Castledown

OPEN ACCESS

Technology in Language Teaching & Learning
ISSN 2652-1687
https://www.castledown.com/journals/tltl/

Technology in Language Teaching & Learning, 6(2), 1–24 (2024)
https://doi.org/10.29140/tltl.v6n2.1136

The Future of Language Education: Teachers' Perceptions About the Surge of Large Language Models like ChatGPT

GIOVANNI ZIMOTTI[a]

CLAIRE FRANCES[b]

LUKE WHITAKER[c]

[a]*Spanish and Portuguese Department, University of Iowa, United States*
giovanni-zimotti@uiowa.edu

[b]*Center for Language and Culture Learning, University of Iowa, United States*
claire-frances@uiowa.edu

[c]*Center for Language and Culture Learning & and the Department of Spanish and Portuguese, University of Iowa, United States*
luke-whitaker@uiowa.edu

Abstract

This study explores the perceptions of second language (L2) educators on the surge of Large Language Models (LLMs) like ChatGPT, and their potential impact on language education. We surveyed over 100 L2 instructors, asking questions about their ideas for AI-proofing assignments, their policies, and their perceptions of how this tool will impact the profession. Data was collected through an anonymous survey and analyzed with a mixed-method approach through a constructivist lens. The results show mixed feeling: ranging from excitement for the pedagogical potential of ChatGPT to fear for potential academic dishonesty and job security in some.

The results analysis provides insight into L2 educators use and policies of ChatGPT. The quantitative data highlighted that instructors' levels of excitement or concern about the adoption of ChatGPT in language education correlate with their personal experience with the tool, in line with Bax's (2003) normalization framework. Those with prior interaction with ChatGPT showed more enthusiasm for its educational potential than those without. Interestingly, this study found no significant difference in attitudes across different age groups or years of teaching experience.

Data Availability Statement: *All relevant data are within this paper.*

The qualitative data show that L2 educators anticipate wide use of ChatGPT in their students' assignments, yet a large majority (90%) is confident in their ability to identify students use of Chat GPT and Google Translate in their work. This study also reviews the current use and policies regarding MT and ChatGPT, highlighting a significant number of instructors prohibiting their uses. The results also highlight different ways instructors have been AI proofing their assignments and how they are currently using ChatGPT in their professions. In conclusion, this study advocates for a proactive use of these tools emphasizing the importance of adopting innovative pedagogies, a student-centered classroom approach, and certain ethical considerations.

Keywords: ChatGPT, computer-assisted language learning (CALL), language education, artificial intelligence for language teaching, teachers' perceptions on AI

The Future of Language Education: Teachers' Perceptions About the Surge of Large Language Models Like ChatGPT

Large Language Models (LLMs) like ChatGPT have been described as a *disruptive technology*; that is, an innovation that significantly alters or even revolutionizes an existing industry, often introducing new and more efficient approaches, challenging methodologies, or rendering older methods obsolete (Haque et al., 2022; García-Peñalvo, 2023). LLMs are computational models that have the capability to understand and generate human language and have the transformative ability to predict the likelihood of word sequences or generate new text based on a given input. This new technology was made possible thanks to a new network architecture discovered by a group of Google employees in 2017 that substantially reduced training time (Vaswani et al., 2017). This capacity to generate human language has the potential to transform language teaching and learning. Teachers can use LLM tools to personalize learning, create syllabi, plan lessons, create quizzes or other learning activities, and to provide feedback. In L2 classes, teachers could use these models to explain grammar, propose improvements in grammar and style, and facilitate conversational practice (Kasneci et al., 2023). Students and teachers both can use LLMs to generate text summaries and outlines, and to structure academic writing. Djikstra et al., (2022) demonstrated that LLM-facilitated automatic quizzing alleviates teachers' workloads and offers students a valuable resource to practice and assess their knowledge.

