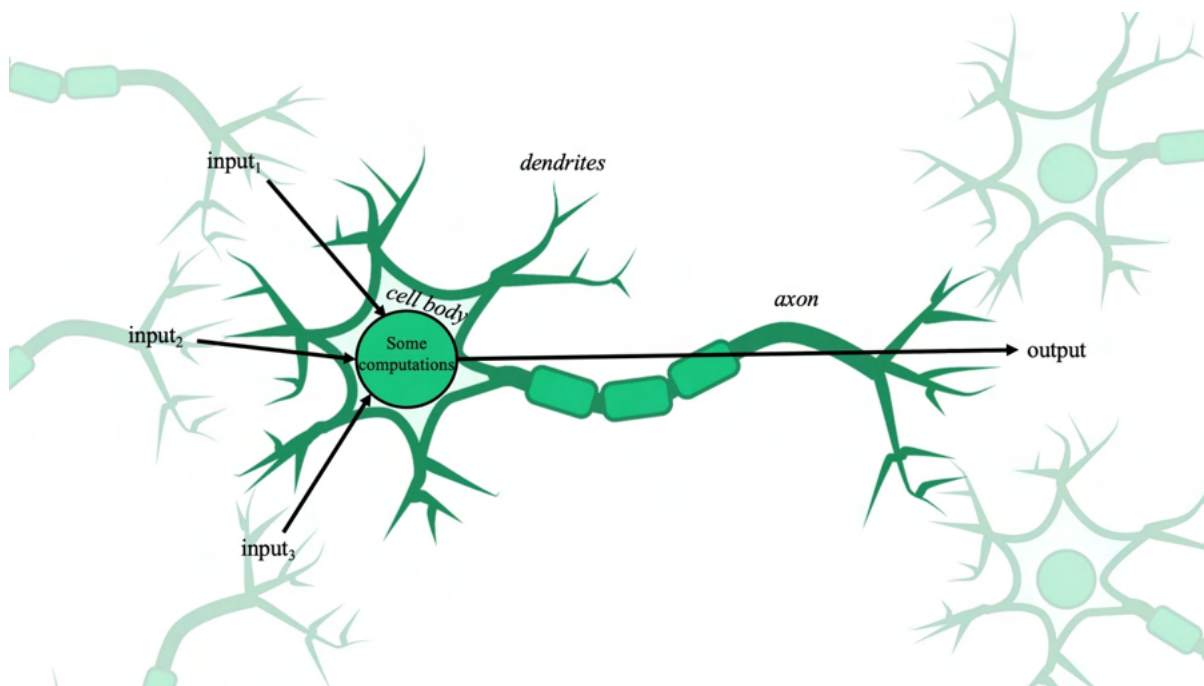
There are several techniques through which ML algorithms learn (Coppin, 2004; Liu et al., 2023; Zhang & Lu, 2021) and one of them is shown in Figure 2 (the details of the techniques used to train ML algorithms are out of the scope of this thesis). Sample by sample, the algorithm tries to say what there is in the image (in this case a cat or a dog) and it is told if its predictions are correct or wrong. In this way, the model learns. If it is trained with enough samples, in the end, the model should be able to distinguish a cat from a dog.

There are various types of algorithms in ML, but now the most relevant algorithms are the Artificial Neural Networks (ANNs) (Holmes et al., 2019). ANNs are statistical models which pretend to imitate a real network of biological neurons (Holmes et al., 2019; Bini, 2018). In order to comprehend how an ANN works, it is relevant to provide the definition of a biological neuron (Stangor & Walinga, 2014; Coppin, 2004). Neurons are cells which form the nervous system in human brains. Their aim is to receive as well as convey information. Each neuron consists of three essential parts: the dendrites, soma (the cell body), and axon. The dendrites receive the input from other neurons and send them to the cell body. Depending on the strength of inputs, which signals have been or have not been received, the neuron will fire or not (Figure 3). In case the neuron fires, it will send an electric impulse through the axon

which connects to the next neurons' dendrites. Basically, the output of a neuron is an input of another neuron (Stangor & Walinga, 2014; Coppin, 2004). Starting from this concept, an artificial neuron works similarly to a biological neuron, but the former (the artificial neuron) receives various inputs in the form of numbers and performs some actions based on several calculations (Holmes et al., 2019).

Figure 3

A representation of how an artificial neuron works

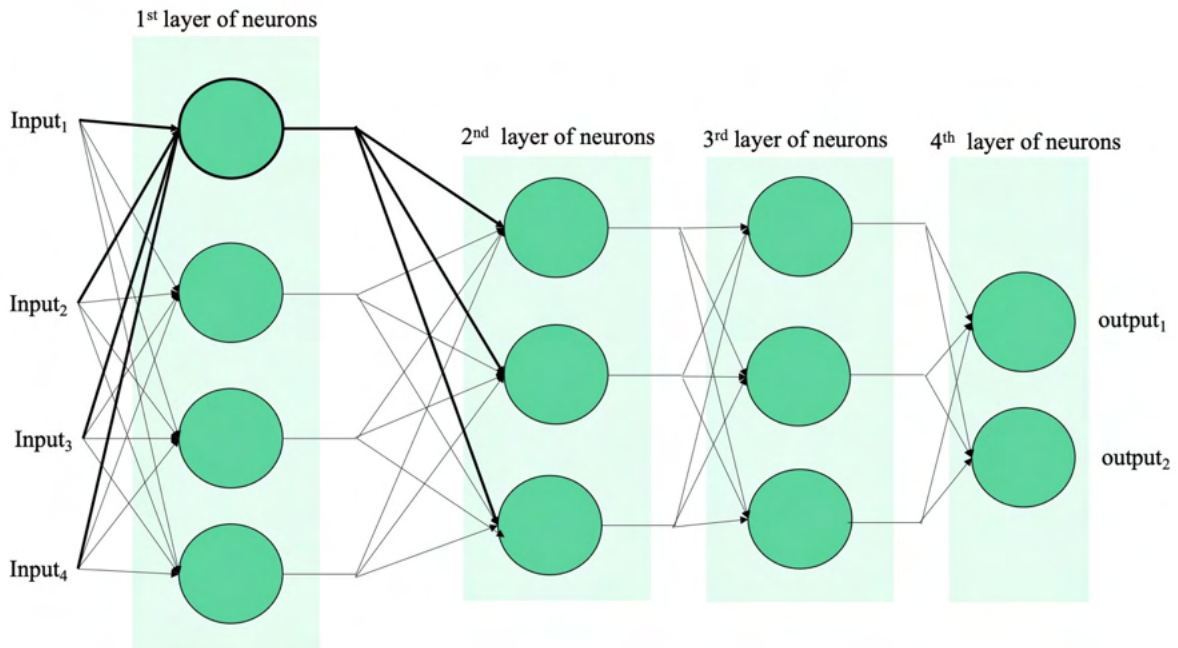


Note. Adapted from: “Artificial Neural Networks Technology” by Anderson and McNeill (1992), pp. 3-4.

However, a single neuron is not able to learn much. Hence, it is necessary to connect many neurons to form a network, as shown in Figure 4. In this case, the network of neurons becomes a powerful ML algorithm capable of learning many features (Holmes et al., 2019).

Figure 4

A representation of a network formed by artificial neurons



Note. Adapted from: “Artificial Intelligence in Education: Promises and Implications for Teaching & Learning” by Holmes et al. (2019). Center for Curriculum Redesign, p. 92

The concept of artificial neuron was already developed in the 1950s, but only in the 1990s ANNs were able to recognize small handwritten letters and numbers (LeCun et al., 1998). The turning point of ANNs was in 2012 during the challenge of classification of images named ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). The competition concerned the AI researchers who had to elaborate the best performing AI model capable of classifying images into one thousand different classes such as sunglasses, fire engines, etc. In order to permit this achievement, the researchers were provided with 1.2 million annotated images (which represented a huge amount of data for that time) to train their AI models. This challenge was won by an ANN by outdistancing the other algorithms considerably (Krizhevsky et al., 2017). Such success was achieved thanks to a huge number of data, a bigger network formed by many layers of neurons, and more powerful hardware able to handle data faster. The ANNs with many layers of neurons (like the ones which won the ILSVRC) have been named Deep Neural Network (DNN) or Deep Learning (DL) (Bini, 2018). The ILSVRC drew the attention of the scientific community who started to employ DNNs and was capable of obtaining enormous advancement in the sectors of AI. On the one hand, DNNs achieved much progress in image classification and detection, text translation and generation, and much

more (Holmes et al., 2019; Zhang & Lu, 2021). On the other hand, DNNs were able to obtain results never seen before such as beating world champions on complex games or videogames (Silver et al., 2016; Silver et al., 2018; Berner et al., 2021). Some of these innovative algorithms were useful to be applied to the real world in order to accelerate various research fields: algorithms were able to predict accurately the shapes adopted by the protein macro-molecules reducing biology researchers' work from several months to a few hours (Tunyasuvunakool et al., 2021); algorithms were also capable of reducing the cooling bill of Google's data centres (Jones, 2018). So far, in the field of text generation, researchers have developed Large Language Models (LLMs) like GPT-3 and GPT-4, on which ChatGPT is based (ChatGPT will be deeply tackled in Chapter II). The current innovations have anew brought AI and its impacts on humanity at the forefront.

AI, however, includes several limitations and may lead the society to some risks (Nadimpalli, 2017). For this reason, it is necessary to use AI tools appropriately and ethically (UNESCO, 2022). Machines, like human beings, make mistakes and are not perfect. Popenici and Kerr (2017) mentioned two catastrophic errors committed by AI over the years: in May 2016 a pilot on a car with an autonomous driving system died as the autopilot software was not able to detect an articulated lorry. A further instance is the AI bot named Tom and created by Microsoft which went out of control on Twitter becoming a racist and narrow-minded account. Likewise, the Apple Card was investigated due to gender discrimination as the credit card had different credit limits for women and men (Vecchione et al., 2021).

In spite of the fast improvements of AI, the idea that people could only rely upon technological tools represents a perilous direction. For this reason, the *National Science and Technology Council of the United States* (2016) suggested that further research is necessary in order to elaborate efficient methods to encourage collaboration between AI and humans.

1.2. Large Language Models

AI research has been divided into different fields depending on the types of data processed (Zhang & Lu, 2021). For example, the field of Computer Vision focuses on images (Holmes et al., 2019; Zhang & Lu, 2021), while Natural Language Processing (NLP) concentrates on processing and manipulating language at various levels (phonology, morphology, syntax, semantics, and pragmatics) (Coppin, 2004). Basically, NLP represents the relation between AI and linguistics (Zhang & Lu, 2021). Language modelling is an existing task in NLP that tries

to build AI models capable of comprehending and learning languages (Russell & Norvig, 2010). Language Models (LMs) represent statistical instruments that can be trained by predicting the subsequent words in an utterance on the basis of the context provided (Russell & Norvig, 2010; García-Peñalvo, 2023). Large Language Models (LLMs) are LMs but trained on a vast quantity of text data to learn patterns of a language (Liu et al., 2023). Recently, LLMs have made relevant progress in NLP as well as had considerable impacts on language teaching and learning. LLMs are able to read, analyse and create human-like texts, write translations into several languages, answer questions as well as generate contents and tasks with high-level precision (Kasneci et al., 2023).

Over the years, many LLMs have been developed and have enhanced better and better performances thanks to continuous training on an increasing amount of data. Some of the leading LLMs are GPT-1,2,3,4 and InstructGPT by OpenAI (Radford et al., 2018; Radford et al., 2019; Brown et al., 2020; Ouyang et al., 2022), BERT (Devlin et al., 2019), PaLM by Google (Chowdhery et al., 2022; Anil et al., 2023), LLaMA by Facebook (Touvron et al., 2023), BLOOM (Scao et al., 2023), among others. All have been trained with huge amounts of data. As can be seen, there have been various LLMs before the release of ChatGPT.

1.3. Artificial Intelligence in Educational Contexts

In recent years, there has been an increasing tendency in the field of education to employ and integrate innovative technological tools (Cloete, 2017; Roll & Wylie, 2016). In fact, the majority of today's learners utilise laptops or tablets instead of notebooks, educators employ several learning platforms, for example, *Moodle*, and *Google Classroom*, and blackboards have been substituted with interactive whiteboards. At the same time, a growing number of Massive Open Online Courses (MOOCs) are present for autonomous study, for instance, *Coursera*, *Udacity*, and *Khan Academy* platforms. Furthermore, enormous progress has taken place in the field of AI, Augmented Reality (AR), Virtual Reality (VR), and robotics applied to educational systems (Chassignol et al., 2018). Clearly, the future of education is deeply connected with the progress of technologies and innovative intelligent machines. In order to deal with the digital needs of education in the future, the *European Framework for the Digital Competence of Educators* (Redecker, 2017) along with the *Digital Competence Framework for Citizens* (Carretero et al., 2017) have highlighted the importance of the acquisition of digital

competences by educators to utilise digital technologies effectively in class and by students to become active citizens in the digital era.

For almost thirty years, Artificial Intelligence in Education (AIED) whose focus is to learn with AI (Holmes et al., 2022), has been one of the most challenging topics in academic research (Tan, 2020). AIED can be defined as an interdisciplinary field which includes education, statistics, computer science, neuroscience as well as cognitive psychology (Chen et al., 2020). According to many experts (Chen et al., 2020; Cardona et al., 2023), AIED has the great power to change as well as facilitate learning and teaching scenarios by providing numerous benefits for the whole society. In fact, several recent studies (Holmes et al., 2019; Pedró et al., 2019; Luckin et al., 2016) have investigated the potential of AI tools applied to education, concluding that AI can aid in enhancing learning and teaching opportunities, reaching better results and promoting personalised learning in various ways. In respect to this, many schools have already begun to cooperate with AI virtual teaching assistants, for example, *Coach Mike* that has been created for aiding students in exploring museums and giving learners support (Du Boulay, 2016). These dual-teacher systems could help teachers, who usually are overburdened, to assess students' tests or in routine tasks, too. In this way, teachers can organize their own duties more efficiently. Whereas other institutions have adopted Intelligent Tutoring Systems (ITS) (Holmes et al., 2019; Nye, 2015) which provide students with individualised tutorials by stages based on learning needs, styles, and strategies of every schoolchild. An instance could be *MATHia* that helps students with maths by giving tailored feedback and assessments. These AI tools are also really important to find out the difficulties and potentialities of each student and based on their own needs, ITSs can recommend instructional alternatives for their learning experiences. Furthermore, Alam (2021) affirms that AI is considered a "neutral" tool in education. In fact, making mistakes in front of teachers might be awkward for some students. This issue is entirely removed when using virtual learning systems.

AIED also plays a crucial role in the realm of the metaverse. In fact, AI permits a fully virtual context called Virtual Reality (VR) or a partial virtual setting named Augmented Reality (AR) (Hwang & Chien, 2022). As regards AR and VR, Mikropoulos and Natsis (2011) have added that using virtual reality with 3D technology can aid learners in comprehending notions. Likewise, other authors (Wartman & Combs, 2018) have considered worthwhile the usage of AI in an augmented setting in order to give students the opportunity to practice medical concepts such as surgeries or explore human anatomy.

Further investigations (Timms, 2016; Sharma et al., 2019) have pointed out that the integration of AI with other technological tools may develop efficient instruments capable of enhancing instruction. For instance, Timms (2016) has mentioned the so-called cobots which are robots designed for collaborating with humans. Consequently, educational cobots represent robots that help teachers differentiate learning and tailor activities based on students' needs and styles. Similarly, Pokrivcakova (2019) has fostered the usage of chatbots which give tailored responses based on students' prompts and boost personalised and autonomous learning. An example of chatbot could be *Memrise*.

Additionally, UNESCO (2019) has assured that AI technologies can foster equity and inclusion in educational settings. In fact, refugees, people with disabilities or living outside their own communities have the great opportunity to benefit from these AI tools by accessing proper learning paths. For instance, the usage of telepresence robotics permits learners with special needs to participate in classes from hospital or home. Furthermore, AIED can boost collaborative learning when students are not physically in the same place (Holmes et al., 2019; Pokrivcakova, 2019; Pedró et al., 2019). For example, AI applications can automatically propose groups of learners who are appropriate for specific tasks or can track online discussion groups by providing teachers with data on students' conversations.

However, some concerns have been raised when handling AIED. Nowadays, both educators and students should be aware of the risks of this sector, particularly regarding security and privacy data issues as the algorithms could be easily hacked (Holmes et al., 2019; Cardona et al., 2023; Kasneci et al., 2023). Therefore, providing educators with training programmes on AIED aiming at assisting their professional development is fundamental in today's society (Miao et al., 2021). According to Holmes et al. (2019), people should understand the key problems with data collection, for instance, what data is gathered or neglected, the computational methods being employed such as which bias can be present, and the ethics of learning and teaching.

In order to address the aforementioned risks, policymakers, educators as well as AI developers should collaborate together (Pedró et al., 2019). Currently, the European Commission is working on legal instruments to regulate AI for sustainable development based on equity, transparency and ethics (García-Peñalvo, 2023; Cardona et al., 2023). Further emerging issues which education should deal with are possible promotion of plagiarism as well as cheating amidst students and researchers (Kasneci et al., 2023). In this manner, the AIED

may adversely influence students' problem-solving as well as critical thinking skills. To address this challenge, Kasneci et al. (2023) have suggested that it is essential to make students aware of AIED limitations and consider AIED as a support to improve the learning experience.

Moreover, some people are worried that AIED can replace teachers, but according to several researchers (Holmes et al., 2019; Popenici & Kerr, 2017; Chassignol et al., 2018; Chan & Tsi, 2023), it is impossible now as social interactions are necessary for effective learning. In fact, teaching practices require much more than transferring knowledge. In the future, experts (Kasneci et al., 2023) foresee that the roles of educators will evolve: their time will be used more efficiently, and their expertise will be augmented. Additionally, two scholars (Popenici & Kerr, 2017) suggested that the focus on education should always be human-centred and not machine-centred. Humans should learn to identify and analyse issues, find solutions, evaluate possibilities and follow the core values of mankind in order to foster their own critical thinking and creativity, while AI and technology should support learners to reach these achievements.

In spite of the above-mentioned preoccupations, AI applied to educational technology can support teachers to diminish their routine workload (Pedró et al., 2019): educators make numerous decisions every day, interact with students beyond class, participate in meeting with families, mediators, or caregivers, collaborate with other teachers and are in charge of administrative tasks. AI can give recommendations to teachers regarding their learners' needs by analysing students' data, aid in assessing learners' tests through the creation of rubrics, reflecting and enhancing their teaching practices. On the other side, AIED can help students with autonomous learning by building up knowledge and improving learning habits. Considering the rapid pace of new technology, it is important to reconsider the role of teachers as well as pedagogies and address these challenges as soon as possible in order to offer positive educational experiences to students (Chassignol et al., 2018; Yang, 2022).

1.4. AI in foreign or second language education

The employment of technology in language education represents a fundamental part of educators' knowledge and a twenty-first-century ability (Kessler, 2018). In the field of language learning and teaching, there are various enthralling AI-powered tools that offer more engaging, inclusive, and personalised learning. These instruments can empower both teachers and students to understand what is being studied, how it is studied, as well as how learners feel.

Moreover, AI-power tools should facilitate the efforts of teachers and students (Pokrivcakova, 2019).

Nowadays, teachers have to face significant challenges in schools (Marzano & Marzano, 2003; Lew & Nelson, 2016): class size, classroom management, presence of students with special needs, and much more. Furthermore, highly tailored support and learning are increasingly needed for each student (Mohammed & Watson, 2019; Liu et al., 2023). As a result, teachers are not always able to constantly analyse and evaluate every single output, identify particular needs of learners and give constructive and rapid feedback. Whereas AI instruments are capable of collecting huge amounts of data concerning progress of the students and providing them with instant criticisms. Hence, AI could be a great support for teachers (Mohammed & Watson, 2019).

The application of AI in language education represents a subset of Computer-Assisted Language Learning (CALL) and its growing popularity has led to the development of Intelligent CALL (ICALL) (Kannan & Munday, 2018; Pokrivcakova, 2019). Pokrivcakova (2019) argued that ICALL has many benefits for each learner, including immediate feedback and assessment, understanding of learning strategies as well as preferences, decreasing of time and frustrations of students when carrying out tasks without receiving instant feedback, repetitions of topics in which the learner feels less uncomfortable, etc. These functions are possible thanks to the data collection, storage, analysis, and the ML algorithms that are able to calculate the weaknesses and strengths of each student, predict their behaviours, and provide customised learning materials.

There are several ways of using AI in language education (Pokrivcakova, 2019). The first instance regards Machine Translation (MT) which is the process of employing AI to translate a text from one language to another in an automatic way such as *Google Translator*. Garcia and Pena (2011) investigated the usefulness of MT in language learning discovering that machine translators can aid beginner language students in improving their language proficiency, while Lee (2019) found out that MT refine learners' revision abilities and help to reduce grammar and lexical mistakes. Various foreign language teachers, however, consider the usage of MT as a failure of the student or a way of cheating (Case, 2015) and decide to forbid its use in the classroom due to their scarce-quality outputs (Steding, 2009). Nevertheless, it has been found that the ban on ML is not effective (Steding, 2009). In fact, it would be better for teachers to guide students in learning how to use machine translators in an effective way.

The second example of AI in language education is chatbot (Pokrivcakova, 2019). Chatbots represent computer programmes which simulate a human conversation in the form of chat such as *Memrize* or *ChatGPT*. These tools can enhance students' language competence as chatbots can provide tailored answers to learners' prompts and suggestions about where to improve. Moreover, chatbots can support autonomous learning (Haristiani et al., 2022) and increase motivation (Kohnke, 2022).

A further example is given by Intelligent Virtual Reality (IVR) such as *Immerse* (Pokrivcakova, 2019). IVR is a virtual environment constituted of speech and facial recognition, AI tools, and NLP where students can complete tasks or speak with avatars that imitate real conversations.

Another option to use AI in language classes is online language learning software such as *Duolingo*, *Babbel*, and *Busuu* (Pokrivcakova, 2019). Some platforms and apps include just digital exercises, for example, matching exercises, and filling the gaps, without giving space to the creativity of students. While others imitate an interaction with the help of a virtual tutor.

In order to improve students' writing skills, AI writing assistants could be introduced throughout language lessons (Pokrivcakova, 2019). They are founded on ML and NLP. Their principal function is to scaffold the writing process of the students. AI writing assistants such as *Grammarly* can identify and correct grammatical mistakes and give tips for further improvement.

Thanks to these AI tools, language students can be active actors in their learning process instead of being passive receivers of information. Additionally, by choosing their own pace of learning, learners have the opportunity to become more autonomous (Pokrivcakova, 2019). On the other hand, teachers should adopt a more student-centred approach by guiding and supporting their students during the learning path (Kessler, 2018). In conclusion, every AI-powered tool used in language education could present potentialities, limits, and risks. It is the responsibility of each teacher to deal with them and instruct students in order to take full advantage of AI in language learning (Kessler, 2018; Pokrivcakova, 2019).

1.5. Conclusion and gaps in the literature on AI in education

The aim of this chapter was to point out the complexity of AI with all its rapid advancement and highlight the importance of employing AI-powered tools in foreign or second language education. In fact, knowing how to use technological instruments properly and ethically is a

digital competence required today in order to become active citizens in the world (Holmes et al., 2022).

Given the complicated nature of AI, educators play a fundamental role in this context (Kessler, 2018; Pokrivcakova, 2019). Teachers, indeed, should integrate the ICALL instruments into their teaching practices and help students be aware of their risks as well as potentialities. Regarding foreign or second language education, language instructors do not need to be AI experts to use Duolingo, chatbots or machine translators in their classes. It is fundamental to mention that the sector of AI does not involve only software engineers, statisticians or mathematicians, but AI is part of our daily life and the whole society should be conscious of its capabilities and dangers in order to use AI-powered tools correctly (UNESCO, 2022).

With the introduction of AI in education, however, further research is needed concerning the new roles and skills of instructors, especially in higher education (Popenici & Kerr, 2017). Moreover, innovative tasks focused more on creativity and critical thinking skills are necessary. Another gap in the literature includes a general lack of research on the effects that AI can entail in teaching and learning (García-Peñalvo, 2023), especially in language education (Liang et al., 2023).

CHAPTER II. CHATGPT

The release of ChatGPT is revolutionising several settings, including education. In this Master's thesis, ChatGPT is considered as an AI tool included in AIED. The current chapter will present a state of the art of ChatGPT. The section will delve into its development, potential and limits. Moreover, the chapter will contain an analysis of the literature on ChatGPT in the field of teaching, learning as well as assessment, particularly in higher education. At the end of the chapter, some research gaps and future research directions will be reported.

2.1. Definition and Development of ChatGPT

Chat Generative Pre-Trained Transformer (ChatGPT) represents an LLM-based AI chatbot designed by one of the leading AI companies named OpenAI (OpenAI, 2022). The firm was founded in 2015, and its headquarters is located in San Francisco. Its principal aim is to develop and promote AI capable of benefitting the whole society (Brockman et al., 2015). OpenAI launched ChatGPT in November 2022, which was based on the so-called language model GPT-3.5 or InstructGPT and trained on a large amount of data in order to comprehend natural language as well as generate answers based on the prompts of the users. In fact, the AI tool is capable of elaborating texts and maintaining human-like conversations in a multilingual context (García-Peñalvo, 2023). Although ChatGPT supports more than fifty languages, investigations into the AI chatbot have been mainly focused on the English language up to now (Kasneci et al., 2023).

Several steps forward have been taken since its release, and OpenAI has pointed out that ChatGPT will be continuously updated so as to become even more powerful (Halaweh, 2023). Nowadays, it is possible to apply ChatGPT to other applications, for instance, some educational-tech firms such as Duolingo (Duolingo, 2023), and Quizlet (Bayer, 2023) have already integrated ChatGPT into their systems.

Concerning the history of the series of language models of Generative Pre-Trained Transformer (GPT), OpenAI was able to develop the first GPT (GPT-1) in 2018 (Radford et al., 2018). GPT-1 was one of the first models to be trained on huge amounts of data by simply assigning it the task of predicting the following word in a text. At that time, results demonstrated that the outputs of GPT-1 matched or were even better than the other models, which were trained only for specific tasks. An example of one of these tasks is provided in Figure 5.

Figure 5

Examples of the Choice of Plausible Alternatives (COPA) task used in the field of NLP to evaluate the trained models on common-sense reasoning

Example 1:

Premise: *The man fell unconscious. What was the cause of this?*

Alternative 1: *The assailant struck the man in the head.*

Alternative 2: *The assailant took the man's wallet.*

Example 2:

Premise: *I knocked on my neighbour's door. What happened as a result?*

Alternative 1: *My neighbour invited me in.*

Alternative 2: *My neighbour left his house.*

Note. In this task, the models were provided with a premise and two alternatives. The models had to choose the most plausible alternative for the premise. Reprinted from: "Choice of Plausible Alternatives: An Evaluation of Commonsense Casual Reasoning" by Roemmele et al. (2011). *AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, pp. 91-92.

In 2019, OpenAI elaborated a GPT of higher quality called GPT-2, which was trained on more written data (Radford et al., 2019). The AI company was concerned about the results obtained with GPT-2 (OpenAI, 2019): the model produced some texts which were difficult to distinguish whether they were written by humans or machines. Thus, OpenAI initially decided to release only the smallest version of GPT-2 and launch the rest of the versions gradually. In the meantime, the researchers of OpenAI worked on its safety (Askell et al., 2019).

In 2020, OpenAI achieved a significant breakthrough: the company released GPT-3 which represented an upgrade of the previous models of GPT. In fact, GPT-3 was more powerful and provided better results as it was more than one hundred times larger than its predecessor (Brown et al., 2020). However, the AI firm took the decision to provide an Application Programming Interface (API) to access GPT-3 for a fee (Brockman et al., 2020). OpenAI adopted a new approach: differently from GPT-2, which was published as an open-source, individuals could use only the API version which did not allow to modify the parameters, and OpenAI could supervise what people requested to the model.

In 2022, OpenAI announced GPT-3.5 or InstructGPT in order to cope with the harmful sentiments, toxic or untrue outputs generated by GPT-3 (Ouyang et al., 2022; Lowe & Leike,

2022). This was partially due to the fact that GPT-3 was trained to foretell the following word on large datasets of texts, instead of performing the language assignment that the users wanted in a safe way. Basically, the models were not aligned with users' requests. InstructGPT, instead, was able to follow users' prompts as this model was based on GPT-3 and also trained with the recent technique named Reinforcement Learning from Human Feedback (RLHF). The details of RLHF are out of the aim of this thesis, for more information see Ouyang et al. (2022) and Christiano et al. (2017).

On 30th November 2022, OpenAI introduced a new AI chatbot named ChatGPT (OpenAI, 2022). Basically, the chatbot was built on InstructGPT and made available for the population through a user-friendly as well as free web chat interface (Rudolph et al., 2023). However, there are other alternatives for ChatGPT, such as Bard developed by Google (Pichai, 2023) and Open Assistant (Kilcher, 2023).

In March 2023, OpenAI presented GPT-4 (OpenAI, 2023). The model can accept images as well as texts as inputs and produce text outputs. Furthermore, GPT-4 is larger than its predecessor and has been trained with the same technique of InstructGPT, and on more data. However, OpenAI has not provided information concerning its architecture and size. Subsequently, OpenAI has launched a version of ChatGPT based on GPT-4. Subsequently (always in March 2023), the company decided to add plugins to ChatGPT, which allow ChatGPT to surf the internet to search for information or execute some simple code programmes (OpenAI, 2023). Figure 6 shows the history of OpenAI's GPT model.

Figure 6

History of OpenAI's GPT model

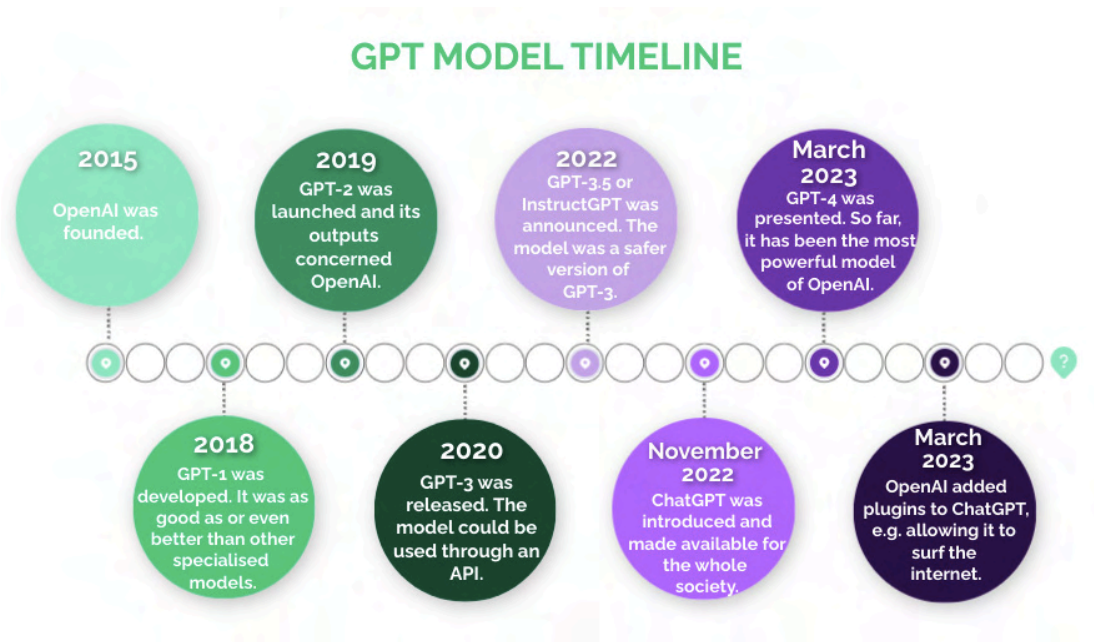


Figure 7 displays the outputs provided by the different GPT versions of the same prompt.

Figure 7

Outputs given by the several versions of GPT answering the same prompt

Input: *In a shocking finding, scientists discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

GPT-1: There were several sightings of the unicorns while the humans didn't . a few weeks later the unicorn who had led them to the site of the encounter was seen flying in a helicopter, and the next day the same pair appeared. at that point, the humans had already been informed that the unicorns were very near the area where the battle between the two groups had first occurred. the discovery, as well as the discovery of the unicorns , was extremely helpful in keeping the humans away from the area. " the unicorn, who was originally made of copper, was actually a unicorn of a different breed. unicorns are quite rare . the unicorns of the past have been considered extinct ever since. "

GPT-2: The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

GPT-3: The herd of 26 unicorns was discovered by a team of researchers from the University of Quebec in Canada while conducting a survey of the region. According to the team's report, the unicorns were "initially shy but quickly warmed up to the researchers and allowed them to approach and photograph them."

The researchers say that they didn't believe their eyes when they first saw the unicorns. "We were completely shaken when we first saw the herd of unicorns.

GPT-3.5: In a state of awe and disbelief, the team of scientists stood at the edge of the valley, watching as the magnificent herd of unicorns grazed peacefully in the lush meadows. Their shimmering, iridescent coats glistened under the sunlight, reflecting an array of colors that seemed to defy the laws of nature. But what truly stunned the researchers was the fact that these mythical creatures were not only real but capable of communicating in flawless English.

GPT-4: The international community was rocked when the news broke out. Footage of the remarkable discovery quickly went viral, showing the majestic creatures gracefully trotting around their isolated paradise, their shimmering horns standing out against the backdrop of the snowy Andean peaks. But the most bewildering part of it all was the unicorns' ability to communicate. They spoke in clear, articulate English, which left scientists puzzled and the public in utter disbelief.

Note. The prompt was taken from "Language Models are Unsupervised Multitask Learners" by Radford et al. (2019), *OpenAI*. The several outputs were generated using: GPT-1: <https://transformer.huggingface.co/doc/gpt>; GPT-2: output taken from "Language Models are Unsupervised Multitask Learners" by Radford et al. (2019), *OpenAI*; GPT-3: <https://platform.openai.com/playground?model=text-curie-001>; GPT-3.5: <https://chat.openai.com/?model=gpt-3.5-turbo>; GPT-4: <https://chat.openai.com/?model=gpt-4>

Since the number of users is growing on ChatGPT, the AI chatbot has caused such a fuss that reactions have been really reverse. For instance, the researcher García-Peñalvo (2023) has stated in his research paper that while some scholars consider ChatGPT as the best disruptive AI model so far, other academics prefer to resist change because they perceive ChatGPT as a danger and threat. The ongoing debate has especially concentrated on the fields of scientific production as well as education (Grassini, 2023).

As regards the awareness of the public, the impact of AI is also influenced by the media. Some analysis carried out by Ouchchy et al. (2020) suggested that the media did not present scientific knowledge when talking about policy and ethical issues regarding emerging technologies, including ChatGPT. Furthermore, Sullivan et al. (2023) collected and examined

one hundred news articles (published at the end of 2022 and the beginning of 2023) to comprehend better to what extent ChatGPT is changing the learning and teaching environments, precisely in Australia, New Zealand, the US and UK. The results showed that there has been a central focus on issues related to ChatGPT usage, such as plagiarism, cheating during exams, and loss of the critical thinking skills of individuals, instead of concentrating on the opportunities this AI tool may give (personalizing assignment tasks, giving simpler explanations, understanding difficult concepts better, providing more constructive and useful feedback, creating rubrics, so on and so forth). Several institutions even chose to ban ChatGPT (García-Peñalvo, 2023; Halaweh, 2023; Sullivan et al., 2023) due to the fear that students could copy during the assignment tasks. According to some experts on AI in education (Kasneci et al., 2023), this prohibition was not the correct solution to the issue, actually, the above-mentioned action represented only a straightforward way out. Hence, it would be important to normalize the fact that AI tools are present, and related advancement will take place in the future.

2.2. ChatGPT in teaching and learning

Over the years, technology has revolutionised the field of education. Especially today, AI is playing a crucial role in changing teaching and learning realms (Chassignol et al., 2018). The hype on ChatGPT has led to profound dismays in education, in particular in the future of higher education. However, numerous scholars (Kasneci et al., 2023) belonging to different research areas (social, data, computer as well as educational sciences) at Ludwig Maximilian University (LMU) and Technical University of Munich (Germany) have collaborated together and displayed that LLMs and chatbots, particularly ChatGPT, can enhance learning and teaching experience at all degrees of education (primary, secondary, university, remote learning, learners with disabilities as well as professional training).