These affordances are a boon to instructors, saving them time and providing personalized instructional support, and many educators are excited to try working with LLMs. But, as is common with any new technology, many others are fearful of LLMs' potential to disrupt the way that students approach classroom learning; still others fear that LLMs may render language teaching obsolete. This research sought to examine these two conflicting emotions of excitement and fear and to draw upon the lessons we've learned as a profession from our experiences with Machine Translators such as Google Translate, which has likewise been described as disruptive (Urlaub and Dessein, 2022).

Fear of new technologies is not uncommon: they require educators to adapt, which can be intimidating; educators may question their pedagogical effectiveness or bemoan the loss of human interaction. As Urlaub and Dessein (2022) noted, "when confronted with a new phenomenon, especially one that can be perceived as an existential threat, a common human reaction is to wish the problem away" (p. 52). These authors urged the profession to engage in a robust debate about how to deal with Google Translate before being "overrun tomorrow by the ongoing proliferation of increasingly sophisticated online translators and their impact on human communicative behavior." (p. 52). The time of proliferation is now, as we face the increasing ubiquity of ever more powerful and sophisticated tools like ChatGPT, one of the most popular LLMs. This study continues to advocate for having that conversation by

taking a professional pulse, examining L2 educators' perceptions and suggested solutions for policies and practices, and offering our own suggestions for embracing LLMs, not only for their ability to streamline the administrative side of teaching, but also for the room they've created for us to explore innovative, student-centered pedagogies. In taking that pulse about perceptions and attitudes, this research aligns with Bax's (2003) studies on technological normalization, or the moment when "a technology is invisible, hardly even recognized as a technology, taken for granted in everyday life" (p. 23). This process happens in stages of early adoption, skepticism, quick forays, second attempts, and alternating feelings of fear and awe, until gradually the new technology is integrated into our teaching practices. Bax (2003) maintained that to understand this process, we need to understand more than hardware and software; we need also to understand teachers' attitudes towards a new technology. To get at where we are in the process of normalization, we posed the following research questions:

- RQ1: What current attitudes and perceptions do L2 instructors have regarding ChatGPT?
- RQ2: What are innovative approaches for employing ChatGPT within language classrooms, their perceived ability to notice students' use, and their current policies regarding this technology?

Definition of Terms

Besides the previously mentioned LLM, this article uses a few technical terms that might not be commonly used in the field of language learning and teaching.

Machine Translation

"Machine translation is the process of using artificial intelligence to automatically translate text from one language to another without human involvement. Modern machine translation goes beyond simple word-to-word translation to communicate the full meaning of the original language text in the target language. It analyzes all text elements and recognizes how the words influence one another" (Amazon Web Service, Inc., n.d.). When MT is discussed throughout this paper, the term is used to refer to the early iterations of machine translators such as Google Translate.

AI Chatbots

AI Chatbots (i.e., AI conversational agents), are software that allow for human-computer interactions in which the computer produces an output of natural language based on the input from the human interacting with it. The inputs given by the human and the outputs produced by the software can both be in the form of either spoken or written language (Nicolescu & Tudorache, 2022; Dale, 2016).

Generative AI

Generative AI is an umbrella term that refers to a subset of AI technologies which can generate new content (e.g., text, images, music, video) based on the large amounts of training data that were used to build the model. This is achieved using deep learning models that provide outputs in response to simple or even highly complex human submitted prompts. Generative AI is distinct from AI chatbots in that the former can produce more than just written or spoken language. "In addition, Generative AI can generate new responses beyond its explicit programming, whereas Conversational AI [(i.e., AI Chatbots)] typically relies on predefined responses. However, not all Generative AI is conversational, and not all Conversational AI lacks the ability to generate content. Augmented AI models, such as ChatGPT, combine both Generative and Conversational AI to enhance their capabilities" (Lim et al., 2023).

When Generative AI is mentioned in this paper, ChatGPT, Stable Diffusion, Sora, Midjourney, etc. are examples of the types of software being referred to.