Concerning learning, the above-mentioned scientists (2023) underlined the fact that every student presents different learning abilities, needs and styles. In these contexts, ChatGPT may act as a tutor or assistant, and aid in providing personalized learning for learners. For instance, ChatGPT may be utilised to create quizzes or simulations of exams, provide further explanations about difficult concepts, outline or summarise texts, proofread writing, help to find information, give tailored feedback, solve doubts, and much more. In this way, ChatGPT might also boost students' autonomy in their learning path. Regarding teaching practices

(Kasneci et al., 2023), ChatGPT could support educators and change teaching methodologies: ChatGPT comes to the aid of teachers and enables them to design personalized learning for every single student and create engaging, interactive materials, lesson plans, as well as exercises. Such assistant may save efforts and time for teachers. As regards research and professional development (Kasneci et al., 2023), ChatGPT could help teachers summarise or explain innovative teaching methodologies or technologies. In this way, educators could be up to date with the most recent developments in the field of education.

When the present Master's thesis was written, very few experiments were carried out concerning the implementation of ChatGPT in class. Some studies (Ali et al., 2023; Muñoz et al., 2023) demonstrated that using ChatGPT in class increases the motivation, and engagement of students, especially in foreign language learning. In fact, outcomes suggested that ChatGPT is able to stimulate students into enhancing writing as well as reading abilities and should be utilised in class as a learning instrument aiming at increasing students' performances.

Additionally, a pilot study was conducted in physics education (Bitzenbauer, 2023). The aim was to demonstrate how ChatGPT can be utilised during lessons in order to boost the critical thinking capabilities of 53 secondary students. The teacher was able to guide the students in the usage of ChatGPT. In the end, results showed that learners had positive perceptions regarding ChatGPT and recognised its benefits. Moreover, data demonstrated that students would like to incorporate this tool into their everyday lives.

Another experiment conducted by Su et al. (2023) illustrated an example of implementing ChatGPT to simplify the argumentative writing process. Evidence revealed that ChatGPT might be a valid support for students concerning linguistic as well as structural features. Furthermore, ChatGPT is able to provide tailored feedback, give advice about content planning, carry out language analysis and proofread passages. Notwithstanding, ChatGPT's responses strictly depend on the prompts and rubrics given by users and cannot provide guiding questions in order to scaffold learners' thinking.

Similarly, Yan (2023) conducted an exploratory investigation into the effects of ChatGPT on undergraduate learners' behaviours as well as attitudes during an L2 writing practical study. Basically, the practicum was divided into three stages adopting the flipped-classroom approach: familiarization of ChatGPT, experiment and exploration. Findings highlighted that ChatGPT could be a potential instrument in L2 writing experiences. Moreover, data showed that students were quite worried about its threats concerning academic honesty.

A further exploratory research was carried out on the usage of ChatGPT in science education (Cooper, 2023). The researcher adopted a self-study approach. Results showed that the answers given by ChatGPT were in agreement with the principal topics of the investigation. In addition, the tool seemed to be useful for creating exercises, rubrics as well as science units for teachers. However, educators should consider ChatGPT as just an assistant, and evaluate any resource designed by the chatbot. Furthermore, the researcher found out that there were several ethical issues regarding AI such as copyrighting. For this reason, educators should instruct students to use ChatGPT responsibly and assign priority to critical thinking.

In another study, Qureshi (2023) used ChatGPT as a learning and assessment tool in an undergraduate curriculum of computer science. The research included a control and an experimental group. Results showed that students who employed ChatGPT achieved higher scores than the control group. However, the group was not capable of achieving perfect scores as the code generated by ChatGPT presented some inaccuracies. The researcher concluded by stating that it is noteworthy to construct a learning environment where hands-on and active learning experiences are fostered, and the reliability of these AI resources, for instance, ChatGPT, is discouraged. Since AI instruments capable of solving learning tasks will not stop appearing, learners may be enticed to cheat in case they are not able to comprehend the core of their learning path. Therefore, teachers have the responsibility to explain the importance of students' work as well as the advantages of acquiring new knowledge and skills through practice.

2.3. ChatGPT on the assessment practices

Nowadays, students have the opportunity to obtain considerable help with their assigned tasks thanks to the introduction of vibrant AI technology, principally ChatGPT. A particular concern about ChatGPT is its impact on language teaching as well as assessment, especially in higher education. For this reason, rethinking innovative approaches to assess students' outcomes is more necessary than ever (Moqbel & Al-Kadi, 2023; Timmis et al., 2015).

Assessing students is an integral part of teaching practices. The term assessment in instruction embraces the process of gathering both quantitative and qualitative evidence from several students' sources in order to aid teachers in planning instruction, making decisions regarding pupils' schooling and guiding learners during their learning process (Brown & Lee, 2015). On the basis of this fundamental source of information, educators have the opportunity

to comprehend both the weaknesses and strengths of each student and then, tailor and adjust the teaching planning to optimise their learning path (Opre^a, 2015). According to Maugeri (2020), one of the principal aims of assessment is to make students aware of their efforts, and mistakes without devaluing learners. It is also important that teachers share assessment modalities (what and how), timing as well as criteria with students in order to avoid stress, and panic amidst the students.

There are two principal types of assessment: summative and formative assessments (Brown & Lee, 2015; Brown & Abeywickrama, 2018). The former regards giving learners a final mark when concluding a specific educational period such as a unit, project, year, etc. in order to demonstrate their accomplishments based on the learning objectives. An example could be final exams at the end of university courses. Summative assessment is also called *assessment of learning*. While the latter takes place several times throughout the course and includes:

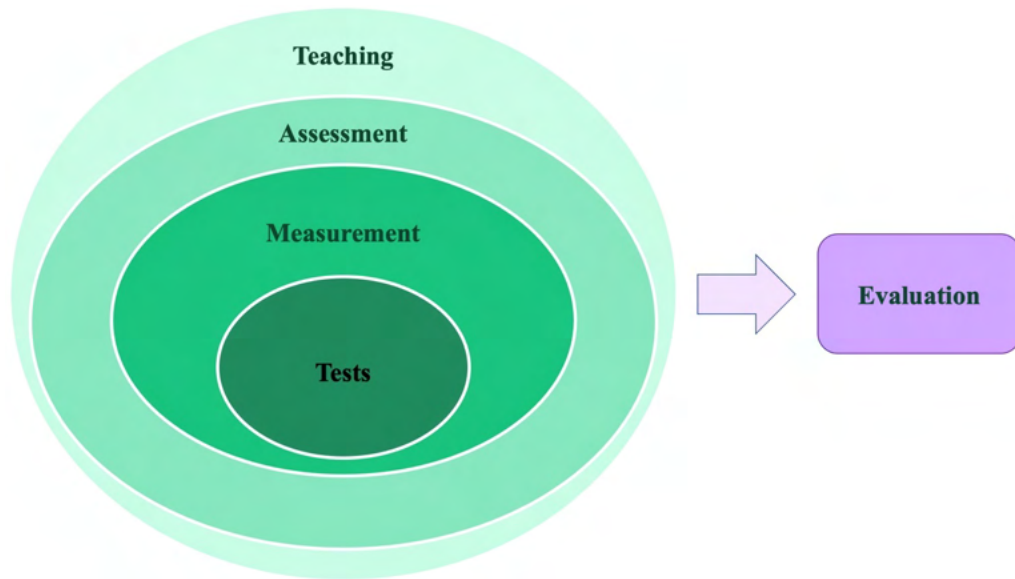
- *Assessment for learning* which concerns providing students with feedback and information on their advancement towards their learning path. The aim is to enhance students' performances as well as the course itself. An example could be a brainstorming of students' prior knowledge at the beginning of the lesson.
- *Assessment as learning* in which learners reflect upon and check their advancement based on their future learning objectives, for instance, keeping a learning diary, self- or peer assessment.

According to Morton and Escobar Urmeneta (2019), and Maugeri (2020), a usual criticism of the approach used to assess all levels of instruction is that educators mainly implement summative assessment. Basically, teaching practices and assessment are separated: firstly, educators teach, and then, they administer a test to verify whether students acquired the knowledge previously taught. The present practice represents a real peril: relying on summative assessment excessively, there could be a risk of learners' failing since receiving only marks represents a selective, and discriminatory teaching practice. A further criticism concerns the misunderstanding of some instructors about the correct implementation of formative assessment: teachers think that they should assess students all the time. Instead of giving only marks, the main goal of formative assessment is to provide learners with constructive feedback in order to enhance their own performances. This could especially happen in foreign language lessons in which learners need multiple occasions to practice the language and obtain feedback on their performance (Morton & Escobar Urmeneta, 2019).

In addition to assessment, it is relevant to comprehend the broad concepts of evaluation, measurement as well as tests. Sometimes these terms are used interchangeably. However, they all are interrelated (Figure 8).

Figure 8

Interrelation of tests, measurement, assessment, teaching, and evaluation



Note. Reprinted from “Language Assessment: Principles and Classroom Practices” by Brown and Abeywickrama (2018). Pearson, p. 7.

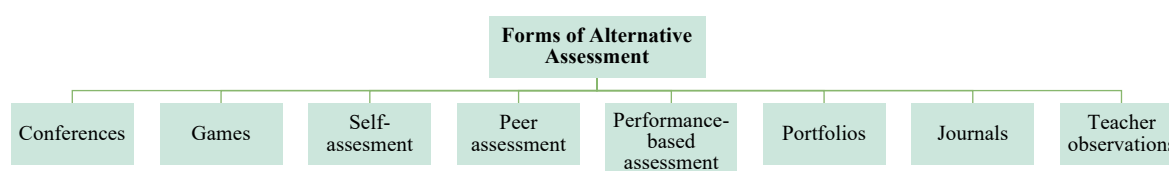
Evaluation entails the interpretation of the pieces of information, for example, evaluation is used to take decisions based on test results. It also refers to making statements about the courses, checking the effectiveness of teaching practices or suitability of instructional materials based on various quality standards (Brown & Abeywickrama, 2018; Morton & Escobar Urmeneta, 2019; Piccardo et al., 2011).

The term *measurement* represents the operation of quantifying the learners’ performances observed in a classroom (Brown & Abeywickrama, 2018). One traditional form of discovering what learners have acquired in their studying process is to do a *test*. According to Brown and Abeywickrama (2018) a test represents a measuring tool that provides information about what students know. However, this method included in the traditional assessment is little used and is not fairly effective in comprehending what skills and knowledge students have acquired (Moqbel & Al-Kadi, 2023). In fact, starting from 1990 the alternative

assessment emerged (Brown & Abeywickrama, 2018). Recently, language educators have not administered only traditional tests, but they have also made numerous efforts in order to elaborate alternative ways of assessment in authentic tasks (Figure 9), for instance, portfolios, journals, conferences, observations, self- and peer-assessments, games, and performance-based assessment (Brown & Lee, 2015; Moqbel & Al-Kadi, 2023). This type of assessment is more authentic and focused on communication as students have to explore, resolve problems, and demonstrate how they communicate with the language (Brown & Abeywickrama, 2018).

Figure 9

Different forms of alternative assessment



Note. Adapted from “Foreign Language Learning Assessment in the Age of ChatGPT: A Theoretical Account” by Moqbel and Al-Kadi (2023). *Journal of English Studies in Arabia Felix*, 2(1), p. 73.

Considering the modern technological progress, especially with the invention of ChatGPT, some researchers (Rudolph et al., 2023) invited language educators to assign tasks to their students that require higher critical thinking levels to be produced. According to the literature, it is not simple to describe the concept of critical thinking (Bellaera et al., 2021; Liu et al., 2014; Uribe Enciso et al., 2017). For the purpose of this thesis, critical thinking refers to the combination of several abilities (analysis, reasoning, evaluation, deduction, problem solving) (Bellaera et al., 2021) which are fundamental to be aware of mistaken information, handle and understand large amounts of information, and recognise personal beliefs from verified facts (Thornhill-Miller et al., 2023). In order to foster critical thinking, the scholars (Rudolph et al., 2023) suggested some examples: elaborating digital materials (videos, animations, audio, etc. in which students discuss their personal works or provide metacognitive detailed analysis of their writing processes), giving oral presentations, focusing on conversational interactions, carrying out more flipped classes to guarantee that the most crucial parts of the tasks are handled in the classroom, permitting learners to express themselves in topics that are interesting for

them by adding personal positions and experiences, using authentic assessment, and fostering teamwork and peer evaluations.

Over the years, there have been growing concerns over academic honesty, plagiarism as well as authenticity (Kasneci et al., 2023; Moqbel & Al-Kadi, 2023; Rudolph et al., 2023; Crawford et al., 2023) as ChatGPT and other AI-powered tools are capable of generating texts rapidly and giving alternative solutions. These AI instruments might be considered a threat to elaborate written assignments such as essays. In fact, some scholars (Fitria, 2023; Yeadon et al., 2023) seem to be sceptical about utilising ChatGPT since students can be dependent upon this chatbot to elaborate their writing assignments. For this reason, several universities chose to ban the usage of ChatGPT from their students. However, Rudolph et al. (2023) argued that ChatGPT could be seen as an opportunity to support learning and introduce both innovative assessments and teaching strategies. They also pinpointed an approach that constructs trustful relationships between students and teachers, uses a student-centred pedagogy and focuses more on assessment *for* and *as* learning instead of conducting only assessment *of* learning. Lastly, learning goals, teaching, assessment practices, and learning should be effectively aligned.

A further point to consider is that even teachers have the possibility of using plenty of AI instruments to assess the works of their students such as Automated Essay Scoring (AES) or Automated Writing Evaluation (AWE). AES are AI-powered tools that are able not only to evaluate but also to give feedback on learners' essays through computer algorithms (Mizumoto & Eguchi, 2023; Rudolph et al., 2023). This AI model can help instructors reduce their workload, evaluation bias such as subjectivity, fatigue, etc., and focus more on the important features of writing, for instance, writing strategies, structure as well as coherence. A study conducted by Mizumoto and Eguchi (2023) discovered that language model chatbots like ChatGPT could be efficiently implemented as AES instruments.

AWE systems, instead, are instruments which give feedback on the argumentative writings as well as language features (Shi et al., 2022). In this way, teachers have the opportunity to reduce their time dedicated to giving complete feedback on both content and language aspects. An investigation carried out by Su et al. (2023) tested ChatGPT as an AWE system. Based on the evidence, the researchers sustained that ChatGPT seems to be a rather helpful instrument to proofread texts and is able to provide tailored feedback.

2.4. Limits and potential of ChatGPT

Since the launch of ChatGPT, dynamic debates have arisen in research communities on its possible dangers and advantages both for teachers and students. In recent times, various academics (Rudolph et al., 2023; Zhou et al., 2023; Jalil et al., 2023; Su et al., 2023) have tried to experiment with this AI LLM chatbot in order to discover its limits as well as potential.

Rudolph together with other researchers (2023) carried out several investigations into ChatGPT. For instance, the scientists asked ChatGPT to provide an explanation of the following concepts: standard deviation, z-scores, and mean. The AI tool was able to give a clear conceptual clarification. On the basis of the experiments' results, the academics made a list of what ChatGPT is able to do and not to do. Generally, the scholars understood that ChatGPT is capable of generating human-like answers in just a few seconds based on the users' prompts, keeping prior conversations, recognising improper questions and admitting its errors in case users point them out. However, it is fundamental to take into consideration that the style and quality of the prompts may influence the outputs of ChatGPT (Jiao et al., 2023).

In spite of its potential, ChatGPT also presents some challenges as well as weaknesses that could affect both its accuracy and performance. With respect to this, Rudolph et al. (2023) also provided some examples of disadvantages, such as biased content, production of wrong information, and no knowledge of events that happened after the year 2021. In fact, ChatGPT has been trained with data till the year 2021. Besides, the same academics (Rudolph et al., 2023) examined whether ChatGPT was able to elaborate an analytical as well as critical essay. Data demonstrated that the AI tool could write the text within two minutes, but the content was too descriptive and generic without any evidence reported. For this reason, it could be stated that ChatGPT has a scarce competence in requests that necessitate higher-order thinking abilities. Furthermore, ChatGPT did not provide any real references. Indeed, the AI instrument could invent the sources.

Likewise, another study (Jalil et al., 2023) analysed how well ChatGPT answers questions included in a software testing curriculum. Findings demonstrated that the outputs provided by ChatGPT to students might not always be correct (in this experiment, ChatGPT replied correctly or quite correctly to only 53% of queries). However, evidence also confirmed that utilising some prompt strategies, which give further context, may enhance the possibilities of proper explanations and answers. Thanks to this experiment, the researchers suggested that ChatGPT could be used in class to scaffold learners through some exercises in order to enhance

their comprehension of the materials. Moreover, results showed that ChatGPT has more difficulty in answering correctly regarding certain settings, for instance, coding. Similarly, Zhou et al. (2023) discovered that ChatGPT often gives wrong output concerning logic or math issues.

Concerning translations, Rudolph et al. (2023) investigated whether the AI tool could communicate in other languages apart from English. Basically, the researchers asked ChatGPT to produce an English essay and then translate it into Chinese. Outcomes revealed that ChatGPT did not translate perfectly into Chinese. Further related studies (Jiao et al., 2023) also confirmed this result: albeit the translation skills of ChatGPT were in competition with the products elaborated by Google Translator, the authors stated that ChatGPT might not maintain high standards in distance languages, for instance, Chinese.

In regard to writing assignments, Su et al. (2023) found out that ChatGPT may help learners outline a text and choose the right language features, can proofread the text, and provide students with personalised feedback or suggestions on the organization of the structure, etc. However, ChatGPT's responses deeply depended on the quality of the inputs and rubrics given by the users. The scholars also pointed out some limits of ChatGPT: the AI LLM chatbot was not able to give proactive directions on writing and may sometimes generate nonsensical or misleading outputs. Another limit of utilising ChatGPT in writing lessons included ethical issues (plagiarism and authorship) that should be addressed soon by the collaboration of educators, policymakers, and software engineers.

A further investigation (Guo et al., 2023) outlined some differences between ChatGPT and human beings. On the one hand, ChatGPT elaborated more descriptive, informative, formal, and objective texts without conveying its feelings. On the other hand, humans tend to be more colloquial and capable of sharing their own emotions. Furthermore, ChatGPT usually gives responses that are rigidly centred on the literal sense of the given prompt. People, instead, are able to diverge and understand possible hidden meanings of the questions on the basis of their knowledge, and common sense.

In conclusion, it could be mentioned that ChatGPT may be a good support, but it would be a mistake to completely rely on its answers for relevant matters. Thus, individuals should consider the above-mentioned limitations of ChatGPT, and employ the AI LLM chatbot together with human judgment as well as analysis (Rudolph et al., 2023). Since ChatGPT has been released lately, a complete description of its potential, drawbacks, and impacts is still to

be provided (Jalil et al., 2023). For this reason, researchers, educators, and policymakers have to carry out a lot of work as well as investigations regarding honesty, strengths, and limits of ChatGPT should be conducted (Alshater, 2022).

2.5. Conclusions and future research directions

By analysing the existing bulk of research, the author has showcased the history and spread of ChatGPT. The AI LLM chatbot is being applied in various domains, in the educational field as well. In recent times, ChatGPT has raised important questions concerning the future of instruction. In fact, ChatGPT is transforming the teaching and learning environments, especially in higher education. As described previously, numerous researchers (Rudolph et al., 2023; Zhou et al., 2023; Jalil et al., 2023; Su et al., 2023) have carried out several investigations into the potential, limits, and impacts of ChatGPT. On the one hand, data showed that ChatGPT could be a useful assistant both for teachers and students. On the other hand, evidence confirmed that regardless of its potentialities, ChatGPT still faces some challenges, and presents some limits. One debate focuses on the ethical usage of ChatGPT in the educational field. Major concerns regard cheating, plagiarism, and the ways ChatGPT may affect assessment practices. A further worry regards the loss of critical thinking skills and creativity of individuals. In order to deal with potential issues related to academic integrity, educators should revolutionise the tasks with the aim of preparing better future citizens. All instructors are advised to give learners higher-order assignments which emphasise logical reasoning, creativity as well as critical thinking. All capabilities that ChatGPT does not seem to possess and perform up to now (Zhai, 2022; Su et al., 2023).

According to some researchers, the quality of ChatGPT's outputs depends on the contexts, quality of the prompts, as well as biased data. However, given the rapid progress achieved by the AI so far, ChatGPT may be soon updated and trained with new data in order to improve its performance (Sohail et al., 2023). For this reason, it is fundamental that educators as well as students should be guided through the process of using ChatGPT in order to develop some digital competences and be able to employ such tools in an ethical, pedagogical and proper manner. This procedure should include how students and teachers can interact with ChatGPT and be aware of its risks (Kohnke et al., 2023). Furthermore, Su, Lin, and Lai (2023) pointed out that learners should also learn to evaluate ChatGPT's outputs critically, while

teachers should give constant feedback and scaffolding to pupils when integrating ChatGPT into lessons.

Concerning the gaps in the literature to address and the future directions of research on ChatGPT, it has been revealed that further research is needed to examine its potentialities applied to language education, and adverse consequences so as to support instructors in diminishing the negative effects of ChatGPT in language programmes (Ali et al., 2023). Moreover, there has been little prior research throughout the advancement of AIED into how teachers utilise, and learners interact with ChatGPT (Rudolph et al., 2023). In conclusion, many scholars also highlight that future studies should concentrate on the investigation of:

- The impact that ChatGPT has on the perceptions of learners, and improvement of critical thinking skills of students (Bitzenbauer, 2023; Su et al., 2023).
- The practices of integrating ChatGPT into classroom lessons (Bitzenbauer, 2023).
- The introduction of innovative pedagogical approaches (Muñoz et al., 2023).

CHAPTER III. METHODOLOGY

This chapter focuses on the methodology adopted in the exploratory study. First, the research context will be delineated considering the previous literature review and the research gaps. Second, the research questions, corresponding hypotheses, as well as objectives of the present study will be outlined. Third, the sample, the data collection instruments, the description of the activity carried out by the participants, and the data collection procedure will be described in depth. At the end of the chapter, the data analysis procedures will be presented.

3.1. The research context

The introduction of ChatGPT has had a dramatic impact on the educational field (Ali et al., 2023; Bitzenbauer, 2023; Cooper, 2023; Kasneci et al., 2023; Muñoz et al., 2023; Qureshi, 2023; Su et al., 2023, Yan, 2023;). As highlighted in Chapter II, an increasing number of investigations into the limitations and potential of ChatGPT have been recently undertaken (Rudolph et al., 2023; Zhou et al., 2023; Jalil et al., 2023; Su et al., 2023). However, considering the literature review as well as the research gaps, limited experiments have been carried out to explore how students interact with ChatGPT, how the AI chatbot may affect the critical thinking skills of students, and how teachers may utilise the AI tool. The present research tries to fill these gaps by comprehending the usefulness of ChatGPT in designing and assessing activities for language students, as well as in exploring the ways university students at Master's level, who may become future teachers, interact with ChatGPT.

The present Master's thesis deals with an exploratory study, conducted with a very specific sample in a very specific context. The research was carried out in the Master's course of *Instructional Design for Language Education* at Ca' Foscari University (Venice). The main goal of this course is to broaden knowledge, improve comprehension in the discipline of instructional design, foster skills to organise curricula and programmes for language instruction as well as design learning materials. Concerning the assessment, throughout the semester there are seven online lessons in which students have the opportunity to carry out some activities and gain 0.3 points for each activity done to add to the final exam. At the end of the semester, there is a final exam divided into two components: a task-based project paper and a written exam. In both components, students have to get the pass mark (18 out of 30).

3.2. Research questions, hypotheses and aims

The research questions (RQs) underpinning the present study emerged from examining the literature, attending webinars and talks, discussing with colleagues, and listening to podcasts concerning ChatGPT in the educational realm. As previously mentioned, there seem to be several research gaps in the usage of ChatGPT in the field of education (Bitzenbauer, 2023; Rudolph et al., 2023; Su et al., 2023). The RQs and the related hypotheses (Hs) are as follows:

RQ1: *To what extent is the use of ChatGPT useful for student teachers in order to design classroom activities for language learners, in terms of:*

- *Planning time?*
- *Quality of the final products?*
- *Development of their critical thinking?*

H1: The usage of ChatGPT in designing classroom activities for language learners makes student teachers spend less time in comparison to creating activities without employing ChatGPT. The employment of ChatGPT positively impacts the quality of the final products realised by student teachers. The use of ChatGPT can lead to higher critical thinking capacities of student teachers.

RQ2: *What are the strategies used by student teachers when interacting with ChatGPT to create a task-based language learning activity?*

RQ3: *How can university professors assess the activities created by their students through the use of ChatGPT?*

RQ4: *To what extent is the use of ChatGPT useful for student teacher educators in order to assess the activities created by their students in terms of planning time?*

H4: The usage of ChatGPT in assessing activities of students helps university professors spend less time in comparison to assessing activities without employing ChatGPT.

In light of the above questions and hypotheses, the research aims are fourfold. The first purpose is to verify whether ChatGPT would be useful (in terms of planning time and quality) for student teachers to create tasks for their language learners, and for professors to assess the tasks realised by the student teachers. The second mission is to analyse whether the usage of ChatGPT affects the critical thinking skills of student teachers. The third objective is to examine how student teachers interact with ChatGPT. The fourth goal is to explore the ways university professors can assess the students' work with the help of ChatGPT.

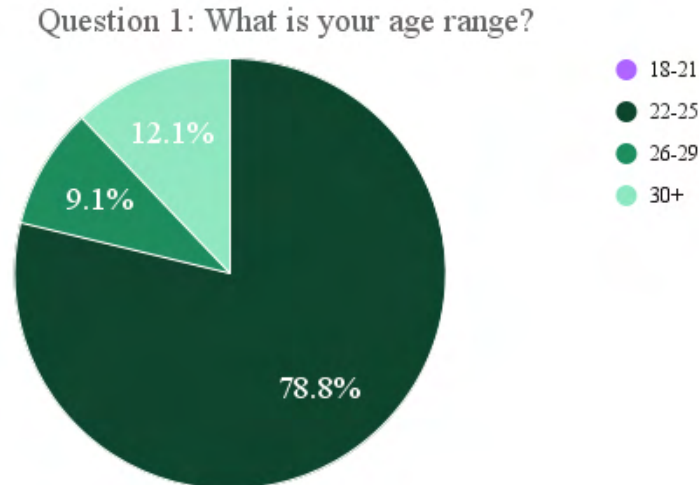
3.3. The participants

The participants of this exploratory study were 39 student teachers attending the Master's degree course in *Instructional Design for Language Education* at Ca' Foscari University (Venice). Participation in this study was voluntary.

The student teachers belonged to the class of the supervisor. The author decided to carry out the experiment with the class of her supervisor as the participants were studying to become language experts or language teachers. From their perspective of both teacher and student, the participants could give deeper responses regarding ChatGPT applied to the field of language education. Due to this key characteristic, and the criterion for easy accessibility, the sample chosen could be considered as an *opportunity or convenience, and purposive sampling* (Dörnyei & Csizér, 2012). Some academics (Cohen et al., 2007) have included this type of sample in the non-probability category and is frequently used in action research, case study research, or ethnographic research.

Figure 10

Age of the participants of the student survey



39 student teachers decided to participate in the online task, of which 33 also completed the online survey after online task completion. According to the student survey, the age range of participants went from 22 to 30 and up years old (Figure 10). As can be seen in the pie chart (Figure 10), 78.8% of participants were between 22 and 25 years old, 9.1% were from 26 to 29 years old, and 12.1% were older than 30 years old.

Figure 11

Nationality of the participants of the student survey

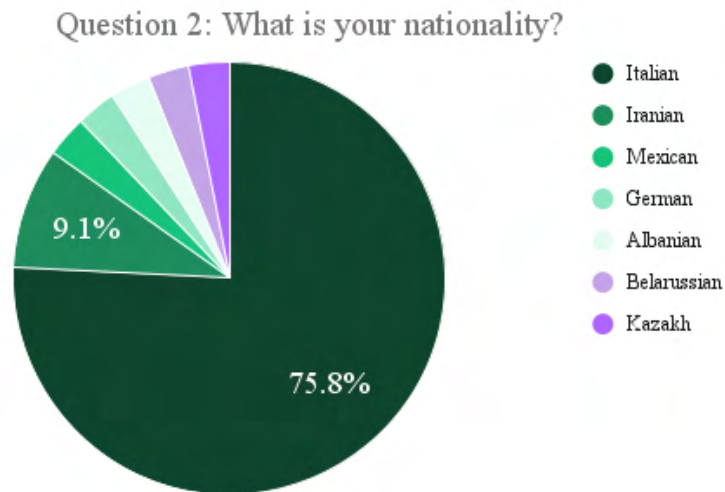
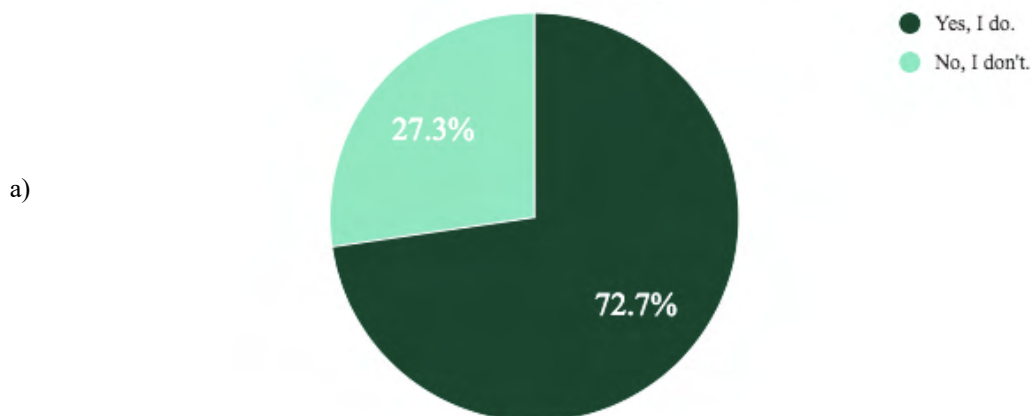


Figure 11 shows the percentage of the student teachers' nationalities that took part in the research project. The majority of the participants were Italian (75.8%). However, there were also other nationalities: Iranian (9.1%), Mexican (3%), German (3%), Albanian (3%), Belarussian (3%), and Kazakh (3%).

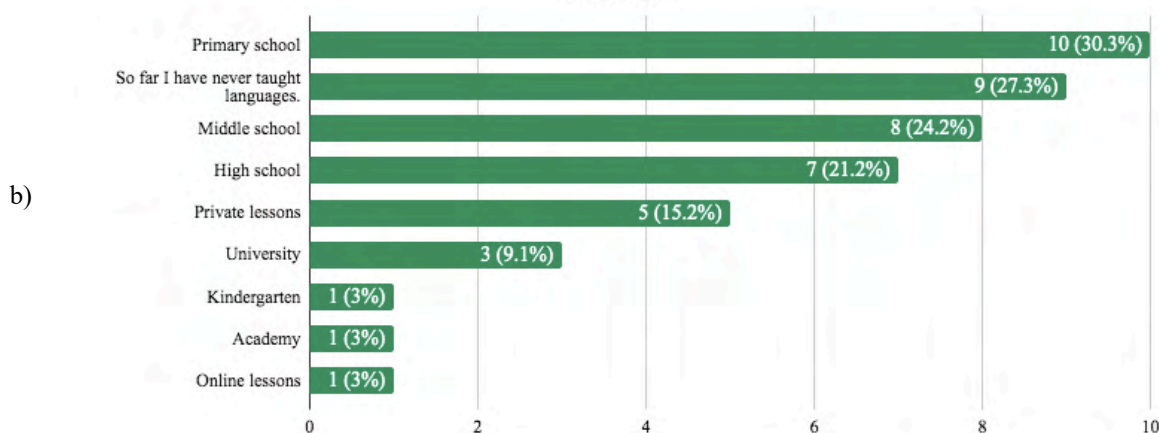
Figure 12

Teaching experience of the participants of the student survey

Question 3: Do you have any work experience in language teaching?



Question 4: In which context are you currently teaching/ did you teach a foreign/second language?



As illustrated in Figure 12, 72.7% of student teachers had work experience in language teaching, especially in primary, middle, and high school. Only 27.3% of participants had no experience in language teaching.

Of this sample, two student teachers belonging to the same group offered to be interviewed online. They were both Italian with little teaching experience in language education.

The course professor was interviewed for the purpose of comprehending in depth the assessment process of the activities created by the student teachers, the structure of feedback, as well as the uses of ChatGPT in the assessment. The professor holds two Master's degrees and a Ph.D. in Language Sciences. Currently, the professor is a researcher in *Educational Linguistics* at Ca' Foscari University and teaches Bachelor's and Master's degree courses. Even in this case, the sample may be considered as an *opportunity or convenience, and purposive sampling* as the professor could have access to the final TBLL activities of the student teachers.

The author was also involved in the research project in relation to assessment. The author holds a Bachelor's degree in *Linguistic Mediation for Tourism Management and International Trade* at CIELS university (Padua) as well as a Master's degree in *Teacher Development for Content and Language Integrated Learning and Foreign Language Education (CLIL & FLE)* at Universidad Autònoma de Barcelona. Currently, she is attending a Master's degree in *Language Sciences* with a specialisation in *Educational Linguistics* at Ca' Foscari University (Venice). Regarding her teaching experience, the author worked as a CLIL teacher intern in a middle school in Barcelona, an English teacher in a language academy in Pamplona,

a Spanish teacher in a kindergarten near Venice, as well as an English and Italian teacher intern in a high school in Venice.

3.4. Data collection instruments

Several research tools were employed to yield information about how student teachers interact with ChatGPT, create TBLL activities, and how professors can assess the activities with the help of ChatGPT. Quantitative data included a student teacher online survey (3.5.1.), while qualitative information encompassed a student teacher online survey (3.5.1.), student teacher online interviews (3.5.2.), student teachers' interactions with ChatGPT (3.5.3.), final TBLL activities realised by student teachers (3.5.4.), professor online interview (3.5.5.), ChatGPT's feedback (3.5.6.), author's feedback (3.5.7.) as well as prompt engineering (3.5.8.). Below the research instruments are explained.

3.4.1. Student teacher online questionnaire

An online survey through *Google Forms* was elaborated. The author chose to administer an online questionnaire as the instrument represents a valid and cost-effective way of gathering large amounts of information during a short period of time (Dörnyei & Csizér, 2012). The student teachers had the opportunity to respond to the questions either in English or Italian. In the end, 33 out of 39 student teachers fulfilled the online survey.

The online questionnaire consisted of 19 English mandatory items with yes/no, multiple choice, as well as open-ended questions. The queries were divided into four main sections:

- Socio-demographic information.
 1. What is your age range?
 2. What is your nationality?
 3. Do you have any work experience in language teaching?
 4. In which context are you currently teaching/ did you teach a foreign/second language?
- Previous knowledge of ChatGPT.
 5. Did you know ChatGPT before this online activity?
 6. Did you use ChatGPT before this online activity?
 7. If yes, in which ways did you use ChatGPT?
- The online activity and ChatGPT.

8. How did you work on this online activity?
 9. How did you work in group?
 10. Do you consider ChatGPT a useful technological tool for your work as a student teacher to design classroom activities for language learners?
 11. Please explain the reasons why.
 12. Do you think that you saved time by using ChatGPT for the activity?
 13. In what ways did you save time by using ChatGPT?
 14. Regarding the activity performed, do you think that you created an activity with better, worse or the same quality as if you had not had access to ChatGPT?
 15. Did you encounter any difficulties or limitations by using ChatGPT? If yes, which ones?
- Further research.
 16. If you consider ChatGPT useful, in what other ways would you use it in language teaching and learning? Apart from the ways you used ChatGPT for this online activity.
 17. Would you like to know more about ChatGPT and Artificial Intelligence in the field of language education?
 18. If you have any teaching experience, do you think that ChatGPT would be a useful tool to help teachers tailor different lessons or activities based on students' needs?
 19. Please explain the reasons why.

3.4.2. Student teacher online interviews

The online interviews were carried out to explore more in-depth the strategies student teachers used to interact with ChatGPT, and the usefulness of ChatGPT for creating a TBLL activity. As participation in the interviews was voluntary, only two students gave their availability to be interviewed. The interviews were conducted through *Google Meet*, carried out completely in Italian, and lasted approximately one hour each. Moreover, they followed a semi-structure consisting of about 12 open-ended questions. However, when the author recognised the necessity for further explications or examples, additional questions were formulated. The author audio-recorded both interviews. During the interview, the author also took some notes of the answers, which seemed to be significant for the queries asked. The author asked the following questions:

- Introductory queries.
 1. Did you like the TBLL activity carried out with the help of ChatGPT? Why?
 2. Did you continue to use this AI chatbot on your own, such as to study autonomously, design materials, ask for difficult concepts, etc.?
- Conversations with ChatGPT.
 3. Regarding the interactions with ChatGPT, it has been noticed that the majority of your prompts concerned adding or asking something, but you rarely asked for modifications or reformulations of the previous outputs given by ChatGPT. Could you explain to me the reasons why?
 4. How reliable do you think ChatGPT is on a scale of 1 to 10? Why?
 5. Concerning the task you uploaded on the Moodle platform, it has been observed that you as a group relied heavily on ChatGPT. Could you tell me why?
 6. When you used ChatGPT, what type of input did you give to the AI chatbot? In your opinion, were you critical of the outputs given by ChatGPT?
 7. How did you discuss in group via Zoom regarding the outputs given by ChatGPT?
 8. Before performing the online task in group, did you have to study the task-based language learning structure to understand whether what ChatGPT wrote was right or wrong?
 9. Why do you think you created a better task with the help of ChatGPT?
 10. Why did you save time with ChatGPT?
- Feedback.
 11. What do you think about the feedback given on Moodle?
- Concluding questions.
 12. Is there anything else you would like to add?

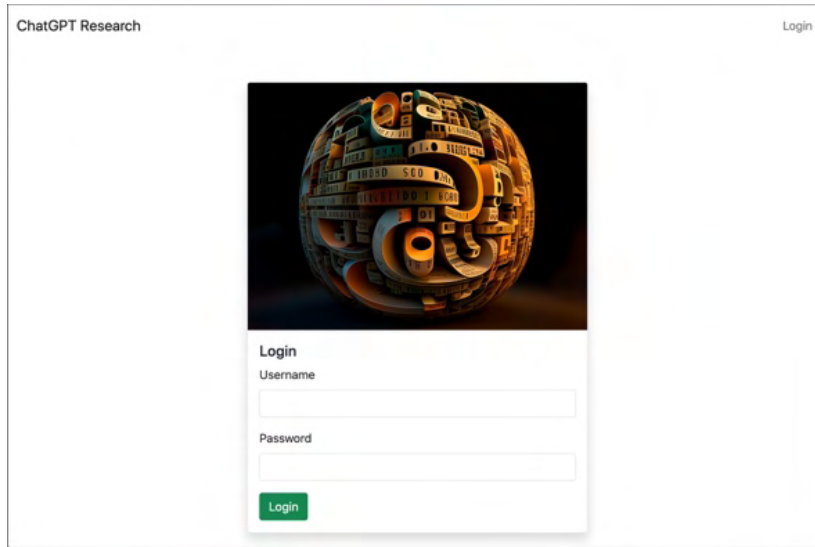
3.4.3. Student teachers' interactions with ChatGPT

In order to carry out the online task, the student teachers had to interact with ChatGPT. A custom research webpage was set up to collect all the conversations between the student teachers and ChatGPT (Figure 13a). In this way, the author could directly analyse the whole interactions produced by the student teachers with ChatGPT, and reduce any bias arising from student teachers' reports.

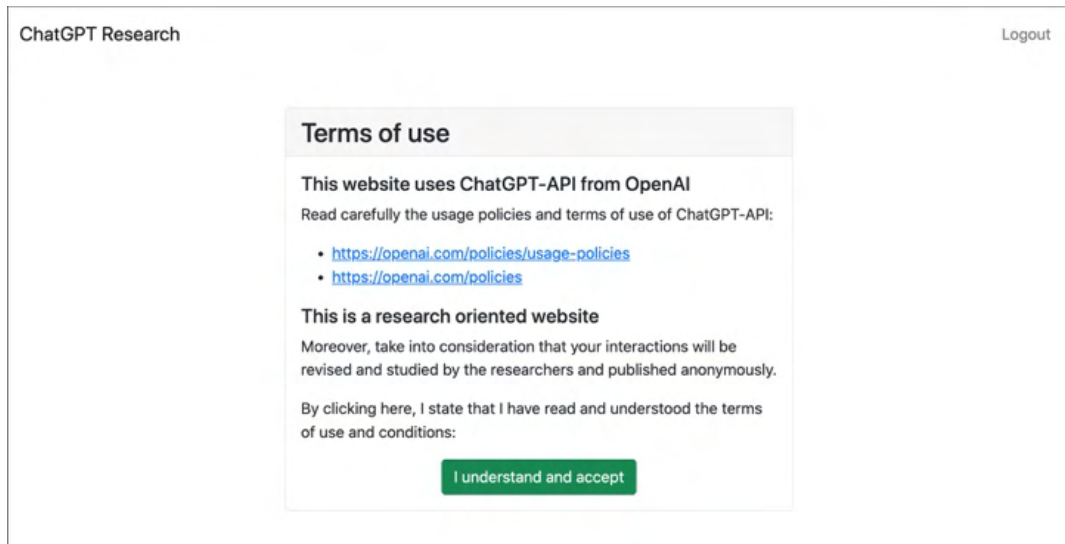
Figure 13

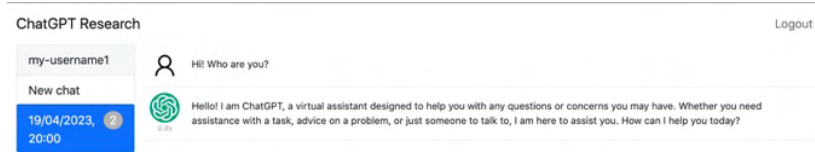
Screenshots of the webpage ChatGPT Research

a)



b)





c)



The web used ChatGPT-API and had an interface similar to ChatGPT (Figure 13c). This web was also developed because the experiment had to be conducted on 18th April 2023, and on 1st April 2023, OpenAI deactivated ChatGPT for users in Italy due to the Italian Garante's¹ concerns about privacy issues. The website was restricted to only those students participating in this study. The subjects who were all adults were invited to accept the terms of use of ChatGPT-API as well as a disclaimer notifying them that their conversations would be analysed by the author and published anonymously (Figure 13b). The students were also informed that they had to use this webpage only for this specific online activity, and they did not have to write any sensitive data. For this research thesis, the author decided to refer to the web as *ChatGPT*. However, the ChatGPT-API parameters may not be equal to the ones of the “standard” ChatGPT.

3.4.4. Student teachers' final TBLL activities

For this online assignment, the student teachers had to design a TBLL activity based on the framework by Willis (1996), which is structured as follows: pre-task, task cycle, and language focus. Student teachers' activities were examined and compared with the chat interactions between the subjects and ChatGPT. The aim was to understand how student teachers relied on the outputs of ChatGPT and used their own critical thinking abilities.

¹For more information: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832> (Garante per la Protezione dei Dati Personali, 2023)

3.4.5. Professor online interview

The author also decided to carry out an online interview with the professor of the course *Instructional Design for Language Education* for several reasons: understanding the criteria to consider, as well as the instruments to use when conducting an assessment of TBLL activities; perceiving what type of assessment would be the best to be provided to the student teachers; comprehending the role of ChatGPT in the assessment practice, and elaboration of feedback; investigating deeper over the quality of the TBLL activities performed by the student teachers. The author sent an email to the course professor in order to organise the online interview. The online structured interview was done through the *Zoom* platform, video-recorded, and conducted in Italian. It lasted about one hour. The author elaborated about 13 questions grouped into four main topics, which were sent to the professor before the online interview:

- Assessment.
 1. What aspects do you usually consider when conducting an assessment of such TBLL activities (for instance, creativity, feasibility of time, etc.)?
 2. What instruments do you usually use to assess such TBLL activities (for instance, checklists, rubrics, etc.)?
 3. If you had an infinite amount of time available, how would you assess the students' activities to help learners improve their mistakes and reinforce their abilities?
 4. In your opinion, what would be the best type of assessment for your students? Why?
- ChatGPT.
 5. Now that you know about the existence of ChatGPT, would you change the assessment mode or type of task?
 6. How would you use ChatGPT to assess the TBLL activities carried out by the students?
- Feedback.
 7. Once you read all the activities, you elaborate the final general feedback to post on Moodle. What aspects do you consider for writing that feedback? Do you usually follow a structure?
 8. What do you think about the feedback given by ChatGPT? Are they high or poor quality?
 9. How would you have elaborated the feedback by using ChatGPT?

10. How long would it have taken you to elaborate feedback similar to the ones written by ChatGPT?

- Task.

11. What do you think of the activities carried out by the students?

12. Did you see a positive influence of ChatGPT on students' activities? Did you notice anything different from last years (for instance, improvement in creativity)?

13. From the exams you administered so far, did you notice that students learnt how to create an appropriate TBLL?

3.4.6. ChatGPT's feedback

In order to respond to RQ4, the author made ChatGPT elaborate feedback on the TBLL activities by giving the following prompts (on the recommendation of the course professor):

- *Could you please assess the following didactic activity in terms of:*
 - 1) coherence among objectives, tasks, and target;*
 - 2) feasibility in terms of time;*
 - 3) suitability in terms of learning environments chosen?*
- *Could you please suggest improvements?*

The objective was to comprehend how ChatGPT could construct and structure feedback based on the input given.

3.4.7. Author's feedback

Based on the professor interview, the author wrote a piece of self-made feedback without the usage of ChatGPT. Afterwards, the author-generated feedback was compared with the feedback generated by ChatGPT. The purpose was to verify what feedback could be developed in shorter time.

3.4.8. Author's prompt engineering

According to White et al. (2023) prompt engineering represents a progressively significant skill set required to communicate efficiently with LLMs. Prompts refer to instructions given to the LLMs to guarantee specific output generation. The author of this Master's thesis tried to employ this technique to elicit more detailed outputs from ChatGPT. Specifically, the prompt

engineering was used to elaborate the best possible input aiming at creating a rubric and evaluating a student teachers' TBLL activity.

3.5. The research methodology

A mixed method approach (Dörnyei, 2007) was applied to the exploratory study, which consisted of the collection of both quantitative and qualitative data. According to Creswell and Creswell (2018), choosing a mixed method research may minimise the limitations of both quantitative and qualitative approaches, and provide a more in-depth comprehension of the research questions and problems.

3.6. Activity description

Before starting the experiment, the student teachers were given a video-tutorial prepared by the author with details about the aim of the study and how to participate in the experiment. Moreover, the student teachers were told what ChatGPT is, and how to use its platform in order to conduct the activity.

The activity student teachers had to design was deeply connected to the one of the lessons previously held in the course, which focused on *Language learning beyond the classroom: Designing tasks in the museum*. The student teachers were asked to create an online Task-Based Language Learning (TBLL) activity with the usage of ChatGPT to be performed individually or in group. They were given the following instructions:

- *The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.*
- *Choose a theme among food education, selfies and portraits, and sustainability.*
- *Choose a museum setting where the activity will take place. It can be a museum that you know or that can be visited in person or you can choose one of the Civic Museums of Venice (<https://www.visitmuve.it/it/virtual-tour/>) that you can visit virtually. The museum can be an art, science or natural museum.*
- *The skill to promote is speaking. Please define the learning objectives (for example: the students will be able to orally present the main characteristics of an Italian garden by using the expression «in my opinion...» and the specific vocabulary related to gardens).*

- *Choose an object or other resource (realia, works of art...) and design a task (see Willis' model: pre-task, task cycle, post-task) specifying the duration of the single stages and the materials and tools you would use.*
- *Describe in detail what you would do in the classroom before and after the museum (for example, activities, materials, ICT tools etc.). Specify the activity/exercises you would do, writing down the text of exercises: e.g., if you create a multiple-choice exercise for reading comprehension, specify both the questions and alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (e.g., I agree with you; I couldn't agree with you more; That's so true, etc.).*

Please feel free to interact with the AI chatbot in the most strategic way, so as to create a teaching activity you are satisfied with!

3.7. Data collection procedure

Data were collected from 18th to 30th April 2023. To design the online task, the student teachers were expected first, to interact with ChatGPT to create the TBLL activity, second, to upload the document with the task on Moodle, and third, to complete the online questionnaire. The participants were finally invited for an interview. Once the tasks were concluded, they were inserted in ChatGPT for evaluation in relation to the assessment criteria normally used by the course professor to evaluate this same activity. In particular, the prompts given to the chatbot were suggested by the course professor as follows:

1) Could you please assess the following didactic activity in terms of:

- *coherence among objectives, tasks, and target;*
- *feasibility in terms of time;*
- *suitability in terms of learning environments chosen?*

2) Could you please suggest improvements?

The feedback texts generated by ChatGPT were revised, slightly modified and posted on Moodle by the author. The time spent to generate, revise and modify the feedback on the TBLL activities was measured. Subsequently, the course professor was interviewed to elicit more information about assessment and feedback. Afterwards, a piece of self-made feedback on a TBLL activity was written by the author without the help of ChatGPT. The time spent to write and revise the self-made feedback was measured.

3.7. Ethics

As research in the educational field deals with the lives of people, it is important to take ethics into consideration (Dörnyei, 2007, p. 63-72). In order to avoid possible ethical issues, the author informed the student teachers that participation in the research was voluntary.

Furthermore, the author wrote the following item on the online questionnaire: *Data will be collected anonymously and used exclusively for research and educational purposes in full respect of privacy, as required by Legislative Decree 163/2017, Ex art. 13 D.L. 196/2003 and Ex art. 13 European Regulation 2016/679.* The student teachers had to agree with it in order to fill in the online questionnaire. Regarding the online interviews, the author asked for permission to audio-record the voices of the student teachers, and to video-record the interview with the professor.

Additionally, the student teachers had to accept the terms of use in order to access as well as use the web (Figure 13b). Moreover, the student teachers were told not to write sensitive information on the webpage.

Anonymity was significant in the data collection of the present study. It is fundamental to specify that the identities of people who participated in the online surveys as well as interviews were omitted, and the participants were apprised of the anonymity. In fact, during the data collection processes, any names, materials or information that could have recognised the nature of the participants were not used and included in the present thesis based on two essential principles: confidentiality as well as consent (Green & Thorogood, 2004, p. 51-75).

3.8. Data analysis procedure

Subsequent to data collection, data were grouped so as to answer the research questions. The quantitative results of the online questionnaire were further integrated with qualitative results of open-ended questions as well as online interviews. The open-ended questions included in the online questionnaire were analysed by applying a content analysis (Dörnyei, 2007; Evans, 2017). In particular, the responses to the open-ended queries were first examined to grasp a general sense of the results, were then divided into categories, and the related percentages were calculated. Since only three online interviews were carried out, the author opted to include the relevant answers directly to the rest of the data, but without conducting a content analysis.

To compare the student teachers' conversations with ChatGPT to the student teachers' TBLL activities, the author carried out a quantitative content analysis composed of four

categories: original words, copied words, rephrased words and copied-exercise words. To perform this analysis, a text-comparison tool was used to further support the study. This instrument is named *Similarity Texter* (<https://people.f4.htw-berlin.de/~weberwu/simtexter/app.html>). Specifically, the author copied and pasted every single student teacher's conversation with ChatGPT as well as the related TBLL activity to compare the two texts and find the copied sequences. The comparison parameters used were a minimum match length of five words, ignoring letter case, numbers, punctuation, and replacing umlaut and ligatures. Subsequently, the author underlined in different colours the four categories and counted the number of words in each category. The numbers were transferred to a *Google Sheet* document in order to construct four histograms. Concerning the prompts given by student teachers to ChatGPT, the author conducted a further content analysis (Dörnyei, 2007; Evans, 2017) categorising the different types of inputs given by the student teachers to ChatGPT (the detailed procedure will be described in section 4.3.). Moreover, the author employed a self-study methodology (Hamilton et al., 2009) through prompt engineering to find ways professors may assess the activities designed by their students (the detailed procedure will be described in section 4.4.). Finally, the two time frames and the lengths of the piece of self-made feedback and the ChatGPT-generated feedback were compared.

CHAPTER IV. DATA ANALYSIS

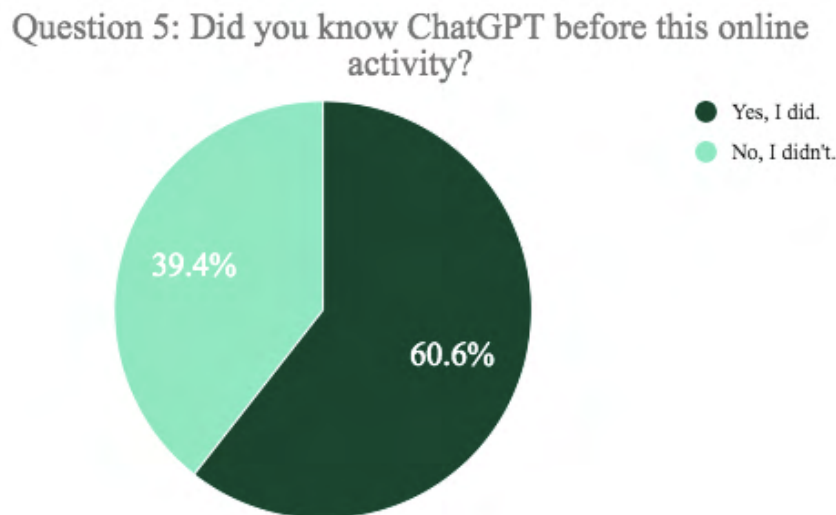
In this section, a detailed data description and analysis will be provided. The results have been grouped according to each research question in order to give a clearer overview to the readers.

4.1. Premises

Before starting the data examination to answer the first research question, it is necessary to make some introductory statements concerning prior knowledge of ChatGPT by the sample.

Figure 14

Prior knowledge of ChatGPT



The pie chart (Figure 14) shows an overview of the percentage of student teachers who knew ChatGPT before the online activity. As can be noticed from the graph, 60.6% of student teachers were informed about the existence of ChatGPT. However, the rest of participants (39.4%) were not aware of the invention of ChatGPT.

Figure 15

Usage of ChatGPT before the online activity

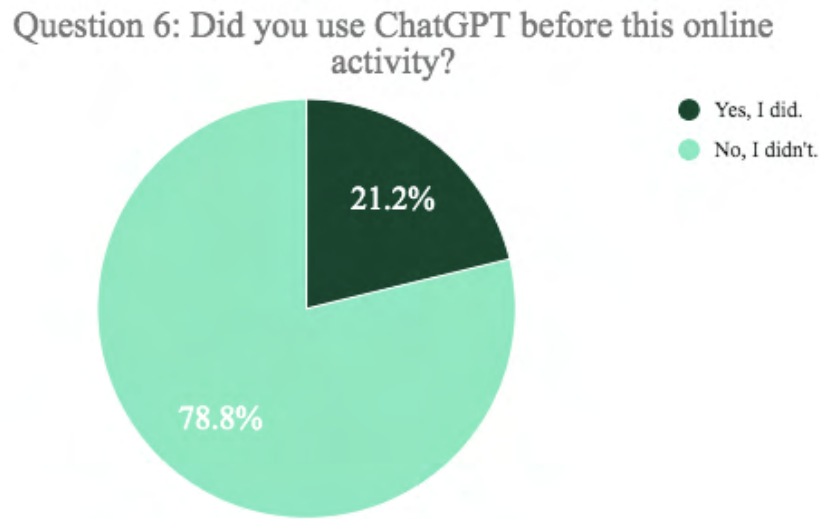


Figure 15 illustrates the percentages of student teachers who used ChatGPT before conducting the online activity. As the pie chart shows, there is a significant difference between the two groups. Surprisingly, 78.8% of subjects reported that they had never utilised ChatGPT before, while 21.2% stated that they had already employed this AI tool.

Table 1

Ways student teachers used ChatGPT before the online activity

7. If yes, in which ways did you use ChatGPT? (If you replied No just write --)	
Asking general questions	1
Self-studying	1
Research	1
Asking for information	1
Translation	1
Completion of Preply account	1
Work	1
No usage	26

Table 1 provides a summary of the ways student teachers used ChatGPT before the online activity. As previously mentioned, most participants (26) did not use ChatGPT. However, other

people tried to utilise the AI tool in different ways, for instance, some student teachers asked ChatGPT for general questions, translations, research, etc.

Figure 16
Organisation of the work

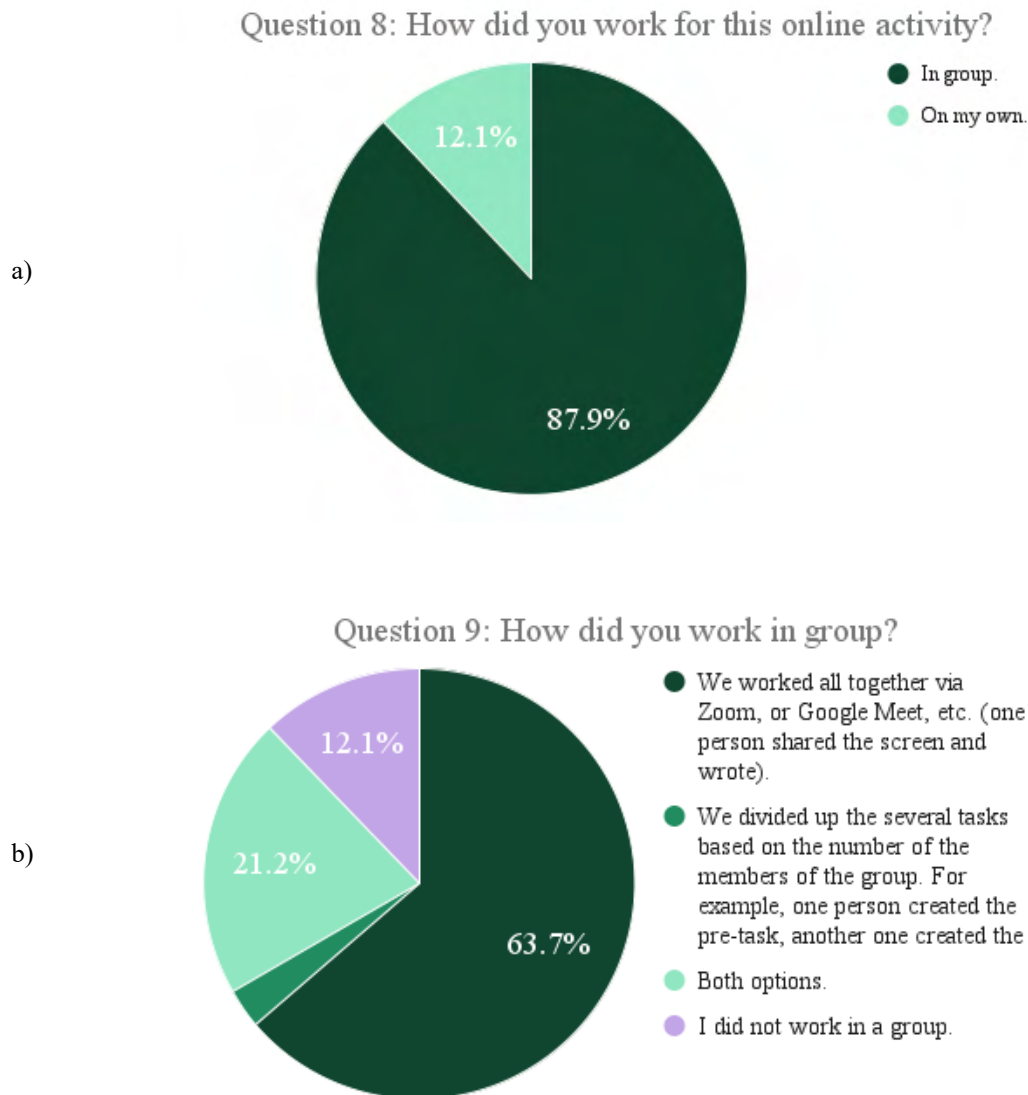


Figure 16a shows that 87.9% of student teachers decided to collaborate in groups instead of working on their own (12.1%). As can be seen from Figure 16b, most student teachers (63.6%) worked all together through an online platform (*Zoom, Google Meet, etc.*), others preferred to divide up the several tasks based on the number of group members (3%), whereas some subjects (21.2%) decided to work with both options: they worked together via an online platform as well as divided up the several tasks. Only 12.1% chose to work individually. In conclusion, 14 TBLL

activities were gathered: 9 assignments were created by teamwork, and 5 activities were designed by single subjects.

4.2. Data analysis of RQ1

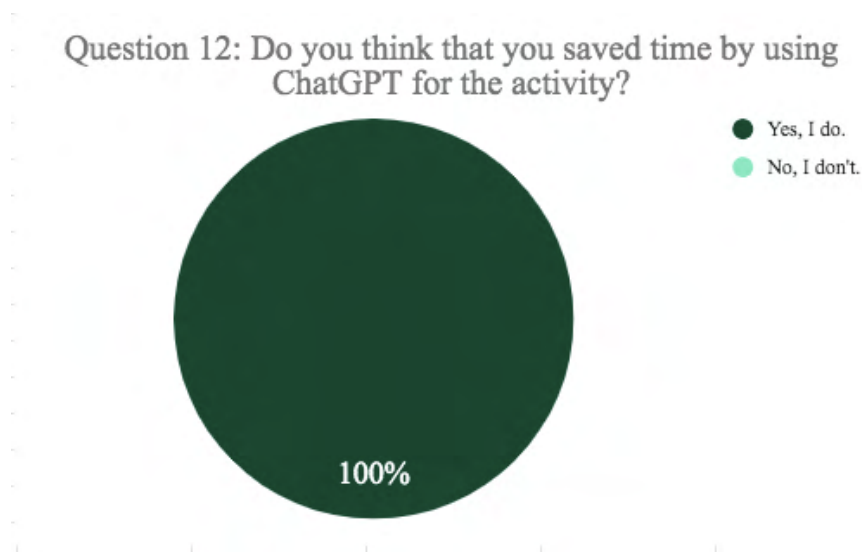
In order to answer the first research question, the author analysed the data collected from student teacher online questionnaires, student and professor online interviews, students' interactions with ChatGPT, and student final TBLL activities. Concerning the online questionnaires and interviews, the queries have been selected, and only the most significant questions have been reported. The first research question is as follows: *To what extent is the use of ChatGPT useful for student teachers in order to design classroom activities for language learners, in terms of:*

- *planning time;*
- *quality of the final products;*
- *development of their critical thinking using ChatGPT?*

Regarding the first point of the research question (in terms of *planning time*), the author of the thesis analysed some questions of the online student teacher survey as well as student teacher online interviews.

Figure 17a

Answer to question number 12 belonging to the student teacher survey



In Figure 17a, it can be clearly seen that the overall response to the question *Do you think that you saved time by using ChatGPT for the activity?* was very positive. In fact, all student teachers (100%) agreed to the fact that ChatGPT helped the subjects to save time in designing their assignments.

Figure 17b

How student teachers saved time by utilising ChatGPT

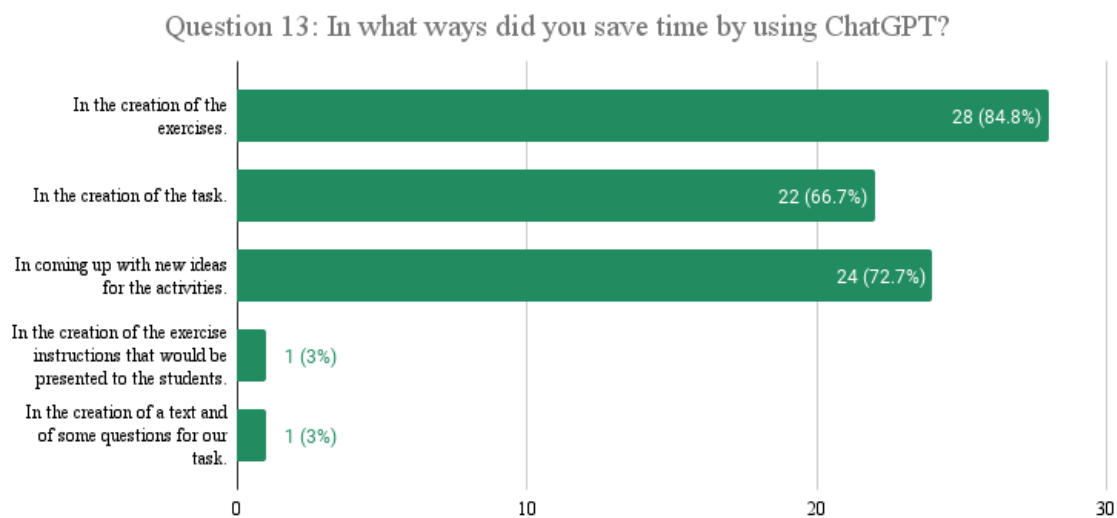


Figure 17b provides information on how the student teachers saved time by using ChatGPT. In this question, the participants were free to give more answers. The majority (84.8%) stated that ChatGPT was helpful in creating exercises, whereas 24 subjects (72.7%) said that the tool was useful to generate new ideas for the activities. Furthermore, 22 participants (66.7%) reported that the AI chatbot aided the subjects in performing the task. In the end, one person (3%) added that ChatGPT facilitated the elaborations of the instructions for the exercises, whereas another teacher-student (3%) commented that the AI tool was worthwhile to generate texts and some questions.

Moreover, in response to the question *Why did you save time with ChatGPT?* a student interviewee reported that ChatGPT represented a very time-saving instrument because it suggested several ideas as initial inputs for the whole activity. The interviewee also added that the AI tool was really helpful in generating the questions for the quiz. In fact, the participant commented: “Writing all the answer options is certainly more time-consuming than receiving a final product that is already ready to implement, and you have just to modify it or make a

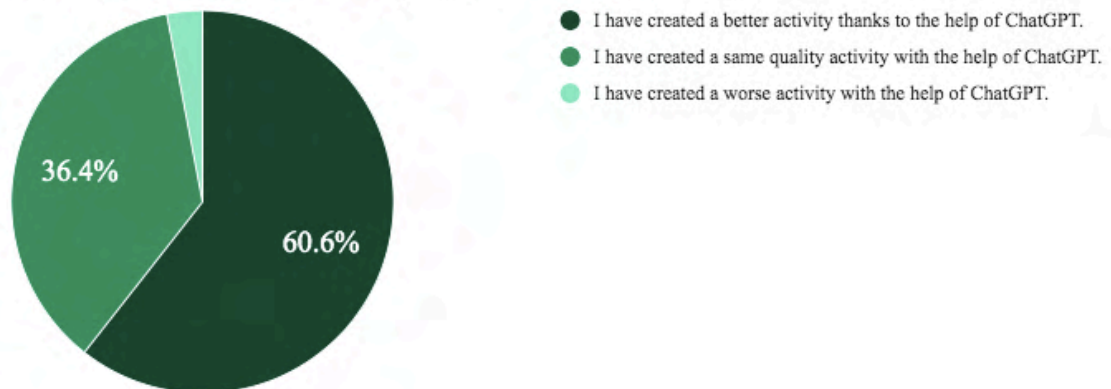
selection process”. Instead, the other interviewee highlighted that her group could save time thanks to ChatGPT to create didactic materials as well as exercises.

Concerning the second objective of the first research question (in terms of the *quality of the final products*), the author focused on the investigation of some responses from the student teacher survey, as well as student teacher and professor interviews.

Figure 18

The quality level of the online assignments with the aid of ChatGPT

Question 14: Regarding the activity performed, do you think that you created an activity with better, worse or the same quality as if you had not had access to ChatGPT?



The pie chart (Figure 18) depicts the perceptions of the student teachers concerning the activity performed with the aid of ChatGPT. The majority (60.6%) of those who answered this question felt that they carried out a better activity thanks to the help of ChatGPT, while 36.4% believed to have designed a same quality activity with the help of ChatGPT. The rest (3%) presumed to have performed a worse activity with ChatGPT.

According to the responses obtained from the student teacher interviews, a participant commented that she might think to have created a better activity with the usage of ChatGPT due to her little experience in the teaching realm and her personal insecurity. She also stated that having ChatGPT available for this online activity gave her a feeling of self-confidence in her teaching skills. Whereas the other interviewee reported believing in the competences and resources of her group, but she also added that ChatGPT aided her group in suggesting such creative ideas that maybe they would never have thought of.

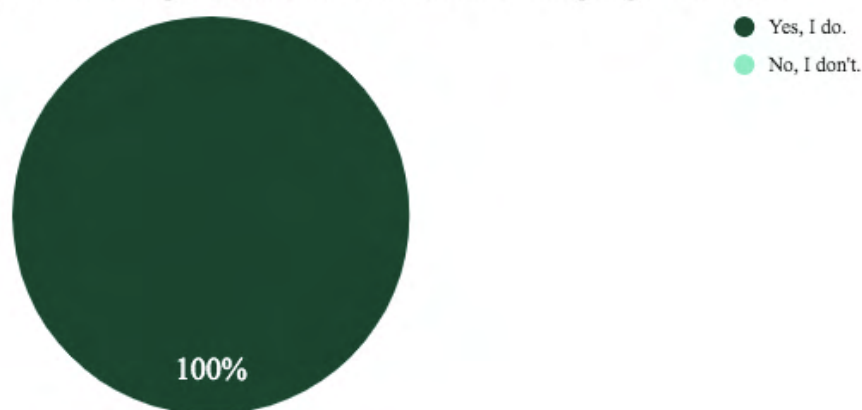
In order to obtain the appropriate results to answer the third part of the first research question (in terms of the *development of their critical thinking of the usage of ChatGPT*), the

author examined several data. Firstly, the author analysed some questions of the online survey and reported the answers in the form of pie charts and tables. Secondly, the author studied and reported some extracts of the online student interviews. Lastly, the author carried out a quantity content analysis between the final TBLL activities designed by the student teachers and their conversations with ChatGPT. The aim was to explore to what extent the student teachers based their final assignments on the interaction with ChatGPT. In order to do that, the author employed a text-comparison tool as further support for her study. This instrument is named *Similarity Texter* (<https://people.f4.htw-berlin.de/~weberwu/simtexter/app.html>).

Figure 19

The usefulness of ChatGPT in designing classroom activities

Question 10: Do you consider ChatGPT a useful technological tool for your work as a student teacher to design classroom activities for language learners?



From the pie chart (Figure 19) it can be noticed that all student teachers (100%) considered ChatGPT as a useful AI instrument to create several activities for language learners.

Table 2

Reasons why student teachers think that ChatGPT is useful for designing lessons

Category	Number of participants	Percentage (%)
Time-saving	8	24.2
Accurate	2	6.1
Creation of exercises, tasks, lesson plans, texts, tests	15	45.5
Brainstorming of ideas	16	48.5

Collection of material	2	6.1
Self-study	1	3.0
Design of creative or interesting tasks or activities instead of traditional ones	5	15.2
Useful for those people who do not have experience in teaching	1	3.0

In response to the question *Please explain the reasons why*, a range of answers was elicited. Based on the data included in Table 2, it can be noticed that the majority considered ChatGPT helpful in suggesting various ideas for class activities (48.5%) and creating texts, exercises, tasks, etc. (45.5%). In respect of this, a participant commented that ChatGPT is “a sort of *creative assistant* that helps to develop ideas”. Some subjects highlighted the fact that they could save time thanks to the help of ChatGPT (24.2%), whereas others emphasised the creativity of the AI instrument (15.2%), and its accurateness (6.1%): “The fact that surprised me the most was that ChatGPT answers to your exact questions: in this sense, if you have doubts in designing an activity, you can ask it for help and it will tell you exactly what you need to do in a clear way”. Furthermore, a person (3%) reported that “I think it is a useful tool for those who have no experience or little experience in teaching”. However, others responded to this question including: “ChatGPT might not be an adequate tool to rely on” and “Teachers should evaluate the outputs given by ChatGPT carefully to guarantee that the ideas, exercises, etc. are suitable for the target as well as learning goals.”

Table 3

Encounter with difficulties or limitations by using ChatGPT

Category	Number of participants	Percentage (%)
Affirmative answers (Yes, I did)	13	38.7
In the beginning phase	2	6.5
Formulation of prompts	6	19.4
Unsatisfying outputs	8	25.8
Negative answers (No, I didn't)	19	61.3
<i>Invalid answers</i>	2	6.5

Note. The category *Invalid answers* includes student teachers' answers that were not understood by the author.

Table 3 illustrates the limitations and difficulties encountered by the participants of the study who employed ChatGPT. 61.3% of student teachers did not have any issues or find any constraints on the usage of ChatGPT, while the rest (41.9%) encountered some limits or impediments. For instance, 25.8% of subjects commented that were not satisfied with the output provided by ChatGPT: “The chat does not always offer a sufficiently accurate answer. Sometimes the result requires revision in accordance with the purposes for which it will be used after”; “We, as a group, looked for some definitions for an exercise in which students had to match words to their definitions, and ChatGPT sometimes was including the word of the definition in the definition itself, so we had to change part of the sentence”. Others, instead, had difficulty in the formulation of the prompt (19.4%), especially in the beginning phase: “The only difficulty that our group had initially was to get used to the tool and to be able to frame the questions in the most complete way to obtain meaningful answers and suggestions”; “At the beginning, it was a little bit difficult to decide what questions to ask and the way in which they should be asked but later it was just intuitive and like you were chatting with a real person”.