Literature Review

Because not many studies have been conducted on this rapidly evolving topic, we expanded the following Literature Review to include two technologies that have some similarities with generative AI and that have already had an impact on language learning and teaching: machine translation (MT) and chatbots.

Teachers' Perceptions of Machine Translation

MT use among students learning a language is widespread and has been going on for over a decade (Clifford et al., 2013). Ata & Debreli (2021) reported that 94% of learners use MT while studying English as a second language. While another study from Briggs (2018) reported that 86% of students use MT inside or outside the classroom as a language learning tool. Similar percentages of MT usage were reported by other recent studies (O'Neill, 2019; Jolley & Maimone, 2015; Niño, 2009). Some of the most common usages of MT tools includes lexical meaning of a single word (Briggs, 2018; O'Neill, 2019), written translations (Ata & Debreli, 2021), help with pronunciation (Ata & Debreli, 2021), complete an assignment (Briggs, 2018), and understanding (O'Neill, 2019). L2 teachers' perceptions of MT have been part of CALL research since at least 1995 when Anderson proposed the use of MT as a tool for L2 learning. While students use MT frequently, most language instructors do not approve (Clifford et al., 2013). The two main issues that instructors face when dealing with machine translation (MT) are identifying academic dishonesty and how to use MT to assess language proficiency (Ducar & Schocket, 2018). Henshaw (2020) observed that educators have developed a variety of approaches to dealing with MT over the years, ranging from incorporating translation activities into their teaching to outright banning it.

Teachers' Perceptions of AI Chatbots

Research on AI tools in language education often centers around chatbots. In their systematic review of conversational AI in language education, Ji, Han, & Ko, (2023) argued for the integration of conversational AI to promote students' cognitive development and decrease teachers' workload. AI chatbots and intelligent voice assistants (e.g., Google Assistant or Alexa) were found to potentially help increase students' conversational proficiency and decrease students' foreign language anxiety (Bao, 2019; Tai & Chen, 2023; El Shazly, 2021; Dizon, G, 2020). Bao (2019) and Tai & Chen (2023) found that AI chatbots designed to enhance conversational skills showed great potential in mitigating speech-related anxieties and learning barriers. But Dizon (2020) found that while students made significant gains in speaking proficiency, they did not similarly improve their listening skills, and El Shazly (2021) found that although AI chatbots could be helpful in increasing linguistic gains, they did not reduce learners' speech-related anxieties. Similarly, Çakmak (2022) found that learners had negative perceptions of and increased anxiety in working with chatbots due to difficulties in being understood. Hew et al. (2022) studied how chatbots can aid in goal setting in EFL classes; they found that learners perceived the chatbots as both useful and easy to use. Jeon (2021) found that chatbots that provide human-like interaction within a dynamic assessment framework are effective in supporting learner growth. But the implementation of AI chatbots in any educational context is still in an early stage, and there are few studies examining the effectiveness of its use in teaching and learning (Hwang & Chang, 2021).

A few studies have examined teachers' perspectives on chatbots. Chocarro et al. (2023) found that teachers are more likely to use them if they perceive them to be relevant and easy-to-use. Pokrivcakova

(2022) surveyed future teachers of English in the Czech Republic and Slovakia about their knowledge of and perceptions of using chatbots for English language teaching and found that teachers rated these tools highly on attributes such as "fun," "accurate," and "entertaining," and a little over a third said they'd integrate chatbots in teaching while close to half of respondents would *not* use chatbots in their future English classes. Belda-Medina & Calvo-Ferrer (2022) similarly studied future language educators' perceptions concerning the integration of conversational AI in language learning; like Pokrivakova, they found that although participants showed interest in learning more about chatbots, they seemed to prioritize human to human over human to chatbot communication.