Table 4

Quantitative content analysis between student teachers’ interactions with ChatGPT and their related final TBLL assignments

Number of the activity	Original words	Copied words	Copied exercise words	Rephrased words	Total words	% Original words	% Copied words	% Copied exercise words	% Rephrased words
1	865	1745	0	0	2610	33	67	0	0
2	192	240	424	37	893	22	27	47	4
3	540	317	260	1224	2341	23	14	11	52
4	423	79	693	44	1239	34	6	56	4
5	300	652	276	220	1448	21	45	19	15
6	205	0	400	169	774	26	0	52	22
7	271	100	448	389	1208	22	8	37	32
8	1	415	0	29	445	0	93	0	7
9	291	1292	0	0	1583	18	82	0	0
10	98	118	0	117	333	29	35	0	35
11	32	133	0	100	265	12	50	0	38
12	929	0	266	216	1411	66	0	19	15
13	98	574	0	166	838	12	68	0	20

14	537	29	406	37	1009	53	3	40	4
----	-----	----	-----	----	------	----	---	----	---

Note. "Original words" represent the words the student teachers wrote themselves. "Copied words" mean the words copied from the output of ChatGPT, excluding comprehension texts and exercises (such as crosswords, quiz and questions). "Copied exercise words" include the words copied from the outputs of ChatGPT, which are texts and exercises. "Rephrased words" represent the sentences wrote by ChatGPT but rewrote by student teachers with their own words.

Table 4 presents a detailed breakdown of the quantitative content analysis between the student teachers' interactions with ChatGPT and their related final TBLL proposals. Data include the number of original, copied, rephrased words, copied exercise and their corresponding percentage found in the 14 final TBLL activities in comparison with the conversations with ChatGPT.

Figure 20

Example of original, copied, and rephrased words of two TBLL activities

- Task
- a)
1. In the museum, in order not to disturb the other visitors, Divide students into pairs or small groups. (3 minutes)
 2. Give each group a brief introduction to the historical context of Venice in the 16th century, including its political and cultural role in Europe at that time. (5 minutes for each group)
 3. Provide students with research resources, such as books, articles, and online databases, to help them investigate the specific historical events and cultural trends that influenced Tintoretto's painting. (2 minutes)
 4. Have each group closely examine the painting and take notes on their assigned theme (portraits). Sharing ideas with their peers. The teacher walking around listening to them and giving them some feedback. (15 minutes)
- b)
3. ACTIVITY
Start by introducing the students to Gustav Klimt's life and work. You can use a short presentation, a video or a poster with information about his career, his style and the most important works he created.

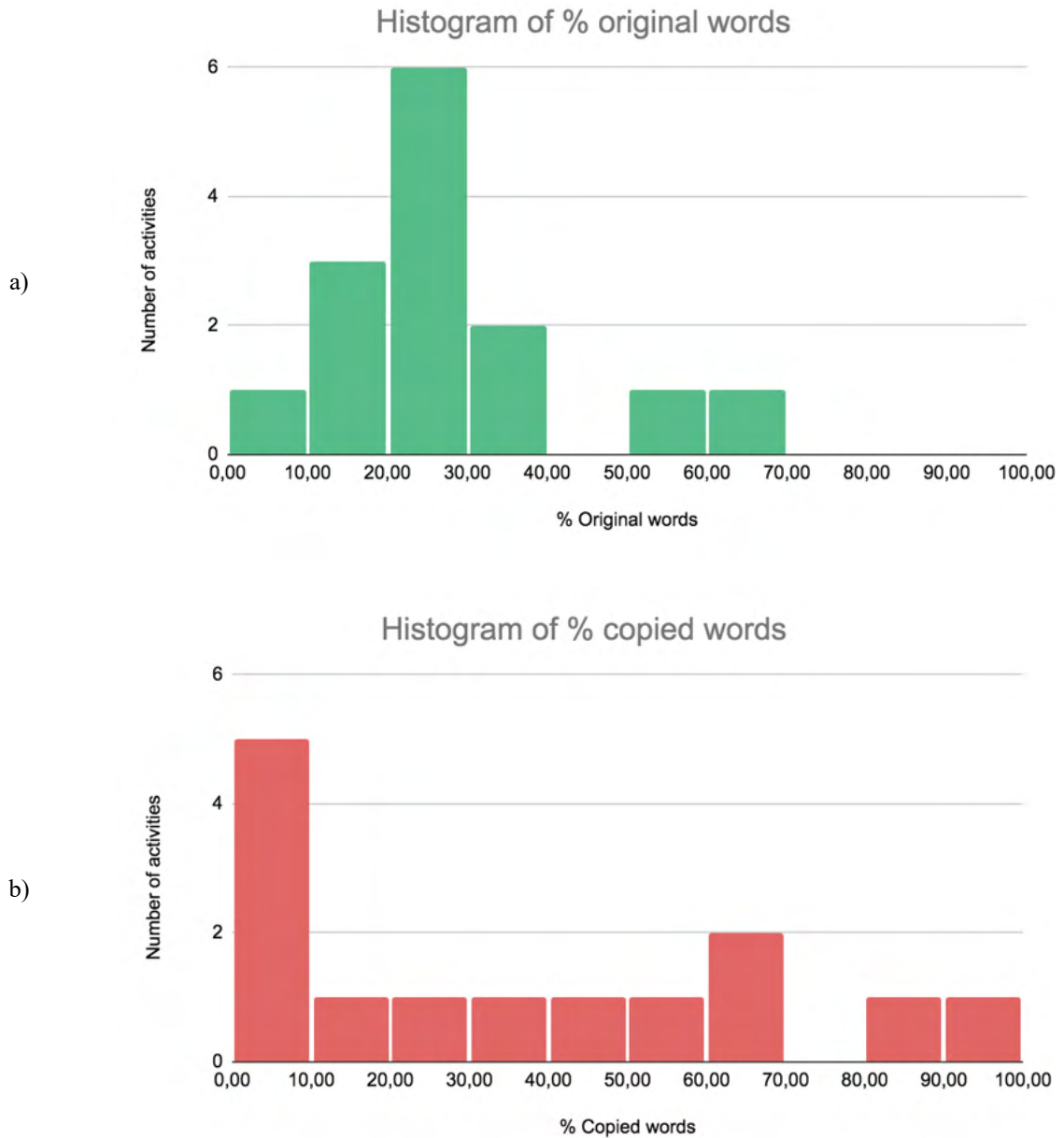
Note. The words highlighted in blue mean rephrased words; the words highlighted in green mean original words written by the student teachers with their own words; the words highlighted in red mean copied words from ChatGPT.

Figure 20 shows two examples of TBLL activities (20a; 20b). It can be noticed that in point 3 of Figures 20a and 20b ChatGPT gave the student teachers a list of several options, but at the

end, the student teachers did not pick the single option they needed for the task. As can be seen, they just copied what ChatGPT suggested.

Figure 21

Histograms including the percentage of original words (a), copied words (b), copied exercise (c), and rephrased words (d) in the 14 final TBLL proposals



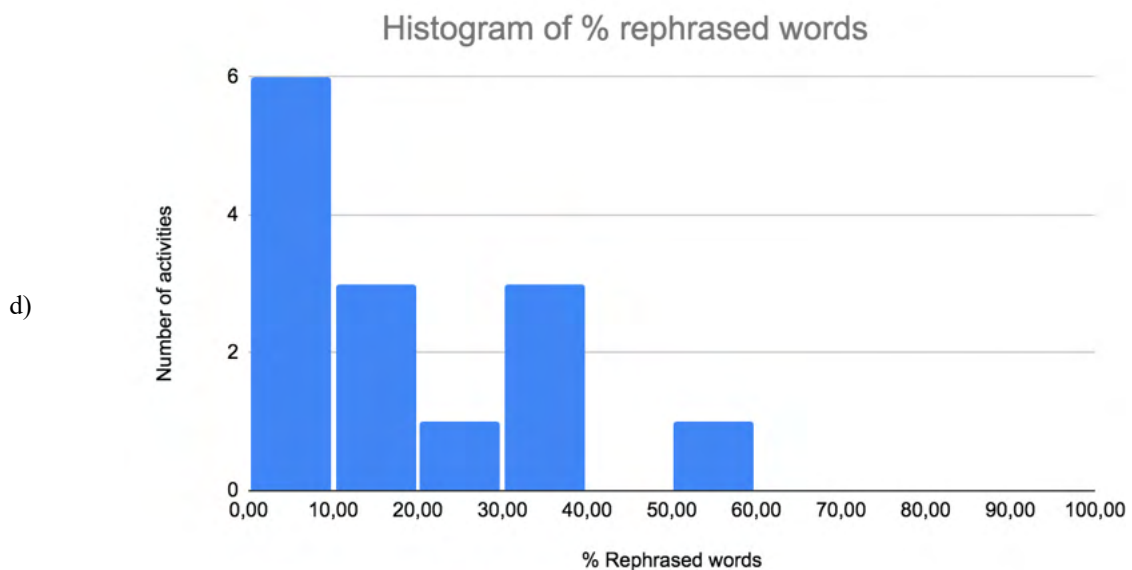
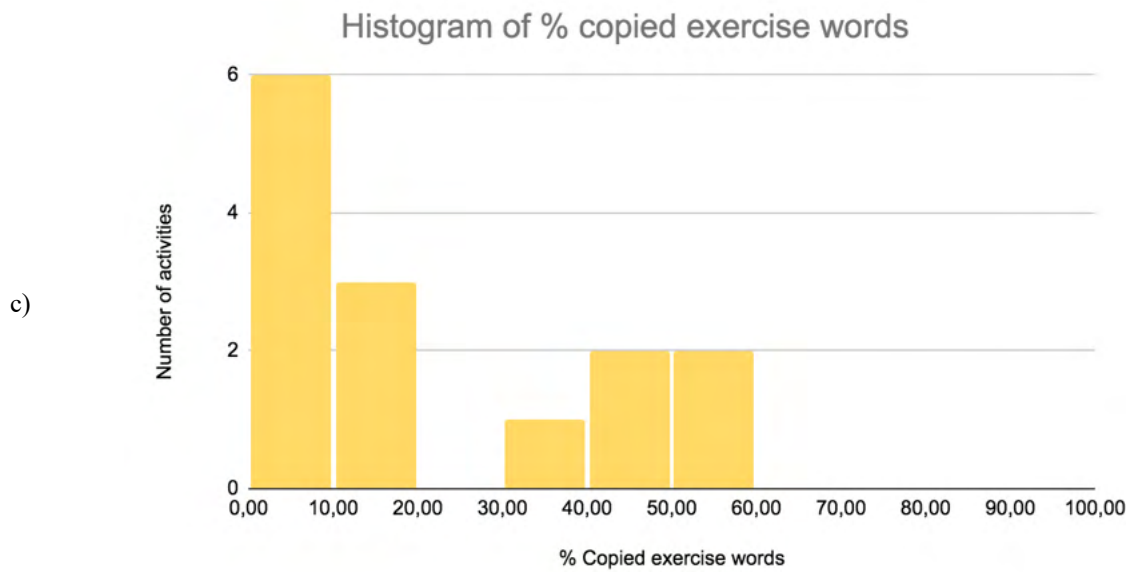


Figure 21 illustrates four histograms showing the percentage of original (a), copied (b), copied exercise (c), and rephrased words (d) in a more illustrative way. The first graph (20a) demonstrates that most final activities created by the student teachers present less than 40% of original words.

The second plot (20b) exhibits a considerable degree of variability in the number of copied words. However, more than a third of the final assignments (5) concentrate between 0% and 10% of the copied words. Additionally, two groups of student teachers with some of the highest percentage of copied words (activity numbers 8 and 9, Table 4) misunderstood the

professor's instructions and created the activity by copying and pasting the whole interaction with ChatGPT. In this regard, one of the interviewees commented that they understood to strictly base their TBLL activity on the outputs of ChatGPT. The third histogram (20c) indicates that almost two-thirds (9 TBLL activities) include less than 20% of copied exercise words and the remaining activities contain between 30% and 60% of copied exercise words.

The fourth histogram (20d) shows that most of the activities present from 0% to 40% of rephrased words, with almost two-thirds (9) below 20%. Overall, Table 4 shows that 8 out of 14 groups used less than 35% of copied words with six of them having below 14% of copied words. However, the remaining six activities contain above 45% of copied words with four of them above 67%. It is also crucial to notice that the activities' lengths range from 265 to 2610 words.

Concerning the critical thinking, the author asked the following questions to the two students interwoven:

- *Concerning the task you uploaded on the Moodle platform, it has been observed that you as a group relied heavily on ChatGPT? Could you tell me why?*
- *In your opinion, were you critical of the outputs given by ChatGPT?*
- *How did you discuss the outputs given by ChatGPT in group via Zoom?*
- *Before performing the online task in group, did you have to study the task-based language learning structure to understand whether what ChatGPT wrote was right or wrong?*

Based on the information yielded through the online student interviews, the subjects explained that they met with their group twice. Before beginning to use ChatGPT, the group studied both the characteristics and structure of TBLL activity in order to comprehend whether the outputs provided by ChatGPT were correct. Subsequently, the group discussed what to do and include in their TBLL. Afterwards, the subjects with the other members of the group reflected on each output by ChatGPT providing their personal thoughts.

4.3. Data analysis of RQ2

Data for answering the second research question that is *What are the strategies used by student teachers when interacting with ChatGPT to create a TBLL activity?* were collected from the prompts written by the student teachers to interact with ChatGPT. In order to analyse these data, the author conducted a content analysis (Evans, 2017; Dörnyei, 2007). Basically, the author

categorised the different types of inputs given by the student teachers to ChatGPT in order to comprehend how they interact with the AI chatbot.

Table 5

Classification of the prompts given by the student teachers to ChatGPT

Types of interactions	Users	Percentage (%)
1. General information		
• <i>Brainstorming of museums</i>	3	12.5
• <i>Request for further information</i>	4	16.7
2. Request for creating the entire task		
• <i>Professor's instructions used as prompts</i>	2	8.3
• <i>Development of the whole TBLL but mentioning or at least explaining Willis' framework to ChatGPT.</i>	5	20.8
3. Organisation of the task		
• <i>Description of learning objectives</i>	6	25.0
• <i>Durations of the activities</i>	5	20.8
4. Creation of the content material		
• <i>Brainstorming of several activities</i>	13	54.2
• <i>Formulation of a single activity</i>	10	41.7
• <i>Creation of worksheets</i>	2	8.3
• <i>Elaboration of a text for a reading comprehension</i>	8	33.3
• <i>Request for specific exercises</i>	14	58.3
• <i>Request for website sources</i>	2	8.3
• <i>Creation of visual representations</i>	1	4.2
• <i>Providing linguistic support</i>	4	16.7
• <i>Enquiry about specific terminology as well as definitions</i>	5	20.8
5. Request for suggestions or some advice	4	16.7
6. Modifications of the outputs	11	45.8

As can be seen from Table 5, the student teachers employed several strategies when interacting with ChatGPT. Seventeen categories were identified and gathered into six macro-groups. Each category includes the number of student teachers who used that type of chat interaction. Afterwards, the percentage of users was calculated based on the total of the pupils, that was 24 student teachers used ChatGPT to design the TBLL. Below the categories are explained, and some corresponding examples of student teachers' conversations with ChatGPT are provided:

- Brainstorming of museums. The student teachers asked ChatGPT for some names of famous museums such as:

- 1) *Do you know some museums related to sustainability?*
- 2) *Are there any museums where teenagers can share the idea of sustainability?*
- 3) *I want to take my students to an art museum or an exhibition in Veneto. What would you recommend?*

According to the data, three student teachers (12.5%) used this type of interaction.

- Request for further information. The student teachers wanted to ask for information such as:

- 1) *Could you name some famous painters who painted self-portraits?*
- 2) *Could you please provide me with a list of all the self-portraits and portraits of others that are on display in the above-mentioned museums? Before taking my students to a museum among those you suggested I want to be sure to find there what I am looking for, that is self-portraits and portraits by others.*
- 3) *Could you please give me information about Mocenico Palace museum?*

Table 5 shows that four student teachers (16.7%) decided to utilise this strategy.

- Professor's instructions used as prompts. The student teachers copied the instructions of the professor published on Moodle and pasted them to generate an input.

- 1) *... The skill to promote is speaking. Please define the learning objectives (for example: the students will be able to orally present the main characteristics of an Italian garden by using the expression «in my opinion...» and the specific vocabulary related to gardens). 5. Choose an object or other resource (realia, works of art...) and design a task (see Willis' model: pre-task, task cycle, post-task) specifying the duration of the single stages and the materials and tools you would use. 6. Describe in detail what you would do in the classroom before and after the museum (for example, activities, materials, ICT tools etc.). Specify the activity/exercises you would do, writing down the text of the exercises: e.g., if you create a multiple-choice exercise for reading comprehension, specify both the questions and the alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (- I agree with you. - I couldn't agree with you more. - That's so true. Etc.)*

- 2) *Specify the activity/exercises you would do, writing down the text of the exercises: e.g., if you create a multiple-choice exercise for reading comprehension, specify both the questions and the alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (- I agree with you. - I couldn't agree with you more. - That's so true. Etc.)*

Based on the data of Table 5, two student teachers (8.3%) employed this type of prompt.

- Development of the whole TBLL but mentioning or at least explaining Willis' framework. In this category, the student teachers referred to or briefly clarified Willis' framework to conduct the TBLL.

- 1) *Can you please design a task based on the topic we mentioned using Willis' (1996) framework?*
- 2) *The activity has to be made up of 3 stages: a first stage that promotes their motivation, a second stage that involves portraits and selfies and a third stage to consolidate or integrate or expand the learning resulting from the exhibition visit. Please describe in detail every stage.*
- 3) *Act as an expert Instructional Designer. Use the Willis' Model; pre-task, task cycle, post-cycle.*

As indicated in Table 5, five student teachers (20.8%) chose to use this category as a prompt.

- Description of learning objectives. The student teachers asked ChatGPT to write some learning objectives for their TBLL.

- 1) *Can you give us some objectives based on our topic?*
- 2) *I will send you the characteristics of this design and I need you to help me find some learning objectives.*
- 3) *Could you write lesson objectives for speaking, students aged 16-18, level B1, theme selfies and portraits?*

As mentioned in Table 5, six student teachers (25%) used this type of interaction.

- Durations of the activities. The student teachers wanted to know the duration of each activity such as:

- 1) *In your opinion, how can I divide these activities in terms of duration?*
- 2) *How much time does this activity take?*

3) *Could you specify the duration of the three categories?*

Based on the data in Table 5, five student teachers (20.8%) wanted to ask ChatGPT for the duration of each activity.

- Brainstorming of several activities. The student teachers requested ChatGPT some ideas of activities in order to draw inspiration, for instance:

1) *What activities could we as teachers deliver in classroom to prepare students for the task in the museum?*

2) *Could you please create for me some activities addressed to 16-18-year-old students, English FL, B1 level, regarding the exhibition of Canaletto's Secret (at Ca' Rezzonico Museum in Venice) and related to sustainability?*

3) *What kind of activities could the students do?*

As indicated in Table 5, thirteen student teachers (54.4%) preferred first to draw their inspiration to then select the appropriate activity.

- Formulation of a single activity. The student teachers requested ChatGPT to create a single activity to carry out in the museum or classroom, to give some examples:

1) *Could you tell me a didactic activity to do in a museum of art?*

2) *Could you create an activity that motivates and engages students dealing with Klimt?*

3) *You have to create an activity for the following target: 16-18-year-old students.*

According to the data of Table 5, ten student teachers (41.7%) wanted to utilise this type of interaction.

- Creation of worksheets. The student teachers asked to create a worksheet to deliver to their learners without specifying any characteristics.

1) *Please, provide the duration of the visit to the museum, the materials students have to use for their activities, the skills involved (most of all speaking), the objectives, the lesson plan and a possible worksheet (if it is necessary for the activities).*

2) *Can you please also specify the duration of the single stages and provide me with a worksheet that I can give my students?*

As can be noticed from Table 5, only two student teachers (8.3%) tried to ask ChatGPT to create worksheets.

- Elaboration of a text for reading comprehension. The student teachers required ChatGPT to create a text on a particular topic, to give some examples:

- 1) *Could you give me a text related to sustainability with a B1 level?*
- 2) *Could you please write a text about selfies and portraits for 16-year-old-students with a B1 level of English?*
- 3) *Could you write a short text about the differences between selfies and portraits?*

Table 5 includes eight student teachers (33.3%) who decided to make ChatGPT generate a text.

- Request for specific exercises. The student teachers asked ChatGPT to design and structure precise exercises, for example:

- 1) *Could you please design a matching exercise related to art movements for students aged 16-18 with a B1 level in English as a foreign language?*
- 2) *I need a comprehension exercise about this text.*
- 3) *Could you please create a gap exercise about the art techniques related to Gustav Klimt?*

As demonstrated in Table 5, fourteen student teachers (58.3%) wanted to create specific exercises for their TBLL.

- Request for website resources. Some student teachers asked ChatGPT to provide links.

- 1) *Could you suggest a brief introductory video about Frida Kahlo?*
- 2) *Can you please link a website where I can find more information about the ongoing exhibition in Padua you mentioned?*
- 3) *And what about the website of the museum in Conegliano?*

Table 5 shows that two student teachers (8.3%) wanted to find some videos. Therefore, they asked ChatGPT.

- Creation of visual representations. In the prompts, the student teachers wanted to create some tables, charts, or graphs such as:

- 1) *Can you create a chart with some questions about paintings, please?*
- 2) *And a table with the same questions?*

According to the data presented in Table 5, only one teacher-student (4.2%) asked ChatGPT for the creation of a chart.

- Providing linguistic support. The student teachers wanted to provide learners with linguistic support. For this reason, they asked ChatGPT to create meaningful sentences to guide students in their conversations, for instance:

- 1) *Could you please provide some useful phrases for the presentation that students can use to talk about what their lives would have looked like if they lived in another decade?*
- 2) *Could you please create some basic questions in order to guide students to describe the subject matter of a painting?*
- 3) *Could you please create a list of phrases for students to describe a portrait or selfie?*

As shown in Table 5, four student teachers (16.7%) used this type of strategy.

- Enquiry about specific terminology as well as definitions. The student teachers asked ChatGPT to provide some vocabularies and definitions, for instance:

- 1) *Could you tell me some specific terms related to museums?*
- 2) *Can you please give me the definition of “colour”?*
- 3) *Could you please make a list of new vocabulary related to the artwork that I could use?*

Table 5 includes five student teachers (20.8%) who adopted this type of interaction.

- Request for suggestions or pieces of advice. The student teachers asked for opinions regarding a task, recommendations for how to use a tool or introduce a topic such as:

- 1) *Is this task appropriate?*
- 2) *Could you suggest a way to introduce Frida Kahlo and her work to a class of 16-18-year-old students?*
- 3) *I want my students to use the interactive tool of making snapshots of themselves in different decades. How can I promote speaking using this tool?*

As indicated in Table 5, four student teachers (16.7%) asked ChatGPT for suggestions.

- Modifications of the outputs. The student teachers were able to interact with ChatGPT by highlighting in their inputs some aspects to modify, for instance:

- 1) *I really like this activity, thank you. Nevertheless, with regard to the follow-up activities, I do not want them to choose one of the activities you proposed. I want to promote students' speaking skills. Can you please focus more on speaking in the follow-up activities and give me new ideas in detail?*
- 2) *I would like to make stage 2 more playful. Is it possible?*
- 3) *This sounds good! However, I want my students to present their selfies after they go to the museum so that they can directly relate it to their new knowledge from the*

museum about the different decades and explain what would have been different if they lived in another decade.

Based on the data of Table 5, eleven student teachers (45.8 %) modified the output provided by ChatGPT.

Table 6

Categorisation of appropriate and inappropriate prompt written by student teachers

Appropriate and inappropriate practices to interact with ChatGPT	Users	Percentage (%)
<i>Appropriate practices</i>		
• <i>Informative writing prompts</i>	2	8.3
• <i>Modifications of the outputs</i>	11	45.8
<i>Inappropriate practices</i>		
• <i>Request for website resources</i>	2	8.3
• <i>Creation of visual representations</i>	1	4.2
• <i>Omission of task details</i>	11	45.8
• <i>Design of the whole TBLL without mentioning Willis' framework</i>	4	16.7

Table 6 presents both appropriate and inappropriate practices to interact with ChatGPT. Even in this case, the author counted the number of student teachers who used a specific prompt and then calculated the percentage on the total of the subjects (24). The explanations of each category and related examples taken from the conversations of student teachers with ChatGPT are provided as follows:

- Informative writing prompts. The student teachers were able to guide ChatGPT by providing examples, explanations or further information to receive an appropriate output, for example:
 - 1) *Furthermore, I need to define the learning objectives: by the end of this activity, what will my students be able to do?*
 - 2) *Can you change the task to something more connected with the lesson? For example, students can write a descriptive paragraph about their favourite artwork from the museum.*

As indicated in Table 6, only two student teachers (8.3%) provided ChatGPT with examples to guide the machine to generate the “correct” output.

- Modifications of the outputs (see explanation of Table 5).
- Request for website resources (see explanation of Table 5).
- Creation of visual representations (see explanation of Table 5).
- Omission of task details. In their prompts, some student teachers did not include complete information regarding the target to address the TBLL, to provide some examples:

- 1) *Can you create an exercise about paintings for students of 18 years old?*
- 2) *Could you please prepare an activity for students at B1 English level about describing a picture and giving an opinion about it?*
- 3) *Could you please suggest a teaching activity about selfies and portraits, which includes the exhibition at Ca' Pesaro in Venice about Klimt? Please, note that the English level is a B1 and that the skill to promote is speaking.*

As noticed in Table 6, eleven student teachers (45.8%) did not mention all the details of the target to design their TBLL.

- Design of the whole TBLL without mentioning Willis' framework. The student teachers asked ChatGPT to develop the entire task without specifying to employ Willis' framework, e.g.:
- 1) *Could you write me a task designed for 16-year-old students which involves studying museum portraits?*
 - 2) *I'd like to create a task in a museum during which students can go through it and find similarities or differences in other artworks about Klimt's Judith II.*
 - 3) *What task do you suggest for this learning objective, considering the level of language B1 and the age of students?*

Table 6 shows that four student teachers (16.7%) did not mention that they should have realised a TBLL based on Willis' framework.

4.4. Data analysis of RQ3

In order to investigate how professors may assess the activities designed by their student teachers (*How can professors assess the activities created by the students through the use of ChatGPT?*), a self-study methodology was employed (Hamilton et al., 2009) through prompt-engineering. According to some academics (Hamilton & Pinnegar, 2000), self-study is a useful methodology to evaluate and comprehend teaching and learning experience better. Moreover,

the same scholars affirmed that self-study may aid in immediately changing teaching practices without awaiting new studies carried out by other researchers.

The author of the thesis was inspired by the research paper written by the researcher Cooper (2023) who also applied a self-study methodology to his exploratory study aiming at exploring ChatGPT. Furthermore, the elaboration of the prompts was based on the research paper written by White, et al. (2023) and the online course carried out by *OpenAI* and *DeepLearning.AI* (Fulford & Ng, 2023).

Firstly, an evaluation rubric was created and used to assess an online TBLL activity performed by the student teachers in line with the method used by the course professor. The academics Allen and Tanner (2006) stated that using a rubric for a specific assignment has several advantages. Firstly, the rubric clarifies the professor's standards as well as resulting grading; secondly, this assessment tool provides the learners with an explanation of how to reach the expectations of a higher level of performance; thirdly, utilising a rubric seems to be especially significant when the learners are novices concerning the design of a particular task. The assessment criteria suggested by the course professor were as follows²:

- Pertinence (with respect to the paper requirement, the target and learning context chosen).
- Originality, suitability and feasibility of the teaching proposal.
- Appropriateness of explanations in support of the proposal.
- Organisational aspects of the paper (appropriate structure and organization of materials, bibliographic references).

While the levels of performance were *no adequate*, *adequate*, *good* and *excellent*. A rubric was created with the help of ChatGPT and its criteria were modified slightly in order to make them more specific for the online TBLL activity carried out by the student teachers. After having obtained a good elaborated rubric with ChatGPT, the rubric and a student teachers' TBLL activity were inserted in the input to make ChatGPT assess the activity based on the assessment criteria and levels of performance included in the rubric (see Appendix A for the whole conversation).

² The criteria for the assessment of the teaching proposal for the course Instructional Design for Language Education (academic year 2022/2023) can be also found in the official webpage of the Ca' Foscari University: <https://www.unive.it/data/course/381868/programma>.

It is essential to keep in mind that the outputs provided by ChatGPT represent a response at that precise moment, signifying that the identical prompts provided to ChatGPT may generate different outputs. Both prompts and outputs below were gathered in July 2023. Due to the fact that in that month ChatGPT-3.5 did not work effectively (Chen et al., 2023), the author chose to employ ChatGPT-4.

Figure 22

Screenshots of the interaction with ChatGPT aiming at designing a rubric

a)

co Hi ChatGPT! I am a university teacher of the course Instructional design for language education. I need to assess a teaching proposal created by my students, which will become future teachers. Now, the majority of them have no experience in language teaching.

The university students have to create a task-based language learning (TBLL) activity following Willis' framework (1996) with the following structure: pre-task, task cycle, and post-task (or language focus).

The instructions for the task given by the professor are as follows (delimited with triple backticks):

```
...
1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
```

(see Appendix A for the whole conversation)

6. Describe in detail what you would do in the classroom before and after the museum (for example, activities, materials, ICT tools etc.). Specify the activity/exercises you would do, writing down the text of the exercises: e.g., if you create a multiple choice exercise for reading comprehension, specify both the questions and alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (e.g.: I agree with you; I couldn't agree with you more; That's so true. Etc.)

```
...
I need you to create some guidelines in a Rubric table to help me with the assessment of the task. Use the following criteria for the assessment to include in the rubric (delimited with triple backticks):
```

```
...
- Pertinence with respect to activity requirements (instructions of the task given by the professor)
- Pertinence with respect to national curriculum, educational program or course, Common European Framework of Reference for Languages (CEFR) according to the target;
- Pertinence with respect to the learning objectives chosen by the university students.
- Originality: activities and creation of materials that you do not normally find in books or that are not copied from books or online resources.
- Suitability and feasibility of the teaching proposal in terms of time (timing for each activity and whether the activity can be completed based on the allotted time decided by the university students).
- Suitability and feasibility of the teaching proposal in terms of target (age, language knowledge, level of languages, cultural background, special needs...).
```

- Suitability and feasibility of the teaching proposal in terms of learning context chosen (type of school, resources present in the school chosen, museum setting, educational approach used).

- The project follows the task-based language learning structure by Willis' framework (1996), that is pre-task, task cycle, and post-task.

- Appropriate organization of the educational material employed in each phase (pre-task, task cycle, and post-task)

...

Structure the Rubric guidelines in a table where each row is one of the above criteria and each column is one of the following grade levels:

- No adequate

- Adequate


- Good

- Excellent

Write in each cell the things I should check very specifically and concisely.

Thanks for the help!


b)

 Here is a rubric table that you can use to evaluate your students' assignments.


Evaluation Criteria	No Adequate	Adequate	Good	Excellent
Pertinence with respect to activity requirements	Task does not meet the professor's instructions	Task meets some but not all of the professor's instructions	Task meets most of the professor's instructions	Task completely meets all of the professor's instructions

(See Appendix A for the whole first rubric)

c)

 Please rewrite the table being more specific in each step


d)

 Absolutely, here's a more specific version of your evaluation rubric:

Evaluation Criteria	No Adequate	Adequate	Good	Excellent
Pertinence with respect to activity requirements	Task misses multiple or key elements of the professor's instructions.	Task includes most instructions but misses or incorrectly implements a few.	Task incorporates nearly all instructions, with minor errors.	Task thoroughly and accurately incorporates all of the professor's instructions.

(See Appendix A for the whole second rubric)

e)

 Can you restructure the table as a list? Following the structure:

...

1. Pertinence with respect to activity requirements:

- No adequate: Task misses multiple or key elements of the professor's instructions.

- Adequate: ...

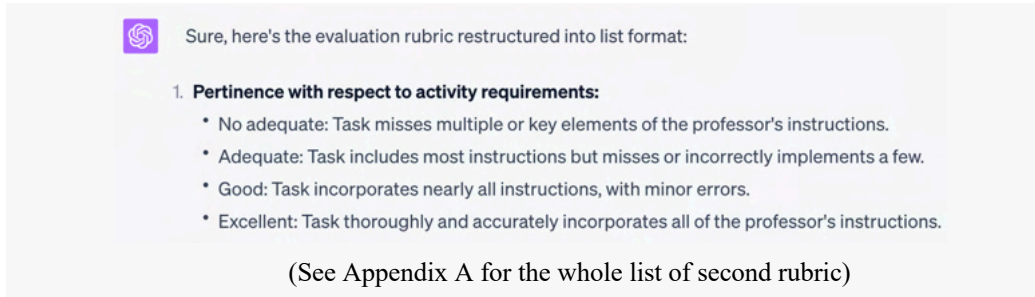
- Good: ...

- Excellent: ...

...

...

f)



Note. The first screenshot (a) presents the prompt for ChatGPT to design a rubric. It includes the instructions of the TBLL activity as well as the guidelines for writing the rubric. The second screenshot (b) is the ChatGPT's answer including the first rubric. The third screenshot (c) includes the second prompt, in which ChatGPT is required to be more precise. The fourth screenshot (d) includes the second rubric designed by ChatGPT. The fifth (e) and sixth (f) screenshots represent the corresponding input and output of transforming the table format of the rubric into a list, as it was not possible in July 2023 to insert a table in a prompt.

Figure 22 shows the three inputs (a, c, e) given by the author and the three corresponding outputs (b, d, f) provided by ChatGPT. The first prompt was written by following a guideline explained in the online course by the scientists Fulford and Ng (2023). The principle applied in the input was to include specific as well as clear instructions by using some delimiters such as triple backticks (```) and ask for a table structure. Moreover, the author implicitly explained what a rubric is to ChatGPT by writing down “I need you to create some guidelines in a rubric table to help me with the assessment of the task. Use the following criteria for the assessment to include in the rubric (delimited with triple backticks) and then, structure the rubric guidelines in a table where each row is one of the above criteria and each column is one of the following grade levels: no adequate; adequate; good; excellent. Write in each cell the things I should check very specifically and concisely”.

It is also noticed in Figure 22 that the first rubric (b) created by ChatGPT is neither specific nor detailed, while the second table is slightly better (d). Afterwards, the author asked ChatGPT to rewrite the table as a list because users cannot give tables as prompts so far (July 2023).

Figure 23

Example of slightly better formulation of evaluation criteria in the second output of ChatGPT

a)	Pertinence with respect to the learning objectives	Learning objectives are unclear or not related to the task	Learning objectives are somewhat clear and related to the task	Learning objectives are clear and closely related to the task	Learning objectives are clear, concise and directly related to the task
	Suitability and feasibility of the proposal in terms of time	Timing is inappropriate or unrealistic for the activities	Timing is somewhat realistic but could be optimized	Timing is realistic and well-planned for most activities	Timing is perfectly planned and feasible for all activities
	Suitability and feasibility of the proposal in terms of target	The proposal is unsuitable for the age, language knowledge, level, cultural background, or special needs of the target	The proposal is somewhat suitable for the target	The proposal is generally suitable for the target	The proposal is perfectly suitable for the target
b)	Pertinence with respect to the learning objectives	Learning objectives are unclear, unrelated to the task, or not included.	Learning objectives are stated, but may not align well with the task or provide clear goals.	Learning objectives align with the task and provide clear goals for most activities.	Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.
	Suitability and feasibility of the proposal in terms of time	The timing suggested is unrealistic, does not allow for task completion, or isn't provided.	Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.	Timing is mostly realistic, with considerations for transitions and possible delays in a few activities.	The suggested timing is completely realistic, provides buffer for transitions and potential delays, and allows for task completion.
	Suitability and feasibility of the proposal in terms of target	The proposal does not match the target students' age, English language level, cultural background, or special needs.	The proposal partially matches the target students but lacks in one or more areas.	The proposal matches the target students in most areas and makes minor errors in others.	The proposal perfectly matches the target students' age, English language level, cultural background, and special needs.

Note. The first screenshot (a) includes three evaluation criteria of the first output of ChatGPT. The second screenshot (b) presents the same three evaluation criteria, written more precisely during the second output.

Figure 23 shows three evaluation criteria written in the first output (a) as well as in the second output (b) provided by ChatGPT: pertinence with respect to the learning objective, suitability and feasibility of the proposal in terms of time, and suitability and feasibility of the proposal in

terms of target. As can be seen in Figure 23b the evaluation criteria of the second rubric seem to be written in a more detailed and precise way. However, Figure 23b does not include specific examples to comprehend better which grade level to choose.

Figure 24

Example of not-so-good formulation of evaluation criteria in the second output of ChatGPT

a)

Evaluation Criteria	No Adequate	Adequate	Good	Excellent
Pertinence with respect to activity requirements	Task does not meet the professor's instructions	Task meets some but not all of the professor's instructions	Task meets most of the professor's instructions	Task completely meets all of the professor's instructions
Pertinence with respect to national curriculum, CEFR	No alignment with the national curriculum or CEFR standards	Some alignment with the national curriculum or CEFR standards	Good alignment with the national curriculum or CEFR standards	Excellent alignment with the national curriculum and CEFR standards
Originality	Materials and activities are copied or unoriginal	Some original activities and materials, but also some copied	Most activities and materials are original	All activities and materials are original

b)

Evaluation Criteria	No Adequate	Adequate	Good	Excellent
Pertinence with respect to activity requirements	Task misses multiple or key elements of the professor's instructions.	Task includes most instructions but misses or incorrectly implements a few.	Task incorporates nearly all instructions, with minor errors.	Task thoroughly and accurately incorporates all of the professor's instructions.
Pertinence with respect to national curriculum, CEFR	Task does not align with national curriculum or CEFR B1 English level standards.	Task shows a basic alignment with national curriculum and CEFR B1 English level standards.	Task aligns well with national curriculum and CEFR B1 English level standards, with minor exceptions.	Task demonstrates an excellent alignment and understanding of national curriculum and CEFR B1 English level standards.
Originality	Materials and activities are copied from sources without any original input.	Some original content but also a significant amount of copied or common activities and materials.	Most activities and materials show original thinking, with a small amount of commonly used or copied content.	All activities and materials are original, innovative, and demonstrate unique thinking.

Note. The first screenshot (a) includes three evaluation criteria of the first output of ChatGPT. The second screenshot (b) presents the same three evaluation criteria, written not so differently from the second output.

Figure 24 includes three further evaluation criteria written in the first output (a) as well as in the second output (b) provided by ChatGPT: pertinence with respect to activity requirements, pertinence with respect to national curriculum, CEFR, and originality. As can be seen in Figure 23b the evaluation criteria of the second rubric do not seem to be so different from the ones of the first rubric.

Figure 25

Two different types of prompt processes (one evaluation criterion at a time versus all evaluation criteria in one single prompt) to assess with ChatGPT a TBLL activity

CO Hi ChatGPT! I need you to assess a task from some university students of the course of Instructional design for language education. The students of this course will become teachers in the future. Now, the majority of them have no experiences in language teaching.

The university students have to create a task-based language learning (TBLL) activity following Willis' framework (1996) with the following structure: pre-task, task cycle, and post-task (or language focus).

The instructions of the task given by the professor are as follows (delimited in triple backticks):
 ```
 1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
 ```

(See Appendix A for the whole conversation)

and the task they created is (delimited in triple backticks):
 ```
 \* Topic: selfies and portraits
 \* Museum: National Portrait Gallery, London (FREE admission)
 \* Skill to promote: speaking
 ```

a) (See Appendix A for the whole TBLL activity designed by a group of student teachers)

2. Students are asked to take a selfie that represents their identity to the fullest, by also using objects and clothes they find representative of their own personality. The selfies will be presented and described to the classmates, who will try to interpret their peers' images and their creative choices. Finally, the creator of the selfie will explain the reasons behind their choices and talk about themselves.

Can you assess it according to the following guidelines delimited in triple backticks?
 First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :
 ```
 Pertinence with respect to activity requirements:
 \* No adequate: The student's activity misses multiple or key elements of the instructions.
 \* Adequate: The student's activity includes most instructions but misses or incorrectly implements a few.
 \* Good: The student's activity incorporates nearly all instructions, with minor errors.
 \* Excellent: The student's activity thoroughly and accurately incorporates all of the professor's instructions.
 ```


b)

CO All right. Now could you assess the task according to the following guidelines delimited in triple backticks?
 First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :
 ...

2. Pertinence with respect to national curriculum, CEFR:
 * No adequate: The student's activity does not align with national curriculum or CEFR B1 English level standards.
 * Adequate: The student's activity shows a basic alignment with national curriculum or CEFR B1 English level standards.
 * Good: The student's activity aligns well with national curriculum and CEFR B1 English level standards, with minor exceptions.
 * Excellent: The student's activity demonstrates an excellent alignment and understanding of national curriculum or CEFR B1 English level standards.
 ...

CO Hi ChatGPT! I need you to assess a task from some university students of the course of Instructional design for language education. The students of this course will become teachers in the future. Now, the majority of them have no experiences in language teaching.

The university students have to create a task-based language learning (TBLL) activity following Willis' framework (1996) with the following structure: pre-task, task cycle, and post-task (or language focus).

(See Appendix A for the whole prompt)

c)

8. Following the Willis' TBLL structure:
 * No adequate: The TBLL structure is not followed or elements are missing.
 * Adequate: The TBLL structure is attempted but lacks a clear pre-task, task cycle, or post-task.
 * Good: The TBLL structure is mostly followed with minor deviations.
 * Excellent: The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

9. Organization of the educational material:
 * No adequate: Materials are disorganized, lacking, or inappropriate for the phases of the task.
 * Adequate: Materials are organized and appropriate for some but not all phases.
 * Good: Materials are mostly well-organized and appropriate for the phases of the task.
 * Excellent: Materials are meticulously organized and perfectly suited for each phase of the task.
 ...

Note. The evaluation criteria used to assess the TBLL activity are those belonging to the rubric created by ChatGPT and slightly modified by the author (Figure 22d). The first screenshot (25a) is the prompt written by the author, in which the professor's instructions, the TBLL task created by a group of student teachers, and the first evaluation criterion to assess are included. The second screenshot (25b) is the consecutive prompt of the previous one (25a), in which the second criterion to assess is included. The third screenshot (25c) represents the other type of prompt, where the professor's instructions, the TBLL task created by a group of student teachers, and all the evaluation criteria are given in one single prompt.

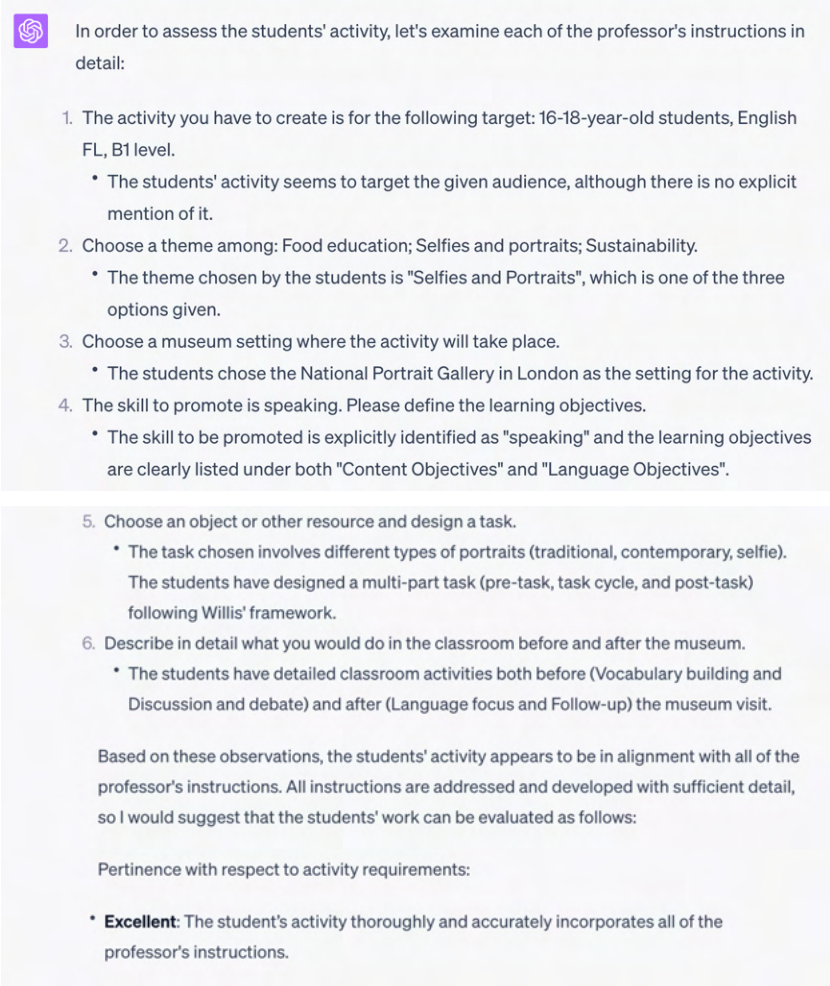
Figure 25 illustrates two different types of prompt processes (a, c), aiming at conducting the assessment of a TBLL activity through the rubric created with the aid of ChatGPT and slightly modified by the author (Figure 22d). Both prompts include the professor's instructions as well


as the TBLL task created by a group of student teachers. However, one process gives nine prompts and, in each prompt, an evaluation criterion of the rubric designed by ChatGPT and slightly modified by the author (Figure 22d) is included (Figures 23a and 23b), while the other process includes the whole rubric in a single prompt (Figure 23c). The prompt (Figure 23a and 23b) was written based on the following principle: giving ChatGPT time to reflect (Fulford & Ng, 2023). By doing so, the author fragmented the evaluation criteria one by one (Figure 23a and b), forced ChatGPT to first analyse and comment on the whole task, and afterwards, she asked it to assign a level of performance.

Figure 26

Evaluation criterion: pertinence of activity requirements

a)



 In order to assess the students' activity, let's examine each of the professor's instructions in detail:

1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
 - The students' activity seems to target the given audience, although there is no explicit mention of it.
2. Choose a theme among: Food education; Selfies and portraits; Sustainability.
 - The theme chosen by the students is "Selfies and Portraits", which is one of the three options given.
3. Choose a museum setting where the activity will take place.
 - The students chose the National Portrait Gallery in London as the setting for the activity.
4. The skill to promote is speaking. Please define the learning objectives.
 - The skill to be promoted is explicitly identified as "speaking" and the learning objectives are clearly listed under both "Content Objectives" and "Language Objectives".
5. Choose an object or other resource and design a task.
 - The task chosen involves different types of portraits (traditional, contemporary, selfie). The students have designed a multi-part task (pre-task, task cycle, and post-task) following Willis' framework.
6. Describe in detail what you would do in the classroom before and after the museum.
 - The students have detailed classroom activities both before (Vocabulary building and Discussion and debate) and after (Language focus and Follow-up) the museum visit.

Based on these observations, the students' activity appears to be in alignment with all of the professor's instructions. All instructions are addressed and developed with sufficient detail, so I would suggest that the students' work can be evaluated as follows:

Pertinence with respect to activity requirements:

- **Excellent:** The student's activity thoroughly and accurately incorporates all of the professor's instructions.

b)

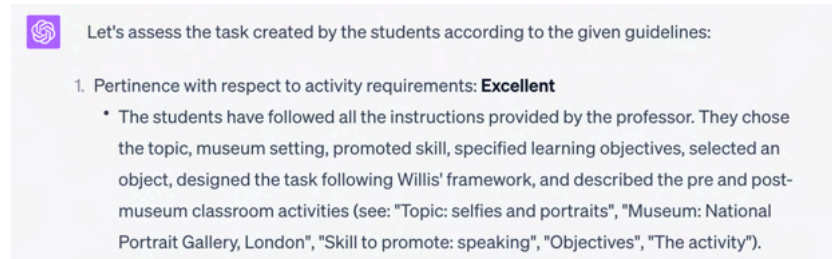
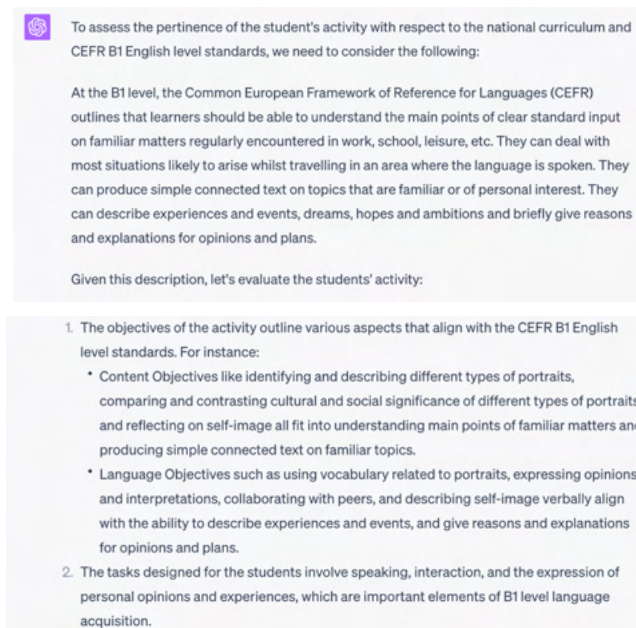


Figure 26 depicts the two different assessment outputs provided by ChatGPT with respect to the section *pertinence of activity requirements*. Comparing the two different assessments (Figures 26a and 26b), it can be claimed that, when giving one criterion at a time (Figure 26a), ChatGPT is able to analyse every single step by pointing out some discrepancies, such as the group did not mention the student target. Whereas, when the AI chatbot is provided with the whole rubric in a single prompt (Figure 26b), ChatGPT summarises the most important features in few lines without being so specific (in this case, it does not indicate that the target is missing).

Figure 27

Evaluation criterion: pertinence of national curriculum and CEFR

a)



b)

3. The activity is designed to encourage learners to deal with a situation (a museum visit), where they have to use their language skills in a real-world context.

Based on this analysis, the activity aligns well with the CEFR B1 English level standards. Regarding the national curriculum, without specific details about which country's curriculum to refer to, I will assume that any curriculum that aligns with the CEFR standards would find this activity suitable.

Hence, the students' work can be evaluated as follows:

Pertinence with respect to national curriculum, CEFR:

- **Excellent:** The student's activity demonstrates an excellent alignment and understanding of national curriculum or CEFR B1 English level standards.

2. Pertinence with respect to national curriculum, CEFR: **Good**

- The task incorporates elements such as speaking, vocabulary building, and interactive activities that align with the B1 English level according to CEFR. However, without explicit referencing to the national curriculum or specific CEFR descriptors, it's hard to determine full alignment (see: "Objectives", "The activity").

In the section *pertinence of national curriculum and CEFR*, when giving one criterion at a time (Figure 27a), ChatGPT starts from a description of what CEFR requires for the B1 level and points out that the prompt did not specify which country's curriculum the activity refers to. ChatGPT proposes "excellent" as the final mark for this section. On the contrary, when the AI chatbot is provided with the whole rubric in one prompt (Figure 27b), it gives "good" as the final mark and explains the reasons behind its choice: "Without explicit referencing to the national curriculum or specific descriptors, it's hard to determine full assignment". Therefore, both versions spotted this error, and maybe the grade given by ChatGPT with the whole rubric in one prompt (Figure 27b) is more adequate for this section.

Figure 28

Evaluation criterion: pertinence of learning objectives

a)

In assessing the pertinence with respect to the learning objectives, the following points are observed:

The learning objectives are clearly stated in two categories: content and language objectives.

- **Content Objectives:**
 - Students will be able to identify and describe different types of portraits.
 - Students will be able to compare and contrast the cultural and social significance of different types of portraits.
 - Students will be able to reflect on their self-image.
- **Language Objectives:**
 - Students will be able to use vocabulary related to portraits.
 - Students will be able to express their opinions and interpretations about different portraits.
 - Students will be able to collaborate with their peers to create a group portrait.
 - Students will be able to use appropriate language to describe their self-image verbally.

The task designed aligns well with these learning objectives:

- The pre-task of brainstorming different types of portraits and discussing what makes a good portrait aligns with the objectives of identifying, describing, and reflecting on portraits.
- The task cycle of analyzing a chosen portrait and creating a group portrait, followed by planning and reporting, aligns with the objectives of comparing, contrasting, expressing opinions, and using appropriate vocabulary.
- The post-task activities, including playing a game of "Taboo" and being provided with language structures to express opinions, align with the objectives of using related vocabulary and expressing opinions.
- The follow-up activity of taking a selfie that represents their identity and explaining it to their peers also aligns with the objective of reflecting on their self-image.

These activities set clear and measurable outcomes related to the stated objectives. For instance, the students are not only expected to learn about portraits but also use the relevant vocabulary and express their own interpretations, which can be clearly observed and assessed during the tasks.

Therefore, I would evaluate the students' work as follows:

Pertinence with respect to the learning objectives:

- **Excellent:** Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.

b)

3. Pertinence with respect to the learning objectives: **Excellent**

- The learning objectives align well with the tasks and set clear and measurable outcomes for all activities (see: "Objectives").

Regarding the *pertinence of learning objectives*, when giving one criterion at a time (Figure 28a), ChatGPT analyses every single learning goal based on the phases of Willis's framework (pre-task, task cycle, and post-task). Moreover, ChatGPT makes a short comment on the outcomes related to the stated learning objectives. On the other side, when ChatGPT is provided with the whole rubric in a single prompt (Figure 28b), it only writes a really short comment on the learning objectives.

Figure 29

Evaluation criterion: originality

a)

In assessing the originality of the task, the following aspects are considered:

- The theme of "Selfies and Portraits" and the setting of the National Portrait Gallery are not uncommon in language learning activities, but the way they are woven together is somewhat unique.
- The structure of the task, following Willis' model, is not original in itself but it is a pedagogically sound and recognized method in task-based language learning.
- The idea of conducting a part of the lesson in a museum is relatively original and can provide a unique learning experience to the students.
- The use of a Kahoot quiz for vocabulary building is a common method used in classrooms, but the questions prepared for the quiz are specific and original.
- The discussions and debates before the museum visit are good strategies for activating prior knowledge and stimulating students' thinking about the theme, and the questions suggested are interesting and thought-provoking.
- The task cycle involving analyzing a chosen portrait and creating a group portrait in the museum, and later presenting their work to the class, demonstrates a degree of originality.
- The post-task activities like the game of "Taboo" are commonly used, but they are well adjusted to the task.
- The follow-up activity where students take a selfie that represents their identity and explain it to their peers is quite original and personal, promoting creative and reflective thinking in students.

b)

Overall, while the task uses some common methods and activities, it also presents a good deal of original content and applies common methods in original ways.

Therefore, I would evaluate the students' work as follows:

Originality:

- **Good:** Most activities and materials of the student's work show original thinking, with a small amount of commonly used content.

4. Originality: **Good**

- The task shows some level of originality, particularly in the use of a museum setting and incorporating portrait analysis and creation. However, use of a Kahoot quiz and discussion questions can be found in many teaching resources (see: "The activity").

As far as *originality* is concerned, when giving one criterion at a time (Figure 29a), ChatGPT examines several points and explains in detail what it considers original and unoriginal. In case ChatGPT receives the whole rubric in one prompt (Figure 29b), the LLM chatbot only summarises the positive and negative aspects of the originality of the group in few lines.

Figure 30

Evaluation criterion: suitability and feasibility of the proposal in terms of time

a)

Let's evaluate the suitability and feasibility of the proposal in terms of time:

- The pre-task activities include a Kahoot quiz and a debate, which are typically manageable in a single class period, but the timing isn't explicitly specified. However, these activities could possibly take more time than expected, especially if students engage deeply with the debate.
- The task cycle involves a visit to the National Portrait Gallery, which includes travel time, time spent at the museum, and a post-visit class. Depending on the location of the school in relation to the museum, travel time could potentially complicate the timing. Additionally, timing may also depend on the size of the class and the complexity of the portraits the students choose to analyze.
- Post-task activities include a game of "Taboo" and a follow-up activity of taking selfies, which are generally feasible within a single class period. However, the timing again isn't explicitly stated.
- Language focus activities are typically conducted over several class periods, and would likely fit within the estimated two weeks. However, depending on the learners' pace and needs, it might take longer.

In general, the proposal seems potentially feasible within the two-week timeframe, although the specifics are somewhat uncertain, and there might be potential delays due to factors like the museum visit or deep engagement with the activities.

Therefore, based on the given scale, I would evaluate this aspect as:

Suitability and feasibility of the proposal in terms of time:

- **Adequate:** Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.

b)

5. Suitability and feasibility of the proposal in terms of time: **Adequate**

- The timing for the tasks seems generally realistic. However, the time allotment for the classroom activities (Kahoot quiz, discussion and debate, language game, and selfie presentation) and the time for transitions and possible delays are not explicitly addressed (see: "The activity").

Concerning the *suitability and feasibility of the proposal in terms of time*, when giving one criterion at a time (Figure 30a), it can be seen that ChatGPT is able to take into consideration several factors that could affect the time allotment, for instance, a debate may need more time than expected: it could depend on students' engagement, travel time, size of the class, complexity of the portraits to analyse, students' needs and learning pace. Additionally, ChatGPT identifies the parts where the duration of the activities is missing. When ChatGPT receives the whole rubric in a single input (Figure 30b), it is able to identify the same positive and negative characteristics, which are recapped in few lines.

Figure 31

Evaluation criterion: suitability and feasibility of the proposal of the target

a)



To evaluate the suitability and feasibility of the proposal in terms of the target, let's look at the task in relation to the specified target audience: 16-18-year-old students, English as a foreign language, B1 level.

- Age: The task is designed with an appropriate level of complexity and engagement for 16-18-year-old students. The theme of "Selfies and Portraits" is likely to be interesting and relevant for this age group, and the activities such as the museum visit, group work, and selfie task are suitable for this age.
- English Language Level: The task appears to be designed with B1 level English learners in mind. The vocabulary and language structures to be used in the task seem to be at the right level of difficulty for B1 learners. Activities like the Kahoot quiz, group discussions, and the presentation should be manageable for B1 students.
- Cultural Background: While the task takes place in the National Portrait Gallery in London, it deals with a universal theme of self-portrait and identity which is likely to be applicable to students from various cultural backgrounds. It could be adapted to other settings if the museum visit is not possible for geographical reasons.

- **Special Needs:** There is no specific mention of accommodations for students with special needs. However, the task does not seem to have any obvious barriers that could prevent students with certain special needs from participating, provided that standard accommodations are in place.

Given these considerations, the proposal appears to match well with the target audience, albeit with minor areas that might need additional considerations or adjustments.

Therefore, based on the given scale, I would evaluate this aspect as:

Suitability and feasibility of the proposal in terms of target:

- **Good:** The proposal matches the target students in most areas and makes minor errors in others.

b)

6. Suitability and feasibility of the proposal in terms of target: **Excellent**

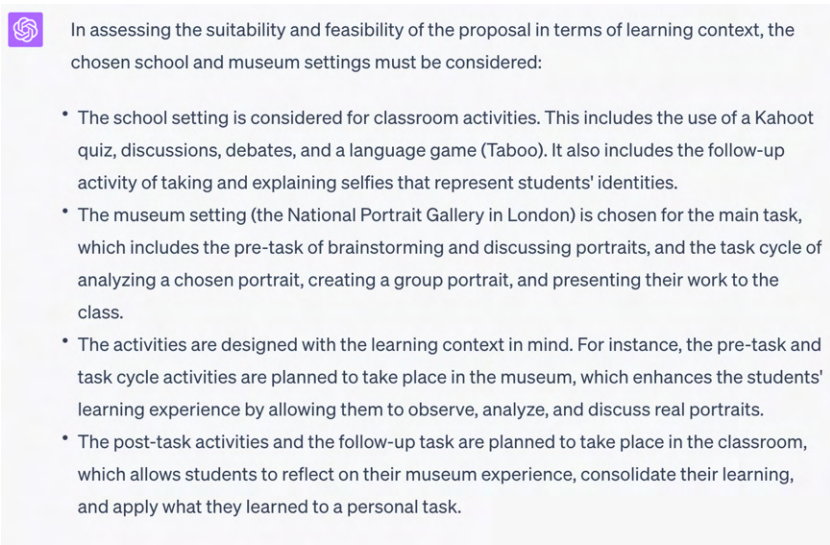
- The activities proposed are very suitable for the target group of 16-18 year-olds, B1 level English learners. The topic is age-appropriate and interesting for teenagers, and the language level seems to match a B1 level (see: "Objectives", "The activity").

As regards the *suitability and feasibility of the proposal of the target*, when giving one criterion at a time (Figure 31a), ChatGPT scrutinises the age, English language level, cultural background, as well as special needs. The data show that ChatGPT is able to identify positive and negative aspects and gives “good” as the final mark. In this circumstance, ChatGPT found that the group did not mention “learners with special needs” (as the target was already given by the professor of the course). It seems that, in this case, ChatGPT has been strict. Maybe ChatGPT could highlight this aspect but in the form of advice, such as “It would be relevant to specify whether there are or aren’t students with special needs”. However, when giving ChatGPT the whole rubric in one prompt (Figure 31b), it assesses this section with “excellent”. In fact, the activity seems to be suitable for 16-18-year-old students with a B1 English level.

Figure 32

Evaluation criterion: suitability and feasibility of learning context

a)



In assessing the suitability and feasibility of the proposal in terms of learning context, the chosen school and museum settings must be considered:

- The school setting is considered for classroom activities. This includes the use of a Kahoot quiz, discussions, debates, and a language game (Taboo). It also includes the follow-up activity of taking and explaining selfies that represent students' identities.
- The museum setting (the National Portrait Gallery in London) is chosen for the main task, which includes the pre-task of brainstorming and discussing portraits, and the task cycle of analyzing a chosen portrait, creating a group portrait, and presenting their work to the class.
- The activities are designed with the learning context in mind. For instance, the pre-task and task cycle activities are planned to take place in the museum, which enhances the students' learning experience by allowing them to observe, analyze, and discuss real portraits.
- The post-task activities and the follow-up task are planned to take place in the classroom, which allows students to reflect on their museum experience, consolidate their learning, and apply what they learned to a personal task.

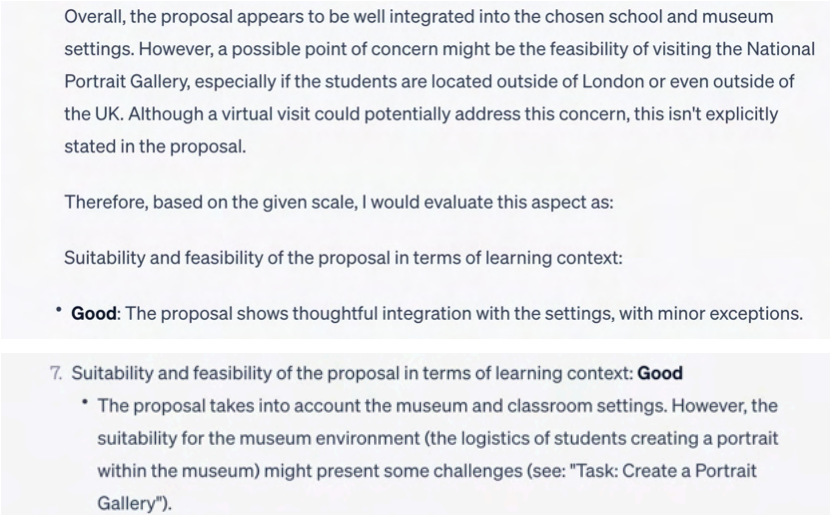
Overall, the proposal appears to be well integrated into the chosen school and museum settings. However, a possible point of concern might be the feasibility of visiting the National Portrait Gallery, especially if the students are located outside of London or even outside of the UK. Although a virtual visit could potentially address this concern, this isn't explicitly stated in the proposal.

Therefore, based on the given scale, I would evaluate this aspect as:

Suitability and feasibility of the proposal in terms of learning context:

- **Good:** The proposal shows thoughtful integration with the settings, with minor exceptions.

b)



7. Suitability and feasibility of the proposal in terms of learning context: **Good**


- The proposal takes into account the museum and classroom settings. However, the suitability for the museum environment (the logistics of students creating a portrait within the museum) might present some challenges (see: "Task: Create a Portrait Gallery").

Referring to *suitability and feasibility of learning context*, when giving one criterion at a time (Figure 32a), ChatGPT is capable of taking into account some particulars, such as the fact that it is not specified where students come from to reach and visit the National Portrait Gallery in London. For this reason, it suggests that maybe it would have been better to conduct a virtual visit (but according to the professor's instructions, virtual visits were only possible in one of the Musei Civici of Venice). However, when giving ChatGPT the whole rubric in one prompt (Figure 32b), it considers another curious aspect: ChatGPT focuses more on the activity carried out inside the museum. In fact, the AI chatbot underlines the difficulty in the logistics of creating a drawing within the museum. Both outputs give "Good" as a mark, but due to two different reasons.

Figure 33

Evaluation criterion: Willis's framework

a)

 In order to evaluate the extent to which the task follows Willis' Task-Based Language Learning (TBLL) structure, we need to look at how the task was designed with regards to the pre-task, task cycle, and post-task phases:

- Pre-task: The students designed a clear pre-task phase that includes introducing and practicing relevant vocabulary through a Kahoot quiz, as well as engaging students in discussions and debates about the topic of "Selfies and Portraits". This phase is meant to prepare students for the main task at the museum.
- Task cycle: The students' task cycle is also clearly outlined. It begins at the museum, where students, in small groups, select and analyze a portrait and then create their own group portrait drawing inspiration from the analyzed portrait. This is followed by a planning and reporting phase in which students plan how they will present their work to the class, and then do so.
- Post-task: The students' post-task phase includes a language focus activity and a follow-up task. In the language focus activity, students play a game of "Taboo" and are provided with language structures to express their ideas and opinions. The follow-up task involves a personal project where students take a selfie that represents their identity, present it to the class, and explain the reasons behind their choices.

Based on these observations, it can be seen that the task adheres to Willis' TBLL structure. All three phases are clearly delineated and thoughtfully designed to scaffold the students' learning.

Therefore, based on the given scale, I would evaluate this aspect as:

Following the Willis' TBLL structure:

- **Excellent:** The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

b)

8. Following the Willis' TBLL structure: **Excellent**

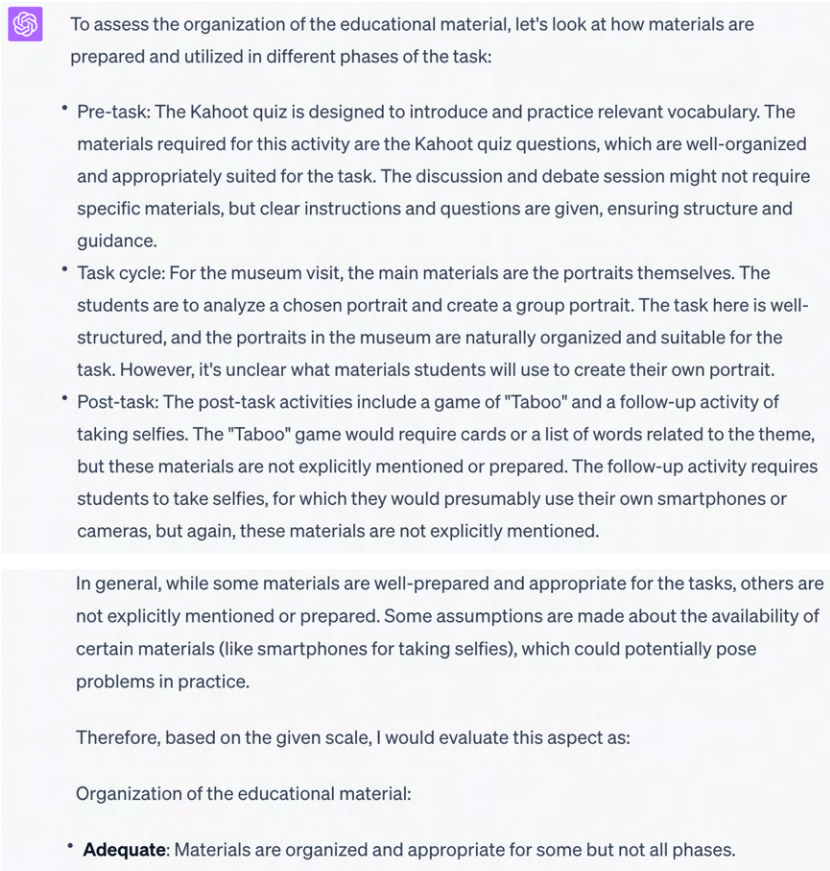
- The task follows the Willis' TBLL structure very well, with a clear pre-task, task cycle, and post-task (see: "Task: Create a Portrait Gallery").

As to *Willis's framework*, when giving one criterion at a time (Figure 33a), ChatGPT recognises through several points that the TBLL task adheres to Willis's framework, and underlines the positive aspects of the task. Whereas, when ChatGPT receives the whole rubric in a single input (Figure 33b), the LLM chatbot seems to only identify the keywords of pre-task, task cycle, and post-task.

Figure 34

Evaluation criterion: organisation of the educational material

a)



To assess the organization of the educational material, let's look at how materials are prepared and utilized in different phases of the task:

- Pre-task: The Kahoot quiz is designed to introduce and practice relevant vocabulary. The materials required for this activity are the Kahoot quiz questions, which are well-organized and appropriately suited for the task. The discussion and debate session might not require specific materials, but clear instructions and questions are given, ensuring structure and guidance.
- Task cycle: For the museum visit, the main materials are the portraits themselves. The students are to analyze a chosen portrait and create a group portrait. The task here is well-structured, and the portraits in the museum are naturally organized and suitable for the task. However, it's unclear what materials students will use to create their own portrait.
- Post-task: The post-task activities include a game of "Taboo" and a follow-up activity of taking selfies. The "Taboo" game would require cards or a list of words related to the theme, but these materials are not explicitly mentioned or prepared. The follow-up activity requires students to take selfies, for which they would presumably use their own smartphones or cameras, but again, these materials are not explicitly mentioned.

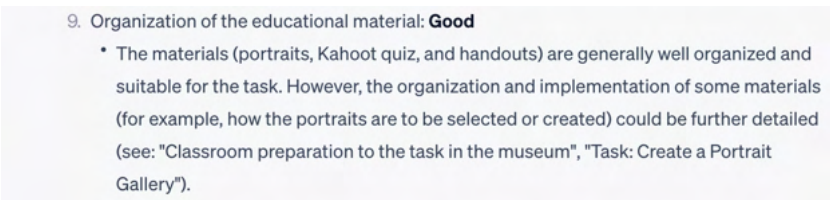
In general, while some materials are well-prepared and appropriate for the tasks, others are not explicitly mentioned or prepared. Some assumptions are made about the availability of certain materials (like smartphones for taking selfies), which could potentially pose problems in practice.

Therefore, based on the given scale, I would evaluate this aspect as:

Organization of the educational material:

- **Adequate:** Materials are organized and appropriate for some but not all phases.

b)



9. Organization of the educational material: **Good**


- The materials (portraits, Kahoot quiz, and handouts) are generally well organized and suitable for the task. However, the organization and implementation of some materials (for example, how the portraits are to be selected or created) could be further detailed (see: "Classroom preparation to the task in the museum", "Task: Create a Portrait Gallery").


Regarding the *organisation of the educational material*, when giving one criterion at a time (Figure 34a), ChatGPT investigates the materials used in each phase (pre-task, task cycle, and post-task). The AI chatbot points out not only the positive features of the activity but also the negative ones. For example, it detects the incomprehensibility of what materials students will use when drawing the portrait within the museum. Furthermore, ChatGPT mentions that the group did not always include the material students have to employ, such as mobile phones or cards for playing Taboo. Hence, ChatGPT decides to assess student teacher activity with “Adequate”. Instead, when giving ChatGPT the whole rubric in one prompt (Figure 34b), it assesses this section with “Good”. The AI chatbot explains that on the one hand, some materials

are explained well, while others are not mentioned. However, in this case, ChatGPT seems to be less specific.

Figure 35

Feedback provided by ChatGPT based on the previous outputs

a) Prompt  Now, could you write a single textual feedback including the 9 assessments you previously gave? Try to first highlight the best parts and then talk about the improvements that could be done

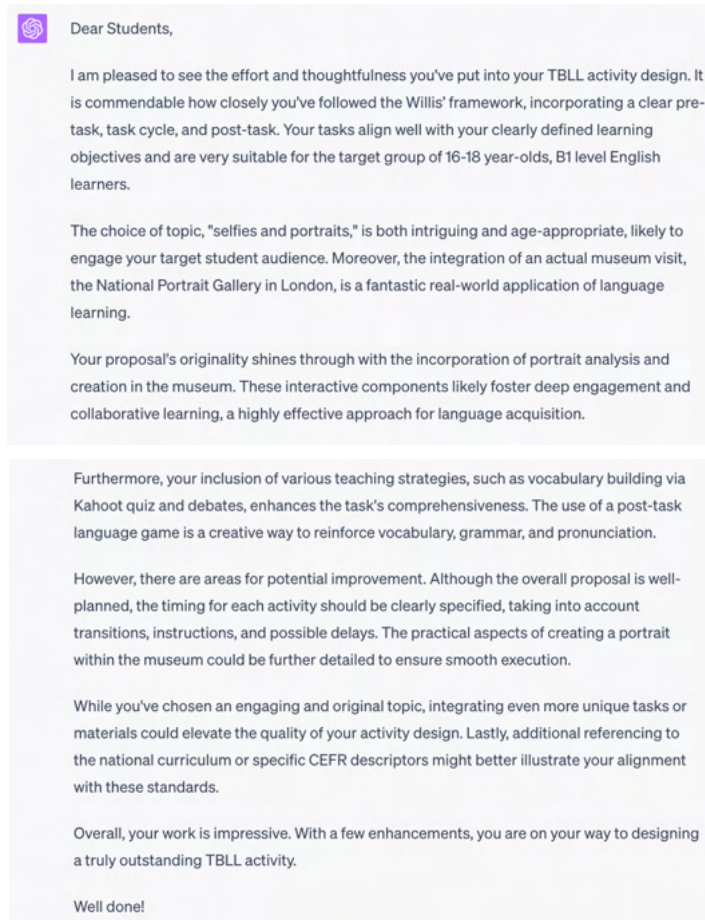
b) Feedback A  Sure, I'd be happy to summarize the assessments into a single feedback.