Teachers' Perception of LLMs

Kasneci et al., (2023) argue that LLMs can negatively impact learner motivation, critical thinking and problem-solving. They suggest strategies like experimenting with LLMs to raise student awareness of their limitations and to critically examine AI-related content; using AI to explore multiple hypotheses rather than to provide answers; and, more generally, incorporating AI in a manner that enriches the learning experience. They similarly argue that teachers may become too reliant on AI and call on them to use AI as a supplement. Educators face an increasingly complex challenge in differentiating between texts that are generated by machines and humans. (Cotton et al., 2023; Elkins & Chun, 2020). While these concerns are valid, they might also reflect a need for evolving pedagogical and assessment methodologies to keep pace with technological advancements. The role of AI in teaching and learning, particularly regarding academic integrity, is a complex issue. Perkins (2023) discusses how LLMs like ChatGPT can be legitimate tools in supporting student education. However, the LLM output is coherent enough that often is not detected by academic staff or traditional plagiarism-detection software. Therefore, Universities must carefully construct policies to deal with student usage. This situation suggests that the concerns about AI could be partly attributed to traditional pedagogical approaches. Additionally, some plagiarism-detection software when analyzing human-generated content "exhibited inconsistencies, produc[ed] false positives and uncertain classifications" (Elkhatat et al., 2023).

Methods

To understand the perceptions and inclinations of language instructors regarding LLMs use, we grounded our research within the constructivist epistemological framework. Constructivism posits that knowledge and understanding are developed through human experiences and interactions, highlighting the subjective nature of reality (Vygotsky, 1978). This study sought to capture the diverse, subjective experiences and interpretations of language instructors, rather than presenting an objective, universal truth. This epistemological orientation informed our methodological choices and the subsequent measures we employed in our study. The axiology behind this study was to provide a "balanced representation of views [of current language instructors] to raise participants' awareness," (Mertens, 2019, p. 11). While the main data collected for this study was qualitative, to provide an overview that tried to encompass all language instructors, quantitative data were also collected.

Participants

A total of 116 individuals responded to the call for participants. Of those that responded, 35 people did not fully complete the survey and their data were not included in the analysis. Of the 81 responses remaining, 16 people responded, "No," to question 1 of the survey, "Are you a Second-Language Learning (L2) Instructor/Teacher?", which immediately led them to the end of the survey. The 65 individuals remaining are the participants of the study and they are all L2 language instructors

from across the world. The participants are mostly instructors teaching in the US. However, responses were also recorded from participants teaching in France, Germany, Belgium, Canada, Turkey, and Japan. Nine languages of instruction are represented in our sample population: French, Russian, English, Spanish, German, Portuguese, (Mandarin) Chinese, Japanese, & Korean. The age range of our participants is from 18 to 77 years old. The years of L2 teaching experience represented in our sample population ranges from less than 5 years to more than 20 years of experience. Most of our participants have experience teaching multiple levels of L2 language instruction including Elementary, Intermediate, Advanced, Literature, Culture, and Linguistics level courses.

Measures

Our survey (See Appendix A) included demographic questions about our participants (questions 1–5), questions prompting a quantitative response from participants (questions 10–14, 17, 18, & 20), as well as questions prompting a qualitative response from participants (questions 6–9, 15–16, 19, & 21). The survey is the result of a collaboration between the three researchers, which first identified and agreed upon the general topics of investigation. The general topics selected were the following: AI use and feelings among language educators, faculty perceptions of students use, ability to detect AI use in students work, current policies surrounding AI, how to AI proof assignments, and current uses for teaching. Then the researchers created questions independently for each topic. Following this, our team gathered again to jointly decide on the inclusion of questions. In this meeting, we collaboratively reviewed each survey question, deliberating on whether a Likert scale would be most appropriate or if an open-ended format might yield more insightful responses. Through constructive discussions, we resolved any differences of opinion, striving to achieve mutual agreement on the questionnaire's structure.

Before sending out the survey, we piloted it within the researchers' main institutions, and small modifications were made thanks to the feedback received; to improve the flow of the survey and to clarify some questions that had the risk of being misinterpreted.

Quantitative Measures

Eight quantitative questions were presented to participants. Participants provided answers on a 5-point Likert scale. Five of the questions (questions 10–14) required responses based on likelihood (i.e., Not at all likely, not likely, unsure, likely, & very likely). Two questions (questions 17 & 18) required responses based on excitement/worry (i.e., Not at all, some, indifferent, a little bit, & a lot) and question 20 required a binary Yes/No response from participants.

Qualitative Measures

Eight qualitative questions were presented to participants. These questions prompted participants to give long form responses of no determined length to questions asking about instructors' impressions about the likelihood their students are to use ChatGPT (or a similar AI tool); the likelihood they might notice when a student has used such a tool; ideas they might have to "AI-proof" assignments; policies they are currently using for MT; policies they are considering/have implemented related to generative AI models; approach when they notice a student has used MT or a LLM tool; impressions about whether LLMs will change the field of language instruction; and experience using ChatGPT. After all responses were recorded, the online open-source document tagging tool for qualitative data analysis, Taguette, was used to identify similar ideas throughout the participant responses. The ideas identified in the qualitative responses were used to find similarities and differences between the recorded responses. The findings from these responses were used to draw conclusions.

Procedure and Analysis

The survey used to collect data for this study was distributed to second-language instructors around the world through personal connections and the MWALLT (Midwest Association of Language Learning Technology) mailing list. The aim of the survey was to collect L2 instructors' voices to inform the field about common fears, exciting new approaches, and common strategies surrounding the use of LLM and MT software effectively in an L2 language classroom context. To understand this complex issue from the point of view of the language educators, data was collected through an anonymous survey that included a demographic section, and mostly open-ended questions with a few quantitative tokens. Data was then analyzed through the lenses of the constructivism paradigm with the goal to present a balanced representation of views.

The quantitative data collected from the survey come from the Likert scale responses participants provided to questions 17 ("How worried are you about the spread of ChatGPT and similar AI tools?") and 18 ("How excited are you about the spread of ChatGPT and similar AI tools?"). These data were cleaned, described, analyzed, and visualized using the Python 3 packages Pandas, Numpy, matplotlib, Seaborn, and OLS in the JupyterLab environment 3.2.1. Data visualizations of the qualitative data were also done using the same Python 3 packages and JupyterLab environment. Some of the visualizations were also modified using OpenAI's Code Interpreter available through GPT-4.

In the qualitative data analysis, our initial step involved a preliminary review of the data to obtain a general overview of the trends without delving into the specifics. Open-ended responses were then imported into Taguette, an online tool for collaborative analysis of qualitative data. All researchers coded responses independently using line-by-line coding. Subsequently, the team reconvened to collaboratively identify broader themes (Charmaz, 2014), any discrepancies among us were addressed through open dialogue, aiming for consensus.

Results

Quantitative Results

The research question addressed in the study using the quantitative data focused on the dependent variable of excitement/worry related to the spread of ChatGPT among the L2 instructors. The researchers looked to see if an instructor's age, teaching experience, and/or experience using ChatGPT predict how excited/worried they are about the spread of ChatGPT and similar LLMs AI tools.

All participant responses that were given via a Likert scale were dummy coded for analysis. A linear regression model was used to look for significant differences in responses to the questions, "How excited/worried are you about the spread of ChatGPT?" between the different levels of each independent variable of interest (i.e., age, years of teaching experience, and whether participants have ever used Chat GPT before).

Chat GPT Experience and Hypotheticals

To fully answer the excited/worried question, the researchers included questions asking participants to share how much experience they have had using Chat GPT. **Figure 1** below illustrates the responses to questions 19 ("Can you talk about a few ways you have used Chat GPT or a similar tool in your classroom or with your students?") and 20 ("Have you asked Chat GPT to complete an assignment, produce any sort of writing, or respond at all in the language you teach?"). Using Taguette, the responses to the more open-ended question 19 were coded to create binary Yes/No responses for ease