Overall, the task you created shows a thorough understanding of the task-based language learning (TBLL) approach and is mostly well-designed, innovative, and pedagogically sound.

- **Strengths:** The strongest aspects of your task are its alignment with Willis' TBLL structure and the pertinence with respect to the professor's instructions, the national curriculum, CEFR B1 English level standards, and the learning objectives. Your task demonstrates an excellent understanding of these frameworks and standards, as well as the ability to create clear, measurable learning objectives that align with your chosen task.
- The theme of "Selfies and Portraits" and the setting of the National Portrait Gallery are not only original and interesting but also aptly chosen for the age group and English language level of the target students. Your activities show thoughtful integration with the chosen settings and a well-considered balance between classroom and museum activities.
- Your use of a Kahoot quiz for vocabulary building is innovative and engaging, and your follow-up task of taking and explaining a selfie that represents a student's identity is an excellent example of a task that is not only engaging and personally relevant but also promotes reflection and deeper learning.

- **Areas for Improvement:** There are a few areas that could be improved. Regarding the timing, while the proposal seems potentially feasible within the two-week timeframe, more detailed consideration should be given to possible delays due to factors like the museum visit or deep engagement with the activities. The organization of the educational material could also be improved by explicitly stating the materials needed for each phase of the task.
- Finally, while the proposal matches the target students in most areas, there is no specific mention of accommodations for students with special needs. Including some options or suggestions for differentiation to accommodate all learners could make your task even more inclusive.

In conclusion, your task represents a strong example of a TBLL activity that is engaging, relevant, and pedagogically sound. With some minor adjustments and more detailed planning, it could be an excellent resource for teaching and learning. Well done!



c) Feedback B

Note. The first screenshot (a) includes the prompt written by the author. The second screenshot (b) includes feedback A, which was generated by the summary of the nine evaluation criteria given step by step. The third screenshot (c) includes feedback B, which was generated by the summary of the nine evaluation criteria given in a single prompt.

As depicted in Figure 35, the author also asked ChatGPT in both prompt processes to reformulate its outputs into a single textual feedback (a). Feedback B (Figure 35c) seems to be more friendly and better written for student teacher audience than feedback A (Figure 35b), which appears to be stricter and composed of a list of bullet points.

4.5. Data analysis of RQ4

To establish whether ChatGPT is useful for educators in order to assess the activities created by the student teachers in terms of time (RQ4), the author generated two different types of data. The first comprehended 14 feedback texts created with the aid of ChatGPT (a comment for each TBLL activity) to upload to the Moodle platform of the course after the conclusion of the online activity. The aim was to provide student teachers with tailored feedback for each group.

The second type of data, instead, regarded one self-made feedback without the assistance of the AI tool on a TBLL chosen by the author. To elaborate the feedback texts with the help of ChatGPT for the student teachers, the author provided the AI instrument with the following two prompts (suggested by the course professor):

1) *Could you please assess the following didactic activity in terms of:*

- *coherence among objectives, tasks, and target;*
- *feasibility in terms of time;*
- *suitability in terms of learning environments chosen?*

2) *Could you please suggest improvements?*

In the case of ChatGPT-created feedback, the process of generation, revision, and modification of all feedback texts took about five hours, which corresponded to approximately 19 minutes for each assignment correction. Most of the time was dedicated to the reading of the feedback provided by ChatGPT and the final activity created by the student teachers, the revision to and correction of the feedback by the author. It is important to mention that the author is not a university professor and has little experience in student assessment. Therefore, the process of revision to the feedback of ChatGPT may not be really comprehensive.

Table 7

Mean and standard deviation of the feedback elaborated by ChatGPT and the corresponding revision and modification by the author

Mean of number of words	472
Standard deviation of number of words	101

As shown in Table 7, 14 feedback texts present a mean of 472 words, and a standard deviation of 101 words.

Table 8

Self-made feedback compared with ChatGPT's feedback

Feedback written by the author	Feedback elaborated by ChatGPT and reviewed by the author
<p>Dear students,</p> <p>Thank you so much for your work and participation in this online activity. You designed a good task-based language learning activity, but now let's assess it in-depth.</p> <p>Regarding the <i>coherence among</i>:</p> <p>1. <i>Objectives</i>: I really appreciate that you divide the content from the language learning objectives. You were really specific. The task follows the majority of your learning objectives. However, in my opinion, they seem to be too many. For this task you chose 6 learning objectives, and then you wrote another aim in the section <i>To Create a Portrait Gallery</i>. Does this aim belong to the learning objectives mentioned before? It is a bit confusing. Apart from this slight oversight, the objectives were described well and coherent with the following activities.</p> <p>2. <i>Task</i>: referring to the task, I noticed that there is the structure of Willi's framework and it is also nice that you created a pre-activity in class before entering the museum. Moreover, I really appreciated your didactic justifications on what to carry out in class and in the museum. Good job!</p> <p>3. <i>Target</i>: to me, you created a really good lesson for 16-18-year-old students with B1 level. You also used ICT (mobile phones, Kahoot), tried to foster their critical thinking skills, and metacognitive processes. I really liked the final activity where students had to take a selfie of themselves as the young generations of today took many selfies and used the technology, especially their mobile phones. So, well done!</p> <p>Concerning the <i>feasibility of time</i>, I realized that you did not mention all the times to conduct the activities. For example, in the "Classroom preparation to the task</p>	<p>1. <i>Coherence among objectives, tasks, and target</i>: The activity is well-structured and coherent in terms of objectives, tasks, and target. The objectives are clearly stated, and the tasks are designed to help students achieve these objectives. The topic of selfies and portraits is relevant and engaging for the target audience of B1 level students. The tasks in the classroom prepare the students for the task in the museum, which is the main focus of the activity. The language objectives are appropriate for the level of the students and are integrated into the task in a meaningful way. The follow-up activities help students reflect on what they have learned and apply it to their own lives.</p> <p>2. <i>Feasibility in terms of time</i>: The activity is feasible in terms of time, as it is divided into three parts, two of which take place in the classroom and one in the museum. The tasks in the classroom can be completed in a reasonable amount of time, and the task in the museum is well-structured and achievable within the allotted time. The follow-up activities can be spread over two consecutive lessons, allowing enough time for reflection and discussion.</p> <p>3. <i>Suitability in terms of learning environments chosen</i>: The learning environments chosen for the activity are appropriate. The classroom tasks allow students to build the necessary vocabulary and skills needed for the task in the museum. The museum visit provides an engaging and authentic learning environment where students can apply their knowledge and skills. The follow-up activities allow students to reflect on what they have learned and apply it to their own lives. The use of technology, such as Kahoot and smartphones, is</p>

in the museum”, it is not specified the time. Additionally, you decided 1 hour to carry out the task in the museum (which is good in order to avoid the museum fatigue), but you did not specify if it also included the language focus and the follow up in the classroom (you did not report the time of these two activity). In my opinion, I would do a pre-task: 10 minutes; task cycle: 60 minutes; language focus: 15 minutes; follow-up: 30/40 minutes (based on the number of students).

Regarding the *suitability of learning environments chosen*, I would say that it was very nice that you chose the National Portrait Gallery in London. I would specify whether the class is on a scholastic trip to visit the museum in person (as required by the instructions). I think that the activity you proposed in the museum are quite practical and you can carry them out for real, but I would mention the materials students had to use to create their own portrait and where they could find this material. In the end, I would specify that students have their own mobile phones, or the class is equipped with computers, or interactive whiteboards...

Now, I'll give you some *suggestions* that could improve your TBLL activity ☺

- When calculating times for activities, also take into consideration the “dead times”, for example: time wasted for creating the groups or pairs, moving the tables...

- Always put yourselves in students' shoes and ask yourself if you comprehend the instructions, task, etc. you wrote (you could also make your friend read your activity).

- Try to reduce the learning objectives and always stick to them.

- Provide students with linguistic support, for example, *how can my students discuss? How can my students give feedback to their classmates?* → in this way, they

also appropriate for the target audience and adds to the engagement and effectiveness of the activity.

Overall, the didactic activity is well-designed and suitable for promoting speaking skills and critical thinking among B1 level students in the topic of selfies and portraits.

There's a *suggestion* for improvement:

Museum environment: while the pre-task and task-cycle are planned to take place in the museum, it is important to consider the suitability of the learning environment for each task. For example, the Kahoot quiz could be completed prior to the museum visit to reduce noise and distractions, and the portrait creation task could be done in a quieter space outside the museum to allow for focused work.

<p>will feel surer and they will not be stuck in the conversation.</p> <p>- Language focus: I would reflect with my students on their oral presentations and maybe write on the board some sentences with mistakes that students had to correct.</p> <p>These are only simple suggestions ☺ I hope you find this feedback useful! For any doubt, I am here for you.</p>	
Total word count: 638	Total word count: 391

In order to verify whether there was a time saving with the aid of ChatGPT, the author wrote a self-made feedback text (without the assistance of the AI chatbot) for a TBLL activity. Even in writing the self-made feedback, the author answered the two aforementioned questions provided by the course professor.

As can be seen from Table 8, the self-made feedback text to the left includes 638 words and was elaborated in 95 minutes (including task reading and revision of the self-made feedback). This means that the feedback process would seem to be accelerated 5 times with the use of ChatGPT.

For the purpose of better understanding the completeness of the two feedback texts (Table 8), the assessment criteria used to evaluate the student teachers' task proposals were compared. Regarding the assessment of the coherence (which refers to the alignment among the learning objectives of the TBLL activity, the several activities created to reach those learning objectives, and the group of students to which activities and learning objectives are addressed) the self-made comment provided specific information about the task, learning objectives and target, whereas ChatGPT wrote a single paragraph containing all the information together.

Concerning the assessment of the appropriateness of the duration of activities designed by the student teachers, it was detected that in the self-made feedback, not all the activities included explicit time duration, which was a specific requirement to satisfy for the creation of the TBLL activity. Instead, ChatGPT's feedback did not mention this lack. Referring to the suitability of the learning environment chosen by the student teachers for their task, it was observed that in the self-made comment, the museum chosen was in London (while it was specifically required that the student teachers had to choose between a virtual tour in one of the Civic Museums of Venice, or a museum that they knew or that can be visited in person) and

the group did not specify if the class was in a study trip. Moreover, the student teacher group did not mention the organisation of the materials to be used by the class in the museum. Instead, ChatGPT provided only a general description highlighting the positive features. In the end, the self-made feedback includes more suggestions than the feedback written with ChatGPT.

However, in response to the question *How long would you take to write a feedback text similar to the ones generated by ChatGPT with and without the help of ChatGPT?*, the professor reported during the online interview that assessing a TBLL assignment in the “traditional” way (self-made feedback) normally takes her approximately 40 minutes. When she tried ChatGPT, it took her about 30 minutes to thoroughly read the AI-generated comment and verify its consistency with the characteristics of students’ tasks (and eventually remove or add some detail if needed).

4.6. Conclusion of the data analysis

In this chapter, the author carried out the analysis of relevant data for each research question. Concerning the first research question in terms of *planning time*, the author investigated the answers of student teachers included in both online surveys and interviews. From the data, it has been seen that 100% of student teachers believed that they saved time in designing the TBLL activity with the aid of ChatGPT. Moreover, the reasons and ways student teachers saved time have been analysed.

In terms of the *quality of the final product*, it has been noticed that most student teachers thought that they designed a better or same quality TBLL activity with the help of ChatGPT. In respect of this, an interviewee further explained how she gained personal confidence about the creation of activities for language learners as she had little teaching experience.

In terms of the *development of student teachers’ critical thinking*, the author analysed the comments on the utility and limits of the AI chatbot from the student teachers. The results show that 100% of student teachers considered ChatGPT useful for their teaching work, but at the same time, 41.9% of participants found some limitations. The author also compared the final assignments of the student teachers with their corresponding conversations with ChatGPT. Overall, these results indicate that almost all the final activities presented less than 40% of original words, six activities included above 45% of copied words, while eight activities contained less than 35% of copied words from the outputs of ChatGPT. However, the interviewees explained that their group studied the theory of TBLL activities based on Willis’

framework, and discussed what to include in their TBLL activity before starting using ChatGPT. Additionally, the group reflected together on each output of ChatGPT to comprehend whether its suggestions could be effective for the target and the activity. An interviewee also commented that her group relied on the outputs of ChatGPT because they were satisfied with ChatGPT answers, and the group understood that the final activity had to be generated based on the ChatGPT outputs.

For the second research question, the prompts generated by the student teachers were analysed and categorised. For each category, some examples were provided. Furthermore, some appropriate and inappropriate practices of interactions were noted and explained.

Regarding the third research question, the author used the indications provided by the professor during the online interview to test ChatGPT's assessment capabilities. In order to do that, the author designed a rubric with the help of ChatGPT and carried out a self-study methodology through prompt engineering.

As to the fourth research question, the needed time for assessing final TBLL activities was examined by comparing the time dedicated to generating feedback with ChatGPT and its revision with a self-made feedback text written by the author. The outcomes show that there was a five acceleration in time with the help of ChatGPT. In addition, some insights about the lengths of the feedback as well as some comments on the correctness of ChatGPT were provided. The next chapter moves on to discuss the analysis of these results.

CHAPTER V. DISCUSSION

In the present chapter, the data analysed in Chapter IV will be discussed by considering the existing literature. The discussion of the results will be divided into the four RQs in order to facilitate the comprehension of the readers. Moreover, limitations to the present research study will be provided.

5.1. RQ1: Usefulness of ChatGPT (time, quality of activities and critical thinking)

The first point of the RQ1 sought to determine whether the use of ChatGPT was useful for student teachers in order to design classroom activities for language learners in terms of *planning time*. The results of the present investigation strongly indicate that ChatGPT was considered a time-saving AI technology by the student teachers. In particular, all participants of the study agreed to the fact that they optimised their time to create their own assignments thanks to the use of ChatGPT (Figure 17a). For this reason, the first point of the first hypothesis (*the usage of ChatGPT in designing classroom activities for language learners makes student teachers spend less time in comparison to creating activities without employing ChatGPT*) was fully confirmed. Furthermore, it is interesting to note that the student teachers could save time not only in one concrete way, but they were also able to identify several manners of gaining time, such as creation of exercises, texts and questions, brainstorming ideas, etc. (Figure 17b). The present findings corroborate the ideas of Kasneci et al. (2023), who suggested that ChatGPT could be used in the role of teacher assistant in order to save time for lesson plans or the creation of materials.

The second part of the RQ1 was designed to detect whether the use of ChatGPT was useful for student teachers in order to design classroom activities for language learners in terms of *quality of the final products*. The related findings suggest that the majority of the student teachers perceived to have designed a better or same quality activity with the aid of ChatGPT (Figure 18). Moreover, all student teachers considered ChatGPT a useful AI assistant capable of helping them during the organisation and creation of task proposals (Figure 19; Table 2). Furthermore, the results obtained in the online interviews suggest that thanks to the aid of ChatGPT, the student teachers could come up with original ideas for their own lessons and increase their personal feeling of self-confidence in their teaching skills. This thesis is also confirmed by Kasneci et al. (2023), who stated that using ChatGPT could be positive for

students as they have the opportunity to take inspiration and learn from ChatGPT's outputs. It can be, therefore, stated that the second point of the first hypothesis (*the employment of ChatGPT positively impacts the quality of the final products created by student teachers*) was confirmed, at least from the perspective of the student teachers. In order to fully verify the hypothesis, a comparison should be made between the ChatGPT tasks produced by the student teachers for this study and the tasks produced by the student teachers of precedent courses.

The third point of the RQ1 concerns the extent to which the use of ChatGPT was useful for student teachers in order to design classroom activities for language learners in terms of the *development of critical thinking*. The results show that only few student teachers made explicit that it is important to analyse and verify ChatGPT's outputs accurately and not to rely upon the AI tool completely (comment on Table 2; Table 3). Besides, only few participants mentioned their dissatisfaction with the answers provided by ChatGPT, and the majority could not find any difficulties or limitations in the AI tool. However, the present results must be interpreted with caution because even if most student teachers did not find any difficulties or limitations, it might not be claimed that those participants did not use their own critical thinking ability. It is possible that ChatGPT provided the student teachers with good outputs or the question formulated by the author in the survey was not so clear. After seeing the results, it would have been better to give a question about limitations and another query concerning the difficulties of using ChatGPT.

It is also interesting to notice that two student teachers with prior teaching experience mentioned that the outputs provided by ChatGPT were accurate. Moreover, another student teacher with teaching experience specified that ChatGPT could be a useful instrument for those who have little experience in teaching. These findings might be considered negative as well as dangerous: in case the student teachers considered ChatGPT as a high-performance AI instrument, they could be exposed to the danger of relying on ChatGPT not only for the present assignment (TBLL activity) but also for future assignments. In fact, depending entirely on ChatGPT would be a mistake for several reasons. First, ChatGPT has been released lately and all its potential, limits, and impacts are still unknown (Jalil et al., 2023). Second, it is crucial to employ ChatGPT with human judgment and analysis, as confirmed by the researchers Rudolph et al. (2023) as well as Cooper (2023).

From the results (Table 3), it also emerges that it would be necessary to organise a training course for student teachers on using ChatGPT, in which its limitations are clearly

specified. In this way, the course may help to foster student teachers' critical thinking abilities. This idea is also in line with some prior studies, for instance, the one carried out by Kasneci together with other researchers (2023), in which ongoing training both for teachers and students should be promoted, aiming at keeping up to date with the latest developments of ChatGPT.

Concerning the interactions with ChatGPT and TBLL activities, the results show that a few TBLL tasks have been designed with a small percentage of copied words (14%). This seems to be positive evidence of using critical thinking skills on behalf of student teachers. In fact, it signifies that some participants revised and modified the outputs provided by ChatGPT completely. Whereas a small number of TBLL activities include a higher percentage of copied words (35%). This may indicate that ChatGPT gave some good ideas for the activity, and therefore, the student teachers decided to copy them. In this case, it can be said that student teachers may have reflected on the ideas proposed by ChatGPT and decided to keep them based on their personal criteria and knowledge. However, some parts seem to be just copied and don't seem to give special attention to details (Figure 20). The rest of TBLL activities present more than 45% of copied words, which corresponds to a negative outcome. It is complicated to interpret this result as a consequence of using ChatGPT because other factors should be taken into account. Actually, the high percentage might be related to the fact that some student teachers did not comprehend the professor and author's instructions, and for this reason, they copied the whole conversation (including prompts) in their final proposal. Another possible cause might be the assessment method regarding the course Instructional Design for Language Education. As explained in section 3.1, throughout the academic semester there are seven online lessons in which students have the opportunity to carry out seven activities and gain 0.3 points for each activity done to add to the final exam. Generally, the student teachers receive 0.3 points with only their participation in the online lesson through the creation of the activity (except in extreme cases of very badly performed activity). This may mean that other participants designed the TBLL task only to obtain 0.3 points to add to the final mark without much interest in producing a high-quality product. In this regard, this idea could be supported by checking the high variability in the length of the TBLL activity (Table 4).

On the assumption that university students, who will probably be future teachers or language experts, fulfil the university assignment only to gain the points to sum to the final mark, it is quite disappointing. Master's degree students should have developed an intrinsic motivation, which does not depend on a numerical digit, and they should work hard to improve

themselves. The fact to have some online lessons in which student teachers can practice what they have learnt in class should be seen as a great opportunity for their learning. However, these findings may conceal significant directions to follow in the future, which were also reported in the investigation carried out by Qureshi (2023). The scholar pointed out that it is crucial to sensitise university students to utterly comprehend the value of their learning path, the meaning as well as consequences of cheating and plagiarism. Moreover, these results confirmed the thesis by the researchers Moqbel and Al-kadi (2023), Timmis et al. (2015), who stated that nowadays, it is extremely necessary to rethink innovative approaches to assess students' outcomes in order to avoid plagiarism or cheating. In fact, according to Zhai (2022), and Su et al. (2023), professors should give the students assignments which require a higher-order critical skills, creativity, and logical reasoning to accomplish the task.

5.2. RQ2: Strategies used by student teachers when interacting with ChatGPT

The RQ2 sought to explore the strategies employed by student teachers when interacting with ChatGPT to create a TBLL activity. The student teachers were not trained to use ChatGPT for the research project as the author wanted to investigate the types of interactions the participants developed on their own and possible innovative and creative ways of interacting with ChatGPT. In the video presentation about ChatGPT (see Chapter III), the author only provided the participants with some examples and suggested that student teachers had to write the prompt step by step. Moreover, it is important to take into account that most participants used ChatGPT for the first time during the research project (Figure 15).

The corresponding results confirm that the student teachers were able to come up with many different ways to interact with ChatGPT, such as asking for general information, creating the material, brainstorming ideas for the activity, and elaborating learning objectives (only to name a few). The findings indicate that most student teachers asked ChatGPT for precise elements, but almost a third of the participants tried to request the whole task in just one prompt (Table 5). However, the results also show that the student teachers did not find or use any original ways to interact with ChatGPT.

Regarding the appropriate and inappropriate practices for writing prompts (Table 6), the results demonstrate that almost half of the student teachers omitted important information about the task in the prompts. Instead, only a low percentage of the participants helped ChatGPT with more informative prompts. According to Jiao et al. (2023), the quality as well style of the

prompts could influence the output of ChatGPT. For this reason, the more the student teachers were precise in the prompt, the more ChatGPT should have been accurate in its output.

Moreover, it is observed that some student teachers asked the AI chatbot to create tables, provide links to website resources, etc., which represented the limits of ChatGPT (at the moment of the experiment, April 2023). From these results, it is comprehensible that the student teachers need a training course on how to interact with ChatGPT properly and understand its limitations as fast as possible. Learning how to utilise ChatGPT may aid students as well as teachers in realising that interaction with this AI tool usually presents different phases. In the beginning phase, ChatGPT seems to be interesting and spark excitement. Subsequently, once human beings get used to employing ChatGPT, individuals comprehend that it is essential to revise ChatGPT's outputs because the AI chatbot can sometimes commit errors. In fact, as stated in the *GPT-4 Technical report* by OpenAI (2023), these mistakes are commonly named *hallucinations*, which represent untruthful contents related to determined sources.

5.3. RQ3: Ways to assess the student teachers' activities with ChatGPT

The third research question aimed to explore the ways professors can assess the activities created by the student teachers through the use of ChatGPT. The author decided to create a rubric based on the information yielded through the online interview with the professor. For this reason, the author tried to write down a prompt as specific as possible by following the guidelines of the online course carried out by OpenAI and DeepLearning.AI (Fulford & Ng, 2023) and the research paper written by White et al. (2023) (Figure 22a). The results obtained are quite disappointing as ChatGPT completely fails the rubric. In fact, the first time ChatGPT was not able to create a sophisticated rubric to assess the TBLL activities. For instance, in *Pertinence with respect to activity requirements*, ChatGPT simply wrote: "Task doesn't meet..."; "Task meets some but not all..."; "Task meets most..."; "Task completely meets..." (Figure 22b). As can be noticed, the AI chatbot only rephrased the evaluation criteria without being detailed as well as precise. After asking ChatGPT to be more precise, the evaluation criteria seem to be designed a little better but still did not include specific examples for each section (Figure 22d).

Overall, the findings suggest that ChatGPT may not be suitable for designing rubrics to assess student teachers' TBLL activities as the AI chatbot was not able to create a detailed rubric. It is crucial to mention that when designing a rubric to evaluate students' work, many

elements have to be taken into consideration. However, the rubrics created by ChatGPT could be used as hints, evaluated and adapted by the professors. For this reason, ChatGPT may always support the work of the teachers, but it could not substitute them in the assessment practices. Similar results were acknowledged by Kasneci et al. (2023) as well as Chan and Tsi (2023).

Moreover, it could be important to mention that the rubric could have been created in a more detailed way if each evaluation criterion had been addressed one at a time. The purpose would have been to verify whether ChatGPT had been capable of designing a better rubric. Creating the rubric step by step by giving ChatGPT one evaluation criterion at a time as well as all informative documents, such as the CEFR of B1 level, or the explanation of the TBLL activity, the rubric could have been much more detailed. However, the difference in saving time between the ChatGPT-assisted rubric and a self-made rubric would not probably have been so significant.

Concerning the interactions with ChatGPT aiming at assessing a TBLL activity (Figures 26-34), the results show that ChatGPT is more specific and detailed when the prompts are given step by step rather than when it is provided with all criteria in a single input. Overall, the findings suggest that both versions (giving one criterion at a time, and all criteria in a single prompt) use the rubric quite properly and precisely to assess the TBLL task designed by student teachers. Therefore, it could be said that these findings seem to be different from the ones provided by the guidelines of the online course (Fulford & Ng, 2023), which taught to give ChatGPT all the instructions in a single prompt. Conversely, the results obtained show that giving ChatGPT an input full of information does not seem to pinpoint where the student teachers make mistakes. From the results, it can also be stated that ChatGPT may be quite trustworthy since it has not invented features (except for a slight error in feedback A, when giving one criterion at a time). Additionally, it seems a good self-assessment tool for students as it is able to spot errors, provide examples, and explain the reasons why the LLM thinks in that way. Even the scholars Su et al. (2023), who tasted ChatGPT as an AWE system in their study, sustained that the AI chatbot seemed to be a helpful tool to proofread texts as well as provide feedback. It is crucial, however, to be able to write a detailed prompt to obtain a more detailed output from ChatGPT. In fact, the AI chatbot could have been precise thanks to the specific input formulated by the author, who spent a little time and did some trials to write a good prompt.

Although it seems that the better outputs of ChatGPT are those provided by giving one criterion at a time, the findings about the feedback given by the two prompt versions, one criterion at a time (Feedback A) versus all criteria in a single input (Feedback B), illustrate that feedback B seems to be more friendly and better written for the audience of student teachers than feedback A, which appears to be stricter and composed of a list of bullet points. Moreover, feedback A stated that the TBLL activity of the student teachers is in line with the national curriculum, which is contrary to what was mentioned in a previous output (Figure 35b). This finding is very interesting because it could indicate that ChatGPT has some issues about summarising the information correctly, when the input context is formed by many long messages.

Despite the aforementioned mistakes by ChatGPT, both feedback texts (Figures 35b and 35c) look to be of good quality and may be useful as self-assessment for student teachers. In case educators want to use ChatGPT to generate feedback, Cooper (2023) suggest that it is important to always do a critical revision to any sources produced by ChatGPT, as the AI tool may commit errors.

5.4. RQ4: Usefulness of ChatGPT in assessing the activities in terms of planning time

The fourth research question wanted to establish whether ChatGPT was useful for educators in order to assess the activities created by the student teachers in terms of *time*. The results obtained confirm that the teachers may save plenty of time in writing feedback with the help of ChatGPT. In fact, thanks to the aid of the AI chatbot, the author was able to create feedback for the group of student teachers five times faster than the self-made comment. However, the self-made feedback is slightly longer than the average ChatGPT's feedback (Table 8).

Moreover, the results show that the feedback created with ChatGPT seems to be really literal, objective, and adhered to the input, for example, it does not include compliments. While the self-made feedback seems to be more amicable and constructive. The present difference was also found in the study conducted by Guo et al. (2023), who outlined some dissimilarities between humans and ChatGPT. In fact, the scholars discovered that ChatGPT produces more descriptive, formal, and objective texts without conveying its feelings, while people are capable of elaborating more colloquial texts expressing emotions.

In conclusion, the hypothesis (*The usage of ChatGPT in assessing activities of students helps university professors spend less time in comparison to assessing activities without*

employing ChatGPT) has been confirmed, but it would have been necessary to conduct a deeper investigation into the quality of the feedback. As the author has little experience in student assessment, she did not prefer to examine this aspect. Furthermore, it was not possible to perform a comprehensive analysis of the quality of the feedback contents with a group of specialists in the assessment field due to time constraints.

5.5. Conclusion of the discussion of the results

Overall, the findings of this exploratory study confirm that the student teachers could save time by creating the TBLL activity with the help of ChatGPT (RQ1). Moreover, most participants stated that they were able to design a better or same quality activity with the aid of ChatGPT, and all student teachers considered ChatGPT a useful AI assistant capable of helping them during the organisation and creation of the task proposals (RQ1). Concerning the TBLL tasks, it is significant to highlight that many student teachers were not able to find the limitations of ChatGPT, and about half of the participants copied the outputs given by the AI tool, which represented a negative outcome (RQ1). However, it is hard to blame ChatGPT since there were other factors involved: some student teachers may not have comprehended the assignment instructions, and the assessment method for the online activity may not have encouraged the design of a well-performed task.

Regarding the strategies used by the student teachers to interact with ChatGPT, the participants were able to come up with many ways to interact with the AI chatbot. However, the student teachers did not find or use any innovative ways to communicate with the AI chatbot (RQ2). Another point to consider included the difficulties the student teachers found while interacting with ChatGPT as well as the incapacity to interact with the AI instrument.

Hence, one possible solution to tackle the issues discovered in RQ1 and RQ2 could be to organise a training course for student teachers, which would include several learning goals. First, the course about ChatGPT would allow the student teachers to discover and comprehend the limitations of the AI chatbot as well as raise their awareness of its constant evolution. This means that OpenAI can change the webpage features or the AI behind ChatGPT at any time. Second, the course would teach the student teachers how to interact with ChatGPT properly. Third, the training would sensitise the participants to avoid plagiarism by reflecting on the reasons behind cheating, the importance of students' practice and future employment, and the benefits of acquiring new knowledge and skills. A further significant solution to solve the

above-mentioned problems could be to rethink the assessment methods and modify the types of assignment by demanding more critical thinking, reasoning, problem-solving as well as creativity from the students.

As far as the ways of assessment are concerned, ChatGPT was not able to create a detailed rubric to assess the TBLL activities. However, it could be used as a hint by educators. It was also noticed that when giving the rubric to ChatGPT, the AI tool was capable of writing a valid feedback (RQ3). By providing ChatGPT with the evaluation criteria one by one, it seemed to be more detailed, although it could not summarise all evaluation criteria very well. Whereas ChatGPT acted in the opposite way when giving it the whole rubric in a single prompt: the AI tool was less detailed when assessing all the evaluation criteria, but it was capable of writing a better feedback text for the student teacher audience. This finding could be interesting for the AI research community because it may indicate that ChatGPT has some issues about summarising the information correctly when the input context is formed by many long messages.

Referring to the feedback texts, the results showed that the feedback written by ChatGPT and revised by the author allowed to save more time than writing a self-made comment (RQ4). However, the feedback written with the help of ChatGPT might not be very precise, detailed as well as constructive.

5.6. Limitations to the research study

The present exploratory study includes a number of significant limitations to be considered. The results obtained from the research cannot be generalised to the wider population as the experiment was conducted with a very specific sample in a very specific context. However, the study may be equally useful given that scholars of different research contexts could draw inspiration from this exploratory study in order to produce further innovative ideas or hypotheses (Swedberg, 2020).

One of the main limitations may be found in using ChatGPT-API instead of “standard” ChatGPT, which was deactivated by OpenAI due to the Italian Garante’s concerns during the data collection period (see Chapter III). The ChatGPT-API parameters may not be equal to the ones of the “standard” ChatGPT. For this reason, the outputs of the webpage may have been slightly different. In spite of this, thanks to this web, the whole student teachers’ interactions with ChatGPT were collected avoiding possible bias by the participants.

A further limitation may be related to the quality of the TBLL proposals: at the beginning phase of the research project, there was the idea of comparing the quality of the TBLL activities of this academic year with the quality of the TBLL proposals of the previous years from the perspective of the course professor. In this way, it could be interesting to establish whether the help of ChatGPT enhanced the quality of students' work. However, due to differences in the assignment instructions and the participants involved, a whole distinct research project would have been necessary.

Another limit encountered during the study may be the formulation of some questions included in the online survey. For instance, the fact that student teachers did not encounter any difficulties or limitations could be due to the lack of clarity of the question formulated by the author. For this reason, it would have been better to split the question and give one question about limitations and another one concerning the difficulties of using ChatGPT. A further important limit was time constraints: it would have been interesting to involve experts in the field of assessment in the analysis of the rubrics and feedback texts in order to conduct a deeper investigation into the assessment, avoiding possible subjective bias.

CONCLUSION

During the last years, the spread of AI in language education has changed the approach to learning and teaching, pioneering further horizons and leading to new challenges for the academic community. Given the modern and ever-changing nature of this sector, the present exploratory study aimed at contributing to the research field of AI applied to language and teaching education. Specifically, the investigation tried to shed light on the usefulness of ChatGPT in designing and assessing TBLL activities addressed to language learners, explored how Master's degree students, who may become future teachers, interact with ChatGPT, and examined the impact of ChatGPT on critical thinking skills.

Concerning the literature review, the realities of AI and ChatGPT in the context of teaching and learning settings have been explored in depth. However, being ChatGPT a new AI instrument, several research gaps have been found. In fact, only a few investigations have been conducted into how teachers can employ ChatGPT as well as how learners interact with the AI tool. Moreover, little has been discovered about how ChatGPT impacts students' critical thinking skills. Taking into consideration the research gaps and the literature review, four research questions were set:

RQ1: To what extent is the use of ChatGPT useful for student teachers in order to design classroom activities for language learners, in terms of planning time, quality of the final products, and development of their critical thinking?

RQ2: What are the strategies used by student teachers when interacting with ChatGPT to create a task-based language learning activity?

RQ3: How can university professors assess the activities created by their students through the use of ChatGPT?

RQ4: To what extent is the use of ChatGPT useful for student teacher educators in order to assess the activities created by their students in terms of planning time?

In order to attempt to address the aforementioned research questions, a group of student teachers were asked to design a TBLL activity with the aid of ChatGPT during an online lesson within a course in language teaching methodology. Both quantitative and qualitative data were gathered from student teacher online survey, student teacher online interviews, student teachers' interactions with ChatGPT, student teachers' final TBLL proposals, professor online

interview, ChatGPT-generated feedback, author-generated feedback as well as author prompt engineering.

The results confirmed that the student teachers could save time and create a same or better quality TBLL activity with the aid of ChatGPT. The participants also deemed the AI tool as an effective assistant to use in the organisation and creation of the TBLL activities. The findings also demonstrated that many participants were not capable of encountering any limitations to ChatGPT, and about half of the student teachers copied the outputs given by ChatGPT, which represented a negative outcome. However, it is difficult to only accuse the use of ChatGPT as there were other aspects involved: some participants may not have understood the instructions of the assignment, and the assessment practice employed for the online activity may not have pushed the student teachers to create a well-performed activity. Furthermore, the results highlighted that the student teachers were able to come up with many strategies to interact with ChatGPT, even if the majority did not use ChatGPT before the experiment. The findings also revealed that the participants encountered some difficulties while interacting with ChatGPT.

One effective solution to mitigate the issues found could be to organise a training course for student teachers aiming at discovering and comprehending the limitations of ChatGPT, raising awareness of ChatGPT constant evolution, interacting properly, and sensitising the participants to avoid cheating and plagiarism. Another practical solution could be to modify the assessment methods and give assignments that necessitate more problem-solving, reasoning, creativity as well as critical thinking skills to be solved.

Moreover, the results indicated that ChatGPT could not create an accurate rubric to assess the TBLL proposals, but it could serve as a useful starting point for teachers. The outcomes also showed that generally, ChatGPT could write valid feedback when giving it the rubric. Precisely, ChatGPT seemed to be more detailed in the assessment when providing it with the evaluation criteria one by one, but then it was not able to summarise all evaluation criteria very well for the feedback. Instead, when giving the AI chatbot the whole rubric in a single prompt, ChatGPT acted in the opposite way. In fact, ChatGPT was less detailed when assessing all the evaluation criteria, but it was able to elaborate a better feedback text for the student teacher audience.

In the end, the findings demonstrated that the feedback elaborated by ChatGPT and then revised by the author permitted to save more time than elaborating a self-made comment.

However, the ChatGPT-generated feedback might not be really detailed, constructive and precise.

Despite its limits, the exploratory study carried out an extensive investigation into ChatGPT. These findings may provide plenty of new opportunities for this research area, which is likely to acquire more and more significance in the future.

Proposals for future research

As for future research proposals, more investigations into how ChatGPT can affect the creativity, problem-solving as well as critical thinking skills of primary and secondary students in the realm of language learning are much required. A further research idea might be to conduct empirical studies on the effectiveness of this AI chatbot in supporting learner autonomy in foreign language education. Lastly, exploring the students' perceptions of the effectiveness and usefulness of ChatGPT automated assessment would be necessary to further improve the assessment practices and the prompts to insert into ChatGPT.

REFERENCES

- Alam, A. (2021). Possibilities and Apprehensions in the Landscape of Artificial Intelligence in Education. *International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 1-8.
- Ali, J., Shamsan, M., Hezam, T., & Mohammed, A. (2023). Impact of ChatGPT on Learning Motivation: Teachers and Students' Voices. *Journal of English Studies in Arabia Felix*, 2(1), 41-49. DOI:10.56540/jesaf.v2i1.51.
- Allen, D., & Tanner, K. (2006). Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learners. *CBE—Life Sciences Education*, 5, 197–203. DOI: 10.1187/cbe.06–06–0168.
- Alshater, M. (2022). Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT. *Social Science Research Network (SSRN)*, <http://dx.doi.org/10.2139/ssrn.4312358>.
- Anderson, D., & McNeill, G. (1992). Artificial neural networks technology. *Kaman Sciences Corporation*.
- Anil, R., Dai, A., Firat, O., Johnson, M., Lepikhin, D., Passos, A., Shakeri, S., Taropa, E., & Wu, Y. (2023). PaLM 2 Technical Report. *arXiv preprint arXiv:2305.10403*.
- Askell, A., Brundage, M., & Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. *OpenAI*. *arXiv:1907.04534*, 1-23. <https://doi.org/10.48550/arXiv.1907.04534>.
- Augustine, R., & Xavier, S. (2021). A Critical Study On Netflix Docudrama ‘The Social Dilemma’. *International Journal of Research Culture Society*, 1-10.
- Baqeel, H., & Saeed, S. (2019). Face detection authentication on Smartphones: End Users Usability Assessment Experiences. *IEEE*, 1-6.
- Bayer, L. (2023, February 28). *Introducing Q-Chat, the world’s first AI tutor built with OpenAI’s ChatGPT*. Retrieved June 2023, from Quizlet Webpage: <https://quizlet.com/blog/meet-q-chat>
- Bellaera, L., Weinstein-Jones, Y., Ilie, S., & Baker, S. (2021). Critical thinking in practice: The priorities and practices of instructors teaching in higher education. *Thinking Skills and Creativity*, 41, 1-16. <https://doi.org/10.1016/j.tsc.2021.100856>.

- Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P. “., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., & Petrov, M. (2021). Dota 2 with Large Scale Deep Reinforcement Learning. *arXiv preprint arXiv:1912.06680v1*.
- Bini, S. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty*, 2358-2361.
- Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology*, 15(3), 1-10. <https://doi.org/10.30935/cedtech/13176>.
- Breck, E., Polyzotis, N., Roy, S., Whang, S., & Zinkevich, M. (2019). Data Validation For Machine Learning. *SysML*.
- Brockman, G., Murati, M., Welinder, P., & OpenAI. (2020, June 11). *OpenAI API: We're releasing an API for accessing new AI models developed by OpenAI*. Retrieved June 2023, from OpenAI Webpage: <https://openai.com/blog/openai-api>
- Brockman, G., Sutskever, I., & OpenAI. (2015, December 11). *Introducing OpenAI*. Retrieved June 2023, from OpenAI Webpage: <https://openai.com/blog/introducing-openai>
- Brown, D. H., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices*. Pearson.
- Brown, D. H., & Lee, H. (2015). Chapter 21: Classroom-Based Assessment. In *Teaching by principles : an interactive approach to language pedagogy* (pp. 513-538). Pearson Education.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Amodei, D. (2020). Language Models are Few-Shot Learners. *OpenAI*, <https://doi.org/10.48550/arXiv.2005.14165>.
- Campbell, M., Hoane Jr, A., & Hsu, F.-h. (2002). Deep Blue. *Artificial Intelligence*, 134, 57-83.
- Cardona, M., Rodríguez, R., & Ishmael, K. (2023). Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations. *U.S. Department of Education, Office of Educational Technology*.
- Carretero, S., Vuorikari, R., & Punie, Y. (2017). DigComp 2.1: The digital competence framework for citizens. *Joint Research Centre (European Commission)*.

- Case, M. (2015). Machine Translation and the Disruption of Foreign Language Learning Activities. *eLearning Papers*(45), 4-16.
- Chan, C., & Tsi, L. (2023). The AI Revolution in Education: Will AI Replace or Assist Teachers in Higher Education? arXiv:2305.01185.
- Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, 136, 16-24.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *Ieee Access*, 8, 75264-75278. DOI: 10.1109/ACCESS.2020.2988510.
- Chen, L., Zaharia, M., & Zou, J. (2023, July 18). How Is ChatGPT's Behavior Changing over Time? *arXiv:2307.09009v1*, 1-8.
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1.
- Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H., Sutton, C., Gehrmann, S., & Fiedel, N. (2022). PaLM: Scaling Language Modeling with Pathways. *arXiv pre-print arXiv:2204.02311v5*.
- Christiano, P., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep Reinforcement Learning from Human Preferences. *31st Conference on Neural Information Processing Systems (NIPS)*, 1-9. Retrieved from: https://proceedings.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf.
- Cloete, A. (2017). Technology and education: Challenges and opportunities. *HTS Teologiese Studies/ Theological Studies*, 73(4), 1-7. <https://doi.org/10.4102/hts.v73i4.4589>.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (Vol. Sixth edition). London New York: Routledge.
- Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, 32(3), 444-452. <https://doi.org/10.1007/s10956-023-10039-y>.
- Coppin, B. (2004). *Artificial Intelligence Illuminated*. Jones and Bartlett Publishers.
- Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of*

- University Teaching & Learning Practice*, 20(3), 1-19.
<https://doi.org/10.53761/1.20.3.02>.
- Creswell, J., & Creswell, J. (2018). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches. Fifth Edition*. SAGE Publications.
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies*. Oxford University Press.
- Dörnyei, Z., & Csizér, K. (2012). How to Design and Analyze Surveys in Second Language Acquisition Research. In A. Mackey, & S. M. Gass, *Research Methods in Second Language Acquisition: A Practical Guide* (pp. 74-94.
<https://doi.org/10.1002/9781444347340.ch5>). Blackwell Publishing Ltd.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805v2*.
- Du Boulay, B. (2016). Artificial Intelligence as an Effective Classroom Assistant. *IEEE Intelligent Systems*, 31(6), 76-81.
- Duolingo. (2023, March 14). *Introducing Duolingo Max, a learning experience powered by GPT-4*. Retrieved June 2023, from Duolingo Webpage:
<https://blog.duolingo.com/duolingo-max/>
- European Parliament. (2023, June 14). *Artificial Intelligence Act*. Retrieved September 2023, from https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf
- Evans, M. (2017). Analysing Qualitative Data. In E. Wilson, *School-based Research: A guide for education students*. Sage publications.
- Fitria, T. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *Journal of English Language Teaching*, 12(1), 44-58.
- Fulford, I., & Ng, A. (2023, July). *ChatGPT Prompt Engineering for Developers*. DeepLearning.AI: <https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/>
- Garante per la Protezione dei Dati Personali. (2023, March 30). *Provvedimento del 30 marzo 2023 [9870832]*. Retrieved July 2023, from [garanteprivacy.it: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832](https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832)

- Garcia, I., & Pena, M. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, 24(5), 471–487. DOI: 10.1080/09588221.2011.582687.
- García-Peñalvo, F. (2023). The perception of Artificial Intelligence in educational contexts after the launch of ChatGPT: Disruption or Panic? *Education in the Knowledge Society*(24), <https://doi.org/10.14201/eks.31279>.
- Garnelo, M., & Shanahan, M. (2019). Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences*, 29, 17-23. <https://doi.org/10.1016/j.cobeha.2018.12.010>.
- Goksel, N., & Bozkurt, A. (2019). Artificial Intelligence in Education: Current Insights and Future Perspectives. In S. Sisman-Ugur, & G. Kurubacak, *Handbook of Research on Learning in the Age of Transhumanism* (pp. 224-236). Hershey, PA: IGI Global.
- Grassini, S. (2023). Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences (Educ. Sci.)*, 13(692), 1-13. <https://doi.org/10.3390/educsci13070692>.
- Green, J., & Thorogood, N. (2004). *Qualitative Methods for Health Research*. SAGE Publications.
- Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. *arXiv:2301.07597v1*.
- Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5-14.
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), 1-11. <https://doi.org/10.30935/cedtech/13036>.
- Hamilton, M., & Pinnegar, S. (2000). On the Threshold of a New Century: Trustworthiness, Integrity, and Self-study in Teacher Education. *Journal of Teacher Education*, 51(3), 234-240. <https://doi.org/10.1177/0022487100051003012>.
- Hamilton, M., Smith, L., & Worthington, K. (2009). Fitting the Methodology with the Research: An exploration of narrative, self-study and auto- ethnography. *Studying Teacher Education*, 17-28. DOI: 10.1080/17425960801976321.

- Haristiani, N., Dewanty, V., & Rifai, M. (2022). Autonomous Learning Through Chatbot-based Application Utilization to Enhance Basic Japanese Competence of Vocational High School Students. *Journal of Technical Education and Training*, 14(2), 143-155. <https://doi.org/10.30880/jtet.2022.14.02.013>.
- Herrmann, H. (2022). The arcanum of artificial intelligence in enterprise applications: Toward a unified framework. *Journal of Engineering and Technology Management*, 66.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence In Education. Promises and Implications for Teaching and Learning*. Boston: Center for Curriculum Redesign.
- Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Council of Europe.
- Huang, A., Lu, O., & Yang, S. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners’ learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194, 1-12. <https://doi.org/10.1016/j.compedu.2022.104684>.
- Hwang, G.-J., & Chien, S.-Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, 3.
- Jalil, S., Rafi, S., LaToza, T., Moran, K., & Lam, W. (2023). ChatGPT and Software Testing Education: Promises & Perils. *2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, 4130-4137. DOI: 10.1109/ICSTW58534.2023.00078.
- Jiao, W., Wang, W., Huang, J.-t., Wang, X., & Tu, Z. (2023). Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine. *arXiv preprint arXiv:2301.08745*, https://www.researchgate.net/profile/Wenxiang-Jiao/publication/367359399_Is_ChatGPT_A_Good_Translator_A_Preliminary_Study/links/6416f968a1b72772e40f9f07/Is-ChatGPT-A-Good-Translator-A-Preliminary-Study.pdf.
- Jones, N. (2018, September 13). How to stop data centres from gobbling up the world’s electricity. *Nature*, 561(7722), pp. 163-166.
- Kannan, J., & Munday, P. (2018). New Trends in Second Language Learning and Teaching through the lens of ICT, Networked Learning, and Artificial Intelligence. *Círculo de*

- Lingüística Aplicada a la Comunicación*, 76, 13-30.
<http://dx.doi.org/10.5209/CLAC.62495>.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 1-9. <https://doi.org/10.1016/j.lindif.2023.102274>.
- Kessler, G. (2018). Technology and the future of language teaching. *Foreign Language Annals*, 51(1), 205-218. <https://doi.org/10.1111/flan.12318>.
- Kilcher, Y. (2023, April 15). *OpenAssistant RELEASED! The world's best open-source Chat AI!* Retrieved June 2023, from Open Assistant Blog: <https://projects.laion.ai/Open-Assistant/blog/2023/04/15/open-assistant-released>
- Kohnke, L. (2022). A Pedagogical Chatbot: A Supplemental Language Learning Tool. *RELC Journal*, 1-11. DOI: 10.1177/00336882211067054.
- Kohnke, L., Moorhouse, B., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, 1-14. DOI: 10.1177/00336882231162868.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6), 84-90.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11).
- Lee, S.-M. (2019). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 33(3), 157–175. DOI: 10.1080/09588221.2018.1553186.
- Lew, M., & Nelson, R. (2016). New Teachers' Challenges: How Culturally Responsive Teaching, Classroom Management, & Assessment Literacy Are Intertwined. *Multicultural Education*, 23, 7-13.
- Liang, J.-C., Hwang, G.-J., Chen, M.-R., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, 31(7), 4270–4296. DOI: 10.1080/10494820.2021.1958348.
- Liu, B., Morales, D., Roser Chinchilla, J., Sabzalieva, E., Valentini, A., Vieira do Nascimento, D., & Yerovi, C. (2023). Harnessing the Era of Artificial Intelligence in Higher

- Education: A Primer for Higher Education Stakeholders. *United Nations Educational, Scientific and Cultural Organization (UNESCO)*, Retrieved from: <https://unesdoc.unesco.org/ark:/48223/pf0000386670>.
- Liu, O., Frankel, L., & Roohr, K. (2014). Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. *ETS Research Report Series*, 1-23. DOI:10.1002/ets2.12009.
- Lowe, R., & Leike, J. (2022, January 27). *Aligning language models to follow instructions*. Retrieved June 2023, from OpenAI Webpage: <https://openai.com/research/instruction-following>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. (2016). *Intelligence Unleashed: An argument for AI in Education*. London: Pearson Education.
- Marzano, R., & Marzano, J. (2003). The Key to Classroom Management. *Educational Leadership*, 61(1), 6-13.
- Maugeri, G. (2020). La costruzione di una prova di verifica e la valutazione. In G. (. Serragiotto, *CEDILS. Certificazione in didattica dell'italiano a stranieri dell'Università Ca' Foscari di Venezia* (pp. 122-137). Torino: Loescher.
- McCarthy, J. (2007). *What is artificial intelligence?* Retrieved May 2023, from Formal Reasoning Group: <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>
- McCarthy, J., Minsky, M., & Shannon, C. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence August 31, 1955. *AI Magazine*, 27(4), 12-14.
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). AI and education: Guidance for policy-makers. (<https://doi.org/10.54675/PCSP7350>, Ed.) *The United Nations Educational, Scientific and Cultural Organization (UNESCO)*, <https://doi.org/10.54675/PCSP7350>.
- Mikropoulos, T., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56, 769–780.
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2, 1-13. <https://doi.org/10.1016/j.rmal.2023.100050>.
- Mohammed, P., & Watson, N. (2019). Towards Inclusive Education in the Age of Artificial Intelligence: Perspectives, Challenges, and Opportunities. In J. Knox, Y. Wang, & M. Gallagher, *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices* (pp. 17-37). Springer Singapore.

- Moqbel, M., & Al-Kadi, A. (2023). Foreign Language Learning Assessment in the Age of ChatGPT: A Theoretical Account. *Journal of English Studies in Arabia Felix*, 2(1), 71–84. DOI: 10.56540/jesaf.v2i1.62.
- Morton, T., & Escobar Urmeneta, C. (2019). Classroom assessment in CLIL and FLE. *MEd in Teacher Development for Foreign Language Education and Content and Language Integrated Learning (CLIL) Module 3*.
- Muñoz, S. A., Gayoso, G. G., Huambo, A. C., Tapia, R. D., Incaluque, J. L., Aguila, O. E., Cajamarca, J. C., Acevedo, J. E., & Rivera, H. V. (2023). Examining the Impacts of ChatGPT on Student Motivation and Engagement. *Social Space*, 23(1), 1-27.
- Nadimpalli, M. (2017). Artificial Intelligence Risks and Benefits. *International Journal of Innovative Research in Science, Engineering and Technology*, 6(6).
- Nye, B. (2015). Intelligent Tutoring Systems by and for the Developing World: A Review of Trends and Approaches for Educational Technology in a Global Context. *International Journal of Artificial Intelligence in Education*, 25, 177-203.
- OpenAI. (2019, February 14). *Better language models and their implications*. Retrieved June 2023, from OpenAI Webpage: <https://openai.com/research/better-language-models>
- OpenAI. (2022, November 30). *Introducing ChatGPT*. Retrieved June 2023, from OpenAI Webpage: <https://openai.com/blog/chatgpt>
- OpenAI. (2023, March 23). *ChatGPT plugins*. Retrieved June 2023, from OpenAI Webpage: <https://openai.com/blog/chatgpt-plugins>
- OpenAI. (2023). GPT-4 Technical Report. *OpenAI*, 1-100. <https://doi.org/10.48550/arXiv.2303.08774>.
- Opre^a, D. (2015). Teachers' Conceptions of Assessment. *Procedia - Social and Behavioral Sciences*, 209, 229-333. doi:10.1016/j.sbspro.2015.11.222.
- Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & Society*, 35, 927-936. <https://doi.org/10.1007/s00146-020-00965-5>.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *OpenAI*, 1-68. <https://doi.org/10.48550/arXiv.2203.02155>.

- Pedró, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development. *UNESCO*, Retrieved from: <https://unesdoc.unesco.org/ark:/48223/pf0000366994>.
- Pedró, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development. *UNESCO*.
- Perez Garcia, M., Saffon Lopez, S., & Donis, H. (2018). Everybody is talking about Virtual Assistants, but how are people really using them? *In Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*, 1-5. <http://dx.doi.org/10.14236/ewic/HCI2018.96>.
- Piccardo, E., Berchoud, M., Cignatta, T., Mentz, O., & Pamula, M. (2011). *Pathways through assessing, learning and teaching in the CEFR*. Strasbourg: Council of Europe.
- Pichai, S. (2023, February 06). *An important next step on our AI journey*. Retrieved June 2023, from Google Webpage: <https://blog.google/technology/ai/bard-google-ai-search-updates/>
- Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, 7(3), 135-153. DOI: 10.2478/jolace-2019-0025.
- Popenici, S., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 1-13.
- Qureshi, B. (2023). Exploring the Use of ChatGPT as a Tool for Learning and Assessment in Undergraduate Computer Science Curriculum: Opportunities and Challenges. *arXiv preprint arXiv:2304.11214.*, <https://doi.org/10.48550/arXiv.2304.11214>.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI*, 1-12. Retrieved from: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI*, Retrieved from: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

- Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu. *Publications Office of the European Union*.
- Roemmele, M., Bejan, C., & Gordon, A. (2011). Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. *AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 90-95.
- Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 26, 582–599. DOI 10.1007/s40593-016-0110-3.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International journal of computer vision*, 211–252.
- Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach. Third Edition*. Pearson Education.
- Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilic, S., Hesslow, D., Castagné, R., & Raffel, C. (2023). BLOOM: A 176B- Parameter Open- Access Multilingual Language Model. *arXiv pre-print arXiv:2211.05100v3*.
- Sharma, R., Kawachi, P., & Bozkurt, A. (2019). The Landscape of Artificial Intelligence in Open, Online and Distance Education: Promises and Concerns. *Asian Journal of Distance Education*, 14(2).
- Shi, Z., Liu, F., Lai, C., & Jin, T. (2022). Enhancing the use of evidence in argumentative writing through collaborative processing of content-based automated writing evaluation feedback. *Language Learning & Technology*, 106-128. <https://doi.org/10125/73481>.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., & Sutsk. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 484-489.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A

- general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362, 1140-1144.
- Sohail, S., Farhat, F., Himeur, Y., Nadeem, M., Madsen, D., Singh, Y., Atalla, S., & Mansoor, W. (2023). The Future of GPT: A Taxonomy of Existing ChatGPT Research, Current Challenges, and Possible Future Directions. *Available at SSRN*, 1-22. <http://dx.doi.org/10.2139/ssrn.4413921>.
- Stangor, C., & Walinga, J. (2014). Chapter 4. Brains, Bodies, and Behaviour. In *Introduction to Psychology - 1st Canadian Edition*.
- Steding, S. (2009). Machine Translation in the German Classroom: Detection, Reaction, Prevention. *Die Unterrichtspraxis/ Teaching German*, 42(2), 178-189.
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 1-11. <https://doi.org/10.1016/j.asw.2023.100752>.
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 1-10. <https://doi.org/10.37074/jalt.2023.6.1.17>.
- Swedberg, R. (2020). Exploratory research. In C. Elman, J. Gerring, & J. Mahoney, *The Production Of Knowledge: Enhancing Progress In Social Science* (pp. 17-41). Cambridge University Press.
- Tan, S. (2020). Artificial Intelligence in education: Rise of the Machines. *Journal of Applied Learning & Teaching*, 3(1), 129-133.
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., Hayek, S. E., Augereau-Landais, M., Mourey, F., & al., e. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, 11(54), 1-32. <https://doi.org/10.3390/jintelligence11030054>.
- Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2015). Rethinking assessment in a digital age: opportunities, challenges and risks. *British Educational Research Journal*, 1-23. DOI: 10.1002/berj.3215.
- Timms, M. (2016). Letting Artificial Intelligence in Education Out of the Box: Educational Cobots and Smart Classrooms. *International Journal of Artificial Intelligence in Education*, 26, 701-712.

- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971v1*.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., & Pritzel, A. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, *596*, 590–596.
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460.
- U.S National Science and Technology Council. (2016). The National Artificial Intelligence Research and Development Strategic Plan. *Networking and Information Technology Research and Development Subcommittee*.
- UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. *the United Nations Educational, Scientific and Cultural Organization*, Retrieved from: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- Uribe Enciso, O., Uribe Enciso, D., & Vargas Daza, M. (2017). Critical Thinking and its Importance in Education: Some Reflections. *Rastros Rostros*, *19*(34), 78-88. <https://doi.org/10.16925/ra.v19i34.2144>.
- Vecchione, B., Barocas, S., & Levy, K. (2021). Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies. *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*, 1-9. <https://doi.org/10.1145/3465416.3483294>.
- Wartman, S., & Combs, C. (2018). Medical Education Must Move From the Information Age to the Age of Artificial Intelligence. *Academic Medicine*, *93*(8), 1107-1109.
- Whitby, B. (2003). *Artificial Intelligence: A Beginner's Guide*. Oxford: Oneworld Publications.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *ArXiv:2302.11382*, <https://doi.org/10.48550/arXiv.2302.11382>.
- Willis, J. (1996). *A framework for task-based learning*. Longman Addison-Wesley.
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 1-25. <https://doi.org/10.1007/s10639-023-11742-4>.

- Yang, W. (2022). Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence*, 3, 1-7. <https://doi.org/10.1016/j.caeai.2022.100061>.
- Yeadon, W., Inyang, O.-O., Mizouri, A., Peach, A., & Testrow, C. (2023). The death of the short-form physics essay in the coming AI revolution. *Physics Education*, 58, 1-13.
- Zhai, X. (2022). ChatGPT User Experience: Implications for Education. *Available at SSRN 4312418*, 1-10. <http://dx.doi.org/10.2139/ssrn.4312418>.
- Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, <https://doi.org/10.1016/j.jii.2021.100224>.
- Zhou, J., Ke, P., Qiu, X., Huang, M., & Zhang, J. (2023). ChatGPT: potential, prospects, and limitations. *Frontiers of Information Technology & Electronic Engineering*, 1-6. <https://doi.org/10.1631/FITEE.2300089>.

APPENDIX A. INTERACTION WITH CHATGPT ON ASSESSMENT PRACTICES

Figure A1

Screenshots of the interaction with ChatGPT aiming at designing a rubric

1

Hi ChatGPT! I am a university teacher of the course Instructional design for language education. I need to assess a teaching proposal created by my students, which will become future teachers. Now, the majority of them have no experience in language teaching.

The university students have to create a task-based language learning (TBLL) activity following Willis' framework (1996) with the following structure: pre-task, task cycle, and post-task (or language focus).

The instructions for the task given by the professor are as follows (delimited with triple backticks):

```
...
```

1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
2. Choose a theme among food education; selfies and portraits; sustainability.
3. Choose a museum setting where the activity will take place. It can be a museum that you know or that you can visit in person or you can choose one of the Civic Museums of Venice (<https://www.visitmuve.it/it/virtual-tour/>) that you can visit virtually. The museum can be an art, science or natural museum.
4. The skill to promote is speaking. Please define the learning objectives (for example: the students will be able to orally present the main characteristics of an Italian garden by using the expression «in my opinion...» and the specific vocabulary related to gardens).
5. Choose an object or other resource (realia, works of art...) and design a task (see Willis' model: pre-task, task cycle, post-task) specifying the duration of the single stages and the materials and tools you would use.

6. Describe in detail what you would do in the classroom before and after the museum (for example, activities, materials, ICT tools etc.). Specify the activity/exercises you would do, writing down the text of the exercises: e.g., if you create a multiple choice exercise for reading comprehension, specify both the questions and alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (e.g.: I agree with you; I couldn't agree with you more; That's so true. Etc.)

```
...
```

I need you to create some guidelines in a Rubric table to help me with the assessment of the task. Use the following criteria for the assessment to include in the rubric (delimited with triple backticks):

```
...
```

- Pertinence with respect to activity requirements (instructions of the task given by the professor)
- Pertinence with respect to national curriculum, educational program or course, Common European Framework of Reference for Languages (CEFR) according to the target;
- Pertinence with respect to the learning objectives chosen by the university students.
- Originality: activities and creation of materials that you do not normally find in books or that are not copied from books or online resources.
- Suitability and feasibility of the teaching proposal in terms of time (timing for each activity and whether the activity can be completed based on the allotted time decided by the university students).
- Suitability and feasibility of the teaching proposal in terms of target (age, language knowledge, level of languages, cultural background, special needs...).

- Suitability and feasibility of the teaching proposal in terms of learning context chosen (type of school, resources present in the school chosen, museum setting, educational approach used).
- The project follows the task-based language learning structure by Willis' framework (1996), that is pre-task, task cycle, and post-task.
- Appropriate organization of the educational material employed in each phase (pre-task, task cycle, and post-task)

Structure the Rubric guidelines in a table where each row is one of the above criteria and each column is one of the following grade levels:

- No adequate
- Adequate
- Good
- Excellent

Write in each cell the things I should check very specifically and concisely.

Thanks for the help!

2



Here is a rubric table that you can use to evaluate your students' assignments.

Evaluation Criteria	No Adequate	Adequate	Good	Excellent
Pertinence with respect to activity requirements	Task does not meet the professor's instructions	Task meets some but not all of the professor's instructions	Task meets most of the professor's instructions	Task completely meets all of the professor's instructions
Pertinence with respect to national curriculum, CEFR	No alignment with the national curriculum or CEFR standards	Some alignment with the national curriculum or CEFR standards	Good alignment with the national curriculum or CEFR standards	Excellent alignment with the national curriculum and CEFR standards
Pertinence with respect to the learning objectives	Learning objectives are unclear or not related to the task	Learning objectives are somewhat clear and related to the task	Learning objectives are clear and closely related to the task	Learning objectives are clear, concise and directly related to the task

Originality	Materials and activities are copied or unoriginal	Some original activities and materials, but also some copied	Most activities and materials are original	All activities and materials are original
Suitability and feasibility of the proposal in terms of time	Timing is inappropriate or unrealistic for the activities	Timing is somewhat realistic but could be optimized	Timing is realistic and well-planned for most activities	Timing is perfectly planned and feasible for all activities
Suitability and feasibility of the proposal in terms of target	The proposal is unsuitable for the age, language knowledge, level, cultural background, or special needs of the target	The proposal is somewhat suitable for the target	The proposal is generally suitable for the target	The proposal is perfectly suitable for the target
Suitability and feasibility of the proposal in terms of learning context	The learning context is not considered or is inappropriate	The learning context is considered but not fully leveraged	The learning context is well-considered and used	The learning context is perfectly leveraged and used

Following the Willis' TBLL structure	TBLL structure is not followed	TBLL structure is somewhat followed	TBLL structure is generally followed	TBLL structure is perfectly followed
Organization of the educational material	Materials are disorganized or inappropriate	Materials are somewhat organized but could be improved	Materials are well-organized for most of the phases	Materials are perfectly organized for all phases

Please remember to consider each of the categories independently of the others. A task can excel in one area while needing improvement in others. You should consider all these factors together when giving overall feedback or assigning an overall grade.

3



Please rewrite the table being more specific in each step

4



Absolutely, here's a more specific version of your evaluation rubric:

Evaluation Criteria	No Adequate	Adequate	Good	Excellent
Pertinence with respect to activity requirements	Task misses multiple or key elements of the professor's instructions.	Task includes most instructions but misses or incorrectly implements a few.	Task incorporates nearly all instructions, with minor errors.	Task thoroughly and accurately incorporates all of the professor's instructions.
Pertinence with respect to national curriculum, CEFR	Task does not align with national curriculum or CEFR B1 English level standards.	Task shows a basic alignment with national curriculum and CEFR B1 English level standards.	Task aligns well with national curriculum and CEFR B1 English level standards, with minor exceptions.	Task demonstrates an excellent alignment and understanding of national curriculum and CEFR B1 English level standards.

Pertinence with respect to the learning objectives	Learning objectives are unclear, unrelated to the task, or not included.	Learning objectives are stated, but may not align well with the task or provide clear goals.	Learning objectives align with the task and provide clear goals for most activities.	Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.
Originality	Materials and activities are copied from sources without any original input.	Some original content but also a significant amount of copied or common activities and materials.	Most activities and materials show original thinking, with a small amount of commonly used or copied content.	All activities and materials are original, innovative, and demonstrate unique thinking.
Suitability and feasibility of the proposal in terms of time	The timing suggested is unrealistic, does not allow for task completion, or isn't provided.	Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.	Timing is mostly realistic, with considerations for transitions and possible delays in a few activities.	The suggested timing is completely realistic, provides buffer for transitions and potential delays, and allows for task completion.
Suitability and feasibility of the proposal in terms of target	The proposal does not match the target students' age, English language level, cultural background, or special needs.	The proposal partially matches the target students but lacks in one or more areas.	The proposal matches the target students in most areas and makes minor errors in others.	The proposal perfectly matches the target students' age, English language level, cultural background, and special needs.
Suitability and feasibility of the proposal in terms of learning context	The chosen school and museum settings are not considered or inappropriate for the proposal.	The proposal shows some consideration for the settings, but could be better integrated.	The proposal shows thoughtful integration with the settings, with minor exceptions.	The proposal is perfectly integrated into the chosen school and museum settings, enhancing the learning experience.
Following the Willis' TBLL structure	The TBLL structure is not followed or elements are missing.	The TBLL structure is attempted but lacks a clear pre-task, task cycle, or post-task.	The TBLL structure is mostly followed with minor deviations.	The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

Organization of the educational material	Materials are disorganized, lacking, or inappropriate for the phases of the task.	Materials are organized and appropriate for some but not all phases.	Materials are mostly well-organized and appropriate for the phases of the task.	Materials are meticulously organized and perfectly suited for each phase of the task.
---	---	--	---	---

Remember, each criterion should be considered independently. While an activity might excel in one area, it might need improvements in others. These criteria should be used as a comprehensive guide for assessment and feedback.

5



Can you restructure the table as a list? Following the structure:

...

1. **Pertinence with respect to activity requirements:**

- No adequate: Task misses multiple or key elements of the professor's instructions.
- Adequate: ...
- Good: ...
- Excellent: ...

...

...

6



Sure, here's the evaluation rubric restructured into list format:

1. **Pertinence with respect to activity requirements:**

- No adequate: Task misses multiple or key elements of the professor's instructions.
- Adequate: Task includes most instructions but misses or incorrectly implements a few.
- Good: Task incorporates nearly all instructions, with minor errors.
- Excellent: Task thoroughly and accurately incorporates all of the professor's instructions.

2. **Pertinence with respect to national curriculum, CEFR:**

- No adequate: Task does not align with national curriculum or CEFR B1 English level standards.
- Adequate: Task shows a basic alignment with national curriculum and CEFR B1 English level standards.
- Good: Task aligns well with national curriculum and CEFR B1 English level standards, with minor exceptions.
- Excellent: Task demonstrates an excellent alignment and understanding of national curriculum and CEFR B1 English level standards.

3. **Pertinence with respect to the learning objectives:**

- No adequate: Learning objectives are unclear, unrelated to the task, or not included.
- Adequate: Learning objectives are stated, but may not align well with the task or provide clear goals.
- Good: Learning objectives align with the task and provide clear goals for most activities.
- Excellent: Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.

4. **Originality:**

- No adequate: Materials and activities are copied from sources without any original input.
- Adequate: Some original content but also a significant amount of copied or common activities and materials.
- Good: Most activities and materials show original thinking, with a small amount of commonly used or copied content.
- Excellent: All activities and materials are original, innovative, and demonstrate unique thinking.

5. Suitability and feasibility of the proposal in terms of time:

- No adequate: The timing suggested is unrealistic, does not allow for task completion, or isn't provided.
- Adequate: Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.
- Good: Timing is mostly realistic, with considerations for transitions and possible delays in a few activities.
- Excellent: The suggested timing is completely realistic, provides buffer for transitions and potential delays, and allows for task completion.

6. Suitability and feasibility of the proposal in terms of target:

- No adequate: The proposal does not match the target students' age, English language level, cultural background, or special needs.
- Adequate: The proposal partially matches the target students but lacks in one or more areas.
- Good: The proposal matches the target students in most areas and makes minor errors in others.
- Excellent: The proposal perfectly matches the target students' age, English language level, cultural background, and special needs.

7. Suitability and feasibility of the proposal in terms of learning context:

- No adequate: The chosen school and museum settings are not considered or inappropriate for the proposal.
- Adequate: The proposal shows some consideration for the settings, but could be better integrated.
- Good: The proposal shows thoughtful integration with the settings, with minor exceptions.
- Excellent: The proposal is perfectly integrated into the chosen school and museum settings, enhancing the learning experience.

8. Following the Willis' TBLL structure:

- No adequate: The TBLL structure is not followed or elements are missing.
- Adequate: The TBLL structure is attempted but lacks a clear pre-task, task cycle, or post-task.
- Good: The TBLL structure is mostly followed with minor deviations.
- Excellent: The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

9. Organization of the educational material:

- No adequate: Materials are disorganized, lacking, or inappropriate for the phases of the task.
- Adequate: Materials are organized and appropriate for some but not all phases.
- Good: Materials are mostly well-organized and appropriate for the phases of the task.
- Excellent: Materials are meticulously organized and perfectly suited for each phase of the task.

This list should help you assess your students' tasks more comprehensively.

Figure A2

Screenshots of the interaction with ChatGPT aiming at assessing a TBLL activity using the rubric (one criterion at a time)

1



Hi ChatGPT! I need you to assess a task from some university students of the course of Instructional design for language education. The students of this course will become teachers in the future. Now, the majority of them have no experiences in language teaching.

The university students have to create a task-based language learning (TBLL) activity following Willis' framework (1996) with the following structure: pre-task, task cycle, and post-task (or language focus).

The instructions of the task given by the professor are as follows (delimited in triple backticks):

```\n

1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
2. Choose a theme among: Food education; Selfies and portraits; Sustainability.
3. Choose a museum setting where the activity will take place. It can be a museum that you know or that you can visit in person or you can choose one of the Civic Museums of Venice (<https://www.visitmuve.it/it/virtual-tour/>) that you can visit virtually. The museum can be an art, science or natural museum.
4. The skill to promote is speaking. Please define the learning objectives (for example: the students will be able to orally present the main characteristics of an Italian garden by using the expression «in my opinion...» and the specific vocabulary related to gardens).
5. Choose an object or other resource (realia, works of art...) and design a task (see Willis' model: pre-task, task cycle, post-task) specifying the duration of the single stages and the materials and tools you would use.
6. Describe in detail what you would do in the classroom before and after the museum (for example, activities, materials, ICT tools etc.). Specify the activity/exercises you would do, writing down the text of the exercises: e.g., if you create a multiple choice exercise for reading comprehension, specify both the questions and alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (e.g.: I agree with you; I couldn't agree with you more; That's so true. Etc.)

```\n

and the task they created is (delimited in triple backticks):

...

* Topic: selfies and portraits

* Museum: National Portrait Gallery, London (FREE admission)

* Skill to promote: speaking

* Objectives:

* Content Objectives:

* Students will be able to identify and describe different types of portraits, including traditional, contemporary, and selfie portraits.

* Students will be able to compare and contrast the cultural and social significance of different types of portraits, including selfies.

* Students will be able to reflect on their self-image and how they present it to the world.

* Language Objectives:

* Students will be able to use vocabulary related to portraits, such as composition, technique and style.

* Students will be able to express their opinions and interpretations about different portraits using appropriate language structures, such as giving reasons and examples.

* Students will be able to collaborate with their peers to create a group portrait using appropriate language for giving and receiving feedback.

* Students will be able to use appropriate language to describe their self-image in a verbal way.

* Objects: portraits (students can choose one of them).

The activity:

This acquisition unit is composed of three parts, two of which will take place in the classroom and one will take place during the visit at the museum. Out of the three sections of the task, only the pre-task and the task-cycle will occur in the museum, whilst the post-task language focus and follow-up activities will be performed in class, because of the lack of instructional tools in a museum setting.

* Classroom preparation to the task in the museum

1. Vocabulary building: Introduce and practice relevant vocabulary related to the topic of "Selfies and Portraits", such as composition, technique, and style. The students are asked to work in teams and participate in a Kahoot quiz.

Examples of Kahoot quiz questions:

1. What is the composition of a portrait?

1. The colors used in the portrait
2. The arrangement of the elements in the portrait
3. The type of camera used to take the portrait
4. The background of the portrait

2. Which of the following is not a technique used in portrait photography?

1. Lighting
2. Posing
3. Facial expressions
4. Video editing

3. What is a selfie?
 1. A portrait taken by someone else
 2. A portrait taken using a smartphone or digital camera
 3. A portrait taken in a studio
 4. A portrait taken using film photography
 4. What is the difference between a traditional portrait and a contemporary portrait?
 1. The type of camera used to take the portrait
 2. The style and composition of the portrait
 3. The time period in which the portrait was taken
 4. The subject of the portrait
 5. Which artist is famous for painting self-portraits?
 1. Vincent van Gogh
 2. Pablo Picasso
 3. Frida Kahlo
 4. Leonardo da Vinci
 6. What is the purpose of a portrait?
 1. To capture the likeness of the subject
 2. To convey the social status of the subject
 3. To express the emotions of the subject
 4. All of the above
 7. What is the difference between a posed portrait and a candid portrait?
 1. A posed portrait is taken without the subject's knowledge, while a candid portrait is taken with the subject's permission
 2. A posed portrait is carefully arranged, while a candid portrait is spontaneous
 3. A posed portrait is always taken in a studio, while a candid portrait is taken in any setting
 4. A posed portrait is always serious, while a candid portrait can be lighthearted
 8. What is the most common type of portrait?
 1. Self-portrait
 2. Family portrait
 3. Group portrait
 4. Pet portrait
 9. What is the difference between a portrait and a selfie?
 1. A portrait is taken by someone else, while a selfie is taken by the subject
 2. A portrait is always serious, while a selfie can be lighthearted
 3. A portrait is carefully arranged, while a selfie is spontaneous
 4. All of the above
2. Discussion and debate: engage students in discussions and debates about the topic of "Selfies and Portraits". In the same small groups, students are asked open-ended questions to encourage critical thinking and opinion-sharing. After discussing in groups the students come together to share what they have discussed.

Examples of questions:

1. What makes a good portrait/selfie?
2. Do you think selfies are a form of art? Why or why not?
3. What do you think is the difference between a selfie and a traditional portrait?
4. Do you think selfies are a reflection of our culture? Why or why not?
5. Is there a difference between the way men and women take selfies? Why or why not?
6. What do you think is the role of portraits in society?
7. How has the portrayal of beauty and body image in portraits evolved over time?

* Task: Create a Portrait Gallery

* Level: B1

* Time: 60 minutes

* Aim: To develop students' ability to analyze and create different types of portraits.

Procedure:

1. Pre-task: (5 minutes, in the museum)

1. Brainstorm different types of portraits with the class (e.g. traditional, contemporary, selfie).
2. Ask students to discuss in pairs or small groups what they think makes a good portrait.

2. Task cycle (in the museum):

1. Task: (35 minutes)

1. Divide the class into small groups (3-4 students per group).
2. Ask each group to select a portrait out of all the portraits present in the gallery
3. Ask each group to analyze the portrait based on its composition, technique, and style.
4. Ask each group to create their own portrait drawing inspiration from one of the portraits of the gallery and discuss how they can incorporate the elements they analyzed in the previous step. The students will then take a picture of their portrait to show to the classmates afterwards.

2. Planning: (5 minutes)

1. Ask each group to plan how they will present their portrait to the class.

3. Report: (15 minutes)

1. Ask each group to present their portrait to the class. The classmates will then try to guess which portrait their peers took inspiration from.
2. The group will then explain their analysis and creative choices in creating the portrait.
3. Encourage the class to ask questions and give feedback to each group.

3. Language focus (in the classroom):

1. Language game: students are asked to play a game of "Taboo" where they have to describe a portrait without using certain keywords or phrases, in order to reinforce vocabulary, grammar, and pronunciation.

2. Language structures: students will be provided with a handout of language structures (such as "in my opinion...", "I believe that...", or "from my perspective...") that they can use to express their ideas and opinions about selfies and portraits.

4. Follow-up (in the classroom, over two consecutive lessons):

1. Students are asked to reflect on what they learned and how they can apply it to their own portrait-taking and appreciation.

2. Students are asked to take a selfie that represents their identity to the fullest, by also using objects and clothes they find representative of their own personality. The selfies will be presented and described to the classmates, who will try to interpret their peers' images and their creative choices. Finally, the creator of the selfie will explain the reasons behind their choices and talk about themselves.

...

Can you assess it according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

Pertinence with respect to activity requirements:

* No adequate: The student's activity misses multiple or key elements of the instructions.

* Adequate: The student's activity includes most instructions but misses or incorrectly implements a few.

* Good: The student's activity incorporates nearly all instructions, with minor errors.

* Excellent: The student's activity thoroughly and accurately incorporates all of the professor's instructions.

...

2



In order to assess the students' activity, let's examine each of the professor's instructions in detail:

1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
 - The students' activity seems to target the given audience, although there is no explicit mention of it.
2. Choose a theme among: Food education; Selfies and portraits; Sustainability.
 - The theme chosen by the students is "Selfies and Portraits", which is one of the three options given.
3. Choose a museum setting where the activity will take place.
 - The students chose the National Portrait Gallery in London as the setting for the activity.
4. The skill to promote is speaking. Please define the learning objectives.
 - The skill to be promoted is explicitly identified as "speaking" and the learning objectives are clearly listed under both "Content Objectives" and "Language Objectives".
5. Choose an object or other resource and design a task.
 - The task chosen involves different types of portraits (traditional, contemporary, selfie). The students have designed a multi-part task (pre-task, task cycle, and post-task) following Willis' framework.
6. Describe in detail what you would do in the classroom before and after the museum.
 - The students have detailed classroom activities both before (Vocabulary building and Discussion and debate) and after (Language focus and Follow-up) the museum visit.

Based on these observations, the students' activity appears to be in alignment with all of the professor's instructions. All instructions are addressed and developed with sufficient detail, so I would suggest that the students' work can be evaluated as follows:

Pertinence with respect to activity requirements:

- **Excellent:** The student's activity thoroughly and accurately incorporates all of the professor's instructions.

3



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

2. Pertinence with respect to national curriculum, CEFR:

- * No adequate: The student's activity does not align with national curriculum or CEFR B1 English level standards.
- * Adequate: The student's activity shows a basic alignment with national curriculum or CEFR B1 English level standards.
- * Good: The student's activity aligns well with national curriculum and CEFR B1 English level standards, with minor exceptions.
- * Excellent: The student's activity demonstrates an excellent alignment and understanding of national curriculum or CEFR B1 English level standards.

...

4



To assess the pertinence of the student's activity with respect to the national curriculum and CEFR B1 English level standards, we need to consider the following:

At the B1 level, the Common European Framework of Reference for Languages (CEFR) outlines that learners should be able to understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. They can deal with most situations likely to arise whilst travelling in an area where the language is spoken. They can produce simple connected text on topics that are familiar or of personal interest. They can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

Given this description, let's evaluate the students' activity:

1. The objectives of the activity outline various aspects that align with the CEFR B1 English level standards. For instance:
 - Content Objectives like identifying and describing different types of portraits, comparing and contrasting cultural and social significance of different types of portraits and reflecting on self-image all fit into understanding main points of familiar matters and producing simple connected text on familiar topics.
 - Language Objectives such as using vocabulary related to portraits, expressing opinions and interpretations, collaborating with peers, and describing self-image verbally align with the ability to describe experiences and events, and give reasons and explanations for opinions and plans.
2. The tasks designed for the students involve speaking, interaction, and the expression of personal opinions and experiences, which are important elements of B1 level language acquisition.
3. The activity is designed to encourage learners to deal with a situation (a museum visit), where they have to use their language skills in a real-world context.

Based on this analysis, the activity aligns well with the CEFR B1 English level standards. Regarding the national curriculum, without specific details about which country's curriculum to refer to, I will assume that any curriculum that aligns with the CEFR standards would find this activity suitable.

Hence, the students' work can be evaluated as follows:

Pertinence with respect to national curriculum, CEFR:

- **Excellent:** The student's activity demonstrates an excellent alignment and understanding of national curriculum or CEFR B1 English level standards.

5



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

3. Pertinence with respect to the learning objectives:

- * No adequate: Learning objectives are unclear, unrelated to the task, or not included.
- * Adequate: Learning objectives are stated, but may not align well with the task or provide clear goals.
- * Good: Learning objectives align with the task and provide clear goals for most activities.
- * Excellent: Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.

...



In assessing the pertinence with respect to the learning objectives, the following points are observed:

The learning objectives are clearly stated in two categories: content and language objectives.

- Content Objectives:
 - Students will be able to identify and describe different types of portraits.
 - Students will be able to compare and contrast the cultural and social significance of different types of portraits.
 - Students will be able to reflect on their self-image.
- Language Objectives:
 - Students will be able to use vocabulary related to portraits.
 - Students will be able to express their opinions and interpretations about different portraits.
 - Students will be able to collaborate with their peers to create a group portrait.
 - Students will be able to use appropriate language to describe their self-image verbally.

The task designed aligns well with these learning objectives:

- The pre-task of brainstorming different types of portraits and discussing what makes a good portrait aligns with the objectives of identifying, describing, and reflecting on portraits.
- The task cycle of analyzing a chosen portrait and creating a group portrait, followed by planning and reporting, aligns with the objectives of comparing, contrasting, expressing opinions, and using appropriate vocabulary.
- The post-task activities, including playing a game of "Taboo" and being provided with language structures to express opinions, align with the objectives of using related vocabulary and expressing opinions.
- The follow-up activity of taking a selfie that represents their identity and explaining it to their peers also aligns with the objective of reflecting on their self-image.

These activities set clear and measurable outcomes related to the stated objectives. For instance, the students are not only expected to learn about portraits but also use the relevant vocabulary and express their own interpretations, which can be clearly observed and assessed during the tasks.

Therefore, I would evaluate the students' work as follows:

Pertinence with respect to the learning objectives:

- **Excellent:** Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.

7



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

4. Originality:

* No adequate: The student's work is largely unoriginal, heavily relying on ideas, activities, or materials that can be commonly found in textbooks or online resources.

* Adequate: The student's work presents some original content, but still relies on ideas, activities, or materials that can be commonly found in textbooks or online resources..

* Good: Most activities and materials of the student's work show original thinking, with a small amount of commonly used content.

* Excellent: All activities and materials are original, innovative, and demonstrate creative and innovative thinking.

...

8



In assessing the originality of the task, the following aspects are considered:

- The theme of "Selfies and Portraits" and the setting of the National Portrait Gallery are not uncommon in language learning activities, but the way they are woven together is somewhat unique.
- The structure of the task, following Willis' model, is not original in itself but it is a pedagogically sound and recognized method in task-based language learning.
- The idea of conducting a part of the lesson in a museum is relatively original and can provide a unique learning experience to the students.
- The use of a Kahoot quiz for vocabulary building is a common method used in classrooms, but the questions prepared for the quiz are specific and original.
- The discussions and debates before the museum visit are good strategies for activating prior knowledge and stimulating students' thinking about the theme, and the questions suggested are interesting and thought-provoking.
- The task cycle involving analyzing a chosen portrait and creating a group portrait in the museum, and later presenting their work to the class, demonstrates a degree of originality.
- The post-task activities like the game of "Taboo" are commonly used, but they are well adjusted to the task.
- The follow-up activity where students take a selfie that represents their identity and explain it to their peers is quite original and personal, promoting creative and reflective thinking in students.

Overall, while the task uses some common methods and activities, it also presents a good deal of original content and applies common methods in original ways.

Therefore, I would evaluate the students' work as follows:

Originality:

- **Good:** Most activities and materials of the student's work show original thinking, with a small amount of commonly used content.

9



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

5. Suitability and feasibility of the proposal in terms of time:

* No adequate: The timing suggested is unrealistic, does not allow for task completion, or isn't provided.

* Adequate: Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.

* Good: Timing is mostly realistic, with considerations for transitions and possible delays in a few activities.

* Excellent: The suggested timing is completely realistic, allows for task completion and takes into considerations possible delays.

...

10



Let's evaluate the suitability and feasibility of the proposal in terms of time:

- The pre-task activities include a Kahoot quiz and a debate, which are typically manageable in a single class period, but the timing isn't explicitly specified. However, these activities could possibly take more time than expected, especially if students engage deeply with the debate.
- The task cycle involves a visit to the National Portrait Gallery, which includes travel time, time spent at the museum, and a post-visit class. Depending on the location of the school in relation to the museum, travel time could potentially complicate the timing. Additionally, timing may also depend on the size of the class and the complexity of the portraits the students choose to analyze.
- Post-task activities include a game of "Taboo" and a follow-up activity of taking selfies, which are generally feasible within a single class period. However, the timing again isn't explicitly stated.
- Language focus activities are typically conducted over several class periods, and would likely fit within the estimated two weeks. However, depending on the learners' pace and needs, it might take longer.

In general, the proposal seems potentially feasible within the two-week timeframe, although the specifics are somewhat uncertain, and there might be potential delays due to factors like the museum visit or deep engagement with the activities.

Therefore, based on the given scale, I would evaluate this aspect as:

Suitability and feasibility of the proposal in terms of time:

- **Adequate:** Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.

11



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

6. Suitability and feasibility of the proposal in terms of target:

* No adequate: The proposal does not match the target students' age, English language level, cultural background, or special needs.

* Adequate: The proposal partially matches the target students but lacks in one or more areas.

* Good: The proposal matches the target students in most areas and makes minor errors in others.

* Excellent: The proposal perfectly matches the target students' age, English language level, cultural background, and special needs.

...

12



To evaluate the suitability and feasibility of the proposal in terms of the target, let's look at the task in relation to the specified target audience: 16-18-year-old students, English as a foreign language, B1 level.

- **Age:** The task is designed with an appropriate level of complexity and engagement for 16-18-year-old students. The theme of "Selfies and Portraits" is likely to be interesting and relevant for this age group, and the activities such as the museum visit, group work, and selfie task are suitable for this age.
- **English Language Level:** The task appears to be designed with B1 level English learners in mind. The vocabulary and language structures to be used in the task seem to be at the right level of difficulty for B1 learners. Activities like the Kahoot quiz, group discussions, and the presentation should be manageable for B1 students.
- **Cultural Background:** While the task takes place in the National Portrait Gallery in London, it deals with a universal theme of self-portrait and identity which is likely to be applicable to students from various cultural backgrounds. It could be adapted to other settings if the museum visit is not possible for geographical reasons.
- **Special Needs:** There is no specific mention of accommodations for students with special needs. However, the task does not seem to have any obvious barriers that could prevent students with certain special needs from participating, provided that standard accommodations are in place.

Given these considerations, the proposal appears to match well with the target audience, albeit with minor areas that might need additional considerations or adjustments.

Therefore, based on the given scale, I would evaluate this aspect as:

Suitability and feasibility of the proposal in terms of target:

- **Good:** The proposal matches the target students in most areas and makes minor errors in others.

13



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

7. Suitability and feasibility of the proposal in terms of learning context:

* No adequate: The chosen school and museum settings are not considered or inappropriate for the proposal.

* Adequate: The proposal shows some consideration for the settings, but could be better integrated.

* Good: The proposal shows thoughtful integration with the settings, with minor exceptions.

* Excellent: The proposal is perfectly integrated into the chosen school and museum settings, enhancing the learning experience.

...

14



In assessing the suitability and feasibility of the proposal in terms of learning context, the chosen school and museum settings must be considered:

- The school setting is considered for classroom activities. This includes the use of a Kahoot quiz, discussions, debates, and a language game (Taboo). It also includes the follow-up activity of taking and explaining selfies that represent students' identities.
- The museum setting (the National Portrait Gallery in London) is chosen for the main task, which includes the pre-task of brainstorming and discussing portraits, and the task cycle of analyzing a chosen portrait, creating a group portrait, and presenting their work to the class.
- The activities are designed with the learning context in mind. For instance, the pre-task and task cycle activities are planned to take place in the museum, which enhances the students' learning experience by allowing them to observe, analyze, and discuss real portraits.
- The post-task activities and the follow-up task are planned to take place in the classroom, which allows students to reflect on their museum experience, consolidate their learning, and apply what they learned to a personal task.

Overall, the proposal appears to be well integrated into the chosen school and museum settings. However, a possible point of concern might be the feasibility of visiting the National Portrait Gallery, especially if the students are located outside of London or even outside of the UK. Although a virtual visit could potentially address this concern, this isn't explicitly stated in the proposal.

Therefore, based on the given scale, I would evaluate this aspect as:

Suitability and feasibility of the proposal in terms of learning context:

- **Good:** The proposal shows thoughtful integration with the settings, with minor exceptions.

15



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

8. Following the Willis' TBLL structure:

* No adequate: The TBLL structure is not followed or elements are missing.

* Adequate: The TBLL structure is attempted but lacks a clear pre-task, task cycle, or post-task.

* Good: The TBLL structure is mostly followed with minor deviations.

* Excellent: The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

...

16



In order to evaluate the extent to which the task follows Willis' Task-Based Language Learning (TBLL) structure, we need to look at how the task was designed with regards to the pre-task, task cycle, and post-task phases:

- Pre-task: The students designed a clear pre-task phase that includes introducing and practicing relevant vocabulary through a Kahoot quiz, as well as engaging students in discussions and debates about the topic of "Selfies and Portraits". This phase is meant to prepare students for the main task at the museum.
- Task cycle: The students' task cycle is also clearly outlined. It begins at the museum, where students, in small groups, select and analyze a portrait and then create their own group portrait drawing inspiration from the analyzed portrait. This is followed by a planning and reporting phase in which students plan how they will present their work to the class, and then do so.
- Post-task: The students' post-task phase includes a language focus activity and a follow-up task. In the language focus activity, students play a game of "Taboo" and are provided with language structures to express their ideas and opinions. The follow-up task involves a personal project where students take a selfie that represents their identity, present it to the class, and explain the reasons behind their choices.

Based on these observations, it can be seen that the task adheres to Willis' TBLL structure. All three phases are clearly delineated and thoughtfully designed to scaffold the students' learning.

Therefore, based on the given scale, I would evaluate this aspect as:

Following the Willis' TBLL structure:

- **Excellent:** The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

17



All right. Now could you assess the task according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...`

9. Organization of the educational material:

* No adequate: Materials are disorganized, lacking, or inappropriate for the phases of the task.

* Adequate: Materials are organized and appropriate for some but not all phases.

* Good: Materials are mostly well-organized and appropriate for the phases of the task.

* Excellent: Materials are meticulously organized and perfectly suited for each phase of the task.

...`

18



To assess the organization of the educational material, let's look at how materials are prepared and utilized in different phases of the task:

- Pre-task: The Kahoot quiz is designed to introduce and practice relevant vocabulary. The materials required for this activity are the Kahoot quiz questions, which are well-organized and appropriately suited for the task. The discussion and debate session might not require specific materials, but clear instructions and questions are given, ensuring structure and guidance.
- Task cycle: For the museum visit, the main materials are the portraits themselves. The students are to analyze a chosen portrait and create a group portrait. The task here is well-structured, and the portraits in the museum are naturally organized and suitable for the task. However, it's unclear what materials students will use to create their own portrait.
- Post-task: The post-task activities include a game of "Taboo" and a follow-up activity of taking selfies. The "Taboo" game would require cards or a list of words related to the theme, but these materials are not explicitly mentioned or prepared. The follow-up activity requires students to take selfies, for which they would presumably use their own smartphones or cameras, but again, these materials are not explicitly mentioned.

In general, while some materials are well-prepared and appropriate for the tasks, others are not explicitly mentioned or prepared. Some assumptions are made about the availability of certain materials (like smartphones for taking selfies), which could potentially pose problems in practice.

Therefore, based on the given scale, I would evaluate this aspect as:

Organization of the educational material:

- **Adequate:** Materials are organized and appropriate for some but not all phases.

19



Now, could you write a single textual feedback including the 9 assessments you previously gave? Try to first highlight the best parts and then talk about the improvements that could be done

20



Sure, I'd be happy to summarize the assessments into a single feedback.

Overall, the task you created shows a thorough understanding of the task-based language learning (TBLL) approach and is mostly well-designed, innovative, and pedagogically sound.

- **Strengths:** The strongest aspects of your task are its alignment with Willis' TBLL structure and the pertinence with respect to the professor's instructions, the national curriculum, CEFR B1 English level standards, and the learning objectives. Your task demonstrates an excellent understanding of these frameworks and standards, as well as the ability to create clear, measurable learning objectives that align with your chosen task.
- The theme of "Selfies and Portraits" and the setting of the National Portrait Gallery are not only original and interesting but also aptly chosen for the age group and English language level of the target students. Your activities show thoughtful integration with the chosen settings and a well-considered balance between classroom and museum activities.
- Your use of a Kahoot quiz for vocabulary building is innovative and engaging, and your follow-up task of taking and explaining a selfie that represents a student's identity is an excellent example of a task that is not only engaging and personally relevant but also promotes reflection and deeper learning.

- **Areas for Improvement:** There are a few areas that could be improved. Regarding the timing, while the proposal seems potentially feasible within the two-week timeframe, more detailed consideration should be given to possible delays due to factors like the museum visit or deep engagement with the activities. The organization of the educational material could also be improved by explicitly stating the materials needed for each phase of the task.
- Finally, while the proposal matches the target students in most areas, there is no specific mention of accommodations for students with special needs. Including some options or suggestions for differentiation to accommodate all learners could make your task even more inclusive.

In conclusion, your task represents a strong example of a TBLL activity that is engaging, relevant, and pedagogically sound. With some minor adjustments and more detailed planning, it could be an excellent resource for teaching and learning. Well done!

Figure A3

Screenshots of the interaction with ChatGPT aiming at assessing a TBLL activity using the whole rubric in a single prompt

1



Hi ChatGPT! I need you to assess a task from some university students of the course of Instructional design for language education. The students of this course will become teachers in the future. Now, the majority of them have no experiences in language teaching.

The university students have to create a task-based language learning (TBLL) activity following Willis' framework (1996) with the following structure: pre-task, task cycle, and post-task (or language focus).

The instructions of the task given by the professor are as follows (delimited in triple backticks):

- ```

1. The activity you have to create is for the following target: 16-18-year-old students, English FL, B1 level.
2. Choose a theme among: Food education; Selfies and portraits; Sustainability.
3. Choose a museum setting where the activity will take place. It can be a museum that you know or that you can visit in person or you can choose one of the Civic Museums of Venice (https://www.visitmuve.it/it/virtual-tour/) that you can visit virtually. The museum can be an art, science or natural museum.
4. The skill to promote is speaking. Please define the learning objectives (for example: the students will be able to orally present the main characteristics of an Italian garden by using the expression «in my opinion...» and the specific vocabulary related to gardens).
5. Choose an object or other resource (realia, works of art...) and design a task (see Willis' model: pre-task, task cycle, post-task) specifying the duration of the single stages and the materials and tools you would use.
6. Describe in detail what you would do in the classroom before and after the museum (for example, activities, materials, ICT tools etc.). Specify the activity/exercises you would do, writing down the text of the exercises: e.g., if you create a multiple choice exercise for reading comprehension, specify both the questions and alternative answers; if you decide to provide your students with the language support to express agreement, write down the sentences you would give them (e.g.: I agree with you; I couldn't agree with you more; That's so true. Etc.)

```

and the task they created is (delimited in triple backticks):

...

- \* Topic: selfies and portraits
- \* Museum: National Portrait Gallery, London (FREE admission)
- \* Skill to promote: speaking
- \* Objectives:
  - \* Content Objectives:
    - \* Students will be able to identify and describe different types of portraits, including traditional, contemporary, and selfie portraits.
    - \* Students will be able to compare and contrast the cultural and social significance of different types of portraits, including selfies.
    - \* Students will be able to reflect on their self-image and how they present it to the world.
  - \* Language Objectives:
    - \* Students will be able to use vocabulary related to portraits, such as composition, technique and style.
    - \* Students will be able to express their opinions and interpretations about different portraits using appropriate language structures, such as giving reasons and examples.
    - \* Students will be able to collaborate with their peers to create a group portrait using appropriate language for giving and receiving feedback.
    - \* Students will be able to use appropriate language to describe their self-image in a verbal way.
- \* Objects: portraits (students can choose one of them).

The activity:

This acquisition unit is composed of three parts, two of which will take place in the classroom and one will take place during the visit at the museum. Out of the three sections of the task, only the pre-task and the task-cycle will occur in the museum, whilst the post-task language focus and follow-up activities will be performed in class, because of the lack of instructional tools in a museum setting.

\* Classroom preparation to the task in the museum

1. Vocabulary building: Introduce and practice relevant vocabulary related to the topic of "Selfies and Portraits", such as composition, technique, and style. The students are asked to work in teams and participate in a Kahoot quiz.

Examples of Kahoot quiz questions:

1. What is the composition of a portrait?
  1. The colors used in the portrait
  2. The arrangement of the elements in the portrait
  3. The type of camera used to take the portrait
  4. The background of the portrait
2. Which of the following is not a technique used in portrait photography?
  1. Lighting
  2. Posing
  3. Facial expressions
  4. Video editing
3. What is a selfie?
  1. A portrait taken by someone else
  2. A portrait taken using a smartphone or digital camera
  3. A portrait taken in a studio
  4. A portrait taken using film photography
4. What is the difference between a traditional portrait and a contemporary portrait?
  1. The type of camera used to take the portrait
  2. The style and composition of the portrait
  3. The time period in which the portrait was taken
  4. The subject of the portrait
5. Which artist is famous for painting self-portraits?
  1. Vincent van Gogh
  2. Pablo Picasso
  3. Frida Kahlo
  4. Leonardo da Vinci

6. What is the purpose of a portrait?
  1. To capture the likeness of the subject
  2. To convey the social status of the subject
  3. To express the emotions of the subject
  4. All of the above
7. What is the difference between a posed portrait and a candid portrait?
  1. A posed portrait is taken without the subject's knowledge, while a candid portrait is taken with the subject's permission
  2. A posed portrait is carefully arranged, while a candid portrait is spontaneous
  3. A posed portrait is always taken in a studio, while a candid portrait is taken in any setting
  4. A posed portrait is always serious, while a candid portrait can be lighthearted
8. What is the most common type of portrait?
  1. Self-portrait
  2. Family portrait
  3. Group portrait
  4. Pet portrait
9. What is the difference between a portrait and a selfie?
  1. A portrait is taken by someone else, while a selfie is taken by the subject
  2. A portrait is always serious, while a selfie can be lighthearted
  3. A portrait is carefully arranged, while a selfie is spontaneous
  4. All of the above

2. Discussion and debate: engage students in discussions and debates about the topic of "Selfies and Portraits". In the same small groups, students are asked open-ended questions to encourage critical thinking and opinion-sharing. After discussing in groups the students come together to share what they have discussed.

Examples of questions:

1. What makes a good portrait/selfie?
2. Do you think selfies are a form of art? Why or why not?
3. What do you think is the difference between a selfie and a traditional portrait?
4. Do you think selfies are a reflection of our culture? Why or why not?
5. Is there a difference between the way men and women take selfies? Why or why not?
6. What do you think is the role of portraits in society?
7. How has the portrayal of beauty and body image in portraits evolved over time?

\* Task: Create a Portrait Gallery

\* Level: B1

\* Time: 60 minutes

\* Aim: To develop students' ability to analyze and create different types of portraits.

Procedure:

1. Pre-task: (5 minutes, in the museum)
  1. Brainstorm different types of portraits with the class (e.g. traditional, contemporary, selfie).
  2. Ask students to discuss in pairs or small groups what they think makes a good portrait.
2. Task cycle (in the museum):
  1. Task: (35 minutes)
    1. Divide the class into small groups (3-4 students per group).
    2. Ask each group to select a portrait out of all the portraits present in the gallery
    3. Ask each group to analyze the portrait based on its composition, technique, and style.
    4. Ask each group to create their own portrait drawing inspiration from one of the portraits of the gallery and discuss how they can incorporate the elements they analyzed in the previous step. The students will then take a picture of their portrait to show to the classmates afterwards.
  2. Planning: (5 minutes)
    1. Ask each group to plan how they will present their portrait to the class.
  3. Report: (15 minutes)
    1. Ask each group to present their portrait to the class. The classmates will then try to guess which portrait their peers took inspiration from.
    2. The group will then explain their analysis and creative choices in creating the portrait.
    3. Encourage the class to ask questions and give feedback to each group.

3. Language focus (in the classroom):

1. Language game: students are asked to play a game of "Taboo" where they have to describe a portrait without using certain keywords or phrases, in order to reinforce vocabulary, grammar, and pronunciation.

2. Language structures: students will be provided with a handout of language structures (such as "in my opinion...", "I believe that...", or "from my perspective...") that they can use to express their ideas and opinions about selfies and portraits.

4. Follow-up (in the classroom, over two consecutive lessons):

1. Students are asked to reflect on what they learned and how they can apply it to their own portrait-taking and appreciation.

2. Students are asked to take a selfie that represents their identity to the fullest, by also using objects and clothes they find representative of their own personality. The selfies will be presented and described to the classmates, who will try to interpret their peers' images and their creative choices. Finally, the creator of the selfie will explain the reasons behind their choices and talk about themselves.

...

Can you assess it according to the following guidelines delimited in triple backticks?

First find (and cite if needed) parts of the text that are relevant for the assessment and then give a grade according to the provided scale :

...

1. Pertinence with respect to activity requirements:

\* No adequate: The student's activity misses multiple or key elements of the instructions.

\* Adequate: The student's activity includes most instructions but misses or incorrectly implements a few.

\* Good: The student's activity incorporates nearly all instructions, with minor errors.

\* Excellent: The student's activity thoroughly and accurately incorporates all of the professor's instructions.

2. Pertinence with respect to national curriculum, CEFR:

\* No adequate: The student's activity does not align with national curriculum or CEFR B1 English level standards.

\* Adequate: The student's activity shows a basic alignment with national curriculum or CEFR B1 English level standards.

\* Good: The student's activity aligns well with national curriculum and CEFR B1 English level standards, with minor exceptions.

\* Excellent: The student's activity demonstrates an excellent alignment and understanding of national curriculum or CEFR B1 English level standards.

3. Pertinence with respect to the learning objectives:

\* No adequate: Learning objectives are unclear, unrelated to the task, or not included.

\* Adequate: Learning objectives are stated, but may not align well with the task or provide clear goals.

\* Good: Learning objectives align with the task and provide clear goals for most activities.

\* Excellent: Learning objectives are clear, directly related to the task, and set measurable outcomes for all activities.

4. Originality:

\* No adequate: The student's work is largely unoriginal, heavily relying on ideas, activities, or materials that can be commonly found in textbooks or online resources.

\* Adequate: The student's work presents some original content, but still relies on ideas, activities, or materials that can be commonly found in textbooks or online resources.

\* Good: Most activities and materials of the student's work show original thinking, with a small amount of commonly used content.

\* Excellent: All activities and materials are original, innovative, and demonstrate creative and innovative thinking.

5. Suitability and feasibility of the proposal in terms of time:

\* No adequate: The timing suggested is unrealistic, does not allow for task completion, or isn't provided.

\* Adequate: Timing is generally realistic but doesn't account for transitions, instructions, or possible delays.

\* Good: Timing is mostly realistic, with considerations for transitions and possible delays in a few activities.

\* Excellent: The suggested timing is completely realistic, allows for task completion and takes into considerations possible delays.

6. Suitability and feasibility of the proposal in terms of target:

\* No adequate: The proposal does not match the target students' age, English language level, cultural background, or special needs.

\* Adequate: The proposal partially matches the target students but lacks in one or more areas.

\* Good: The proposal matches the target students in most areas and makes minor errors in others.

\* Excellent: The proposal perfectly matches the target students' age, English language level, cultural background, and special needs.

7. Suitability and feasibility of the proposal in terms of learning context:

\* No adequate: The chosen school and museum settings are not considered or inappropriate for the proposal.

\* Adequate: The proposal shows some consideration for the settings, but could be better integrated.

\* Good: The proposal shows thoughtful integration with the settings, with minor exceptions.

\* Excellent: The proposal is perfectly integrated into the chosen school and museum settings, enhancing the learning experience.

8. Following the Willis' TBLL structure:

\* No adequate: The TBLL structure is not followed or elements are missing.

\* Adequate: The TBLL structure is attempted but lacks a clear pre-task, task cycle, or post-task.

\* Good: The TBLL structure is mostly followed with minor deviations.

\* Excellent: The TBLL structure is followed exactly, with clear pre-task, task cycle, and post-task phases.

9. Organization of the educational material:

\* No adequate: Materials are disorganized, lacking, or inappropriate for the phases of the task.

\* Adequate: Materials are organized and appropriate for some but not all phases.

\* Good: Materials are mostly well-organized and appropriate for the phases of the task.

\* Excellent: Materials are meticulously organized and perfectly suited for each phase of the task.

...



Let's assess the task created by the students according to the given guidelines:



1. Pertinence with respect to activity requirements: **Excellent**
  - The students have followed all the instructions provided by the professor. They chose the topic, museum setting, promoted skill, specified learning objectives, selected an object, designed the task following Willis' framework, and described the pre and post-museum classroom activities (see: "Topic: selfies and portraits", "Museum: National Portrait Gallery, London", "Skill to promote: speaking", "Objectives", "The activity").
2. Pertinence with respect to national curriculum, CEFR: **Good**
  - The task incorporates elements such as speaking, vocabulary building, and interactive activities that align with the B1 English level according to CEFR. However, without explicit referencing to the national curriculum or specific CEFR descriptors, it's hard to determine full alignment (see: "Objectives", "The activity").
3. Pertinence with respect to the learning objectives: **Excellent**
  - The learning objectives align well with the tasks and set clear and measurable outcomes for all activities (see: "Objectives").
4. Originality: **Good**
  - The task shows some level of originality, particularly in the use of a museum setting and incorporating portrait analysis and creation. However, use of a Kahoot quiz and discussion questions can be found in many teaching resources (see: "The activity").
5. Suitability and feasibility of the proposal in terms of time: **Adequate**
  - The timing for the tasks seems generally realistic. However, the time allotment for the classroom activities (Kahoot quiz, discussion and debate, language game, and selfie presentation) and the time for transitions and possible delays are not explicitly addressed (see: "The activity").
6. Suitability and feasibility of the proposal in terms of target: **Excellent**
  - The activities proposed are very suitable for the target group of 16-18 year-olds, B1 level English learners. The topic is age-appropriate and interesting for teenagers, and the language level seems to match a B1 level (see: "Objectives", "The activity").
7. Suitability and feasibility of the proposal in terms of learning context: **Good**
  - The proposal takes into account the museum and classroom settings. However, the suitability for the museum environment (the logistics of students creating a portrait within the museum) might present some challenges (see: "Task: Create a Portrait Gallery").
8. Following the Willis' TBLL structure: **Excellent**
  - The task follows the Willis' TBLL structure very well, with a clear pre-task, task cycle, and post-task (see: "Task: Create a Portrait Gallery").
9. Organization of the educational material: **Good**
  - The materials (portraits, Kahoot quiz, and handouts) are generally well organized and suitable for the task. However, the organization and implementation of some materials (for example, how the portraits are to be selected or created) could be further detailed (see: "Classroom preparation to the task in the museum", "Task: Create a Portrait Gallery").



Now, could you write a single textual feedback including the 9 assessments you previously gave? Try to first highlight the best parts and then talk about the improvements that could be done



Dear Students,

I am pleased to see the effort and thoughtfulness you've put into your TBLL activity design. It is commendable how closely you've followed the Willis' framework, incorporating a clear pre-task, task cycle, and post-task. Your tasks align well with your clearly defined learning objectives and are very suitable for the target group of 16-18 year-olds, B1 level English learners.

The choice of topic, "selfies and portraits," is both intriguing and age-appropriate, likely to engage your target student audience. Moreover, the integration of an actual museum visit, the National Portrait Gallery in London, is a fantastic real-world application of language learning.

Your proposal's originality shines through with the incorporation of portrait analysis and creation in the museum. These interactive components likely foster deep engagement and collaborative learning, a highly effective approach for language acquisition.

Furthermore, your inclusion of various teaching strategies, such as vocabulary building via Kahoot quiz and debates, enhances the task's comprehensiveness. The use of a post-task language game is a creative way to reinforce vocabulary, grammar, and pronunciation.

However, there are areas for potential improvement. Although the overall proposal is well-planned, the timing for each activity should be clearly specified, taking into account transitions, instructions, and possible delays. The practical aspects of creating a portrait within the museum could be further detailed to ensure smooth execution.

While you've chosen an engaging and original topic, integrating even more unique tasks or materials could elevate the quality of your activity design. Lastly, additional referencing to the national curriculum or specific CEFR descriptors might better illustrate your alignment with these standards.

Overall, your work is impressive. With a few enhancements, you are on your way to designing a truly outstanding TBLL activity.

Well done!