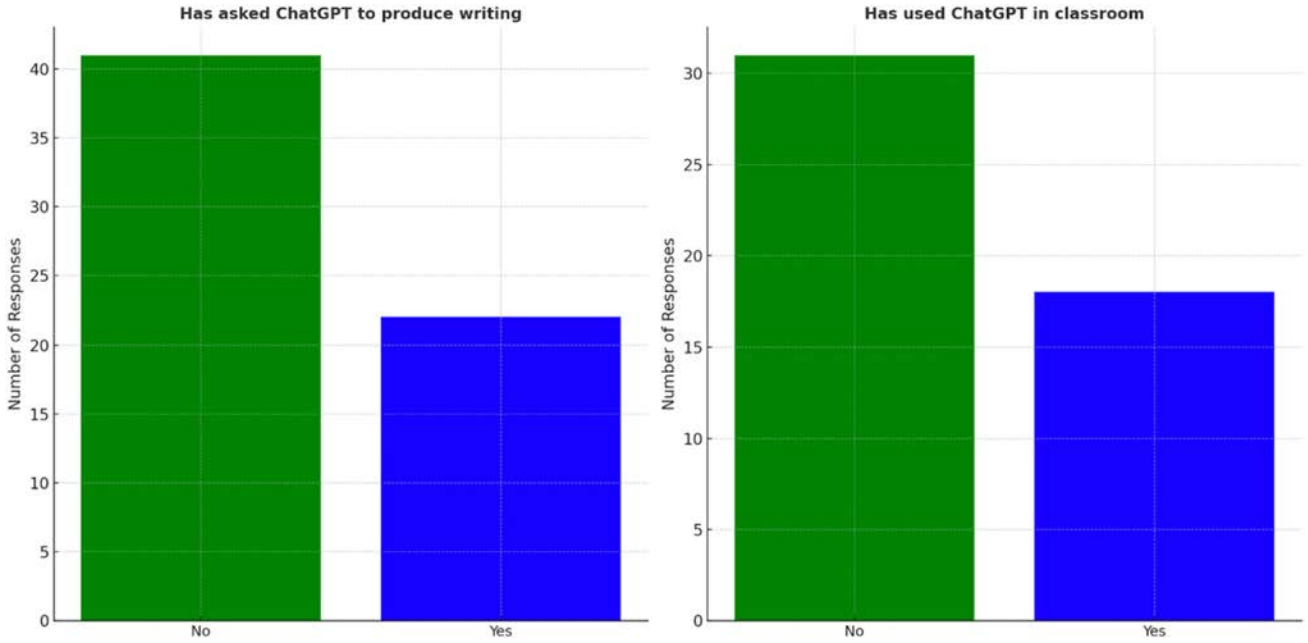

Figure 1 *L2 Instructors' Experience Using ChatGPT.*

of visualization and interpretation. While quite a few participants did have experience using Chat GPT (#19 – Yes = 18; #20 – 22 = Yes), at the time they participated in the current study, most participants had not yet used ChatGPT in either scenario (#19 – No = 31; #20 – No = 41).

In a 5-part section related to the potential integration of ChatGPT in the language classroom (questions 10–14 in **Appendix A**), participants were asked to indicate how likely they would be to implement 5 different hypothetical uses of ChatGPT in the classroom. The responses illustrated in **Figure 2** above indicate varying levels of enthusiasm and uncertainty. With respect to using ChatGPT as a conversation partner for speaking practice, opinions were almost evenly split, with 24 instructors expressing likelihood, 22 unsure, and 19 not likely to use this application. The concept of using ChatGPT to create language-learning games garnered the most positive response, with 34 instructors likely to use it, 19 unsure, and only 11 not likely. The use of ChatGPT for cultural insights related to a student's target language was also met with positive reception, with 27 likely and 22 unsure, though 16 expressed unlikelihood. The application of ChatGPT in preparing for oral exams was viewed favorably by 26 instructors, with 21 unsure and 17 not likely. Lastly, asking ChatGPT to generate discussion questions for in-class presentations was seen as likely by 31 instructors, with 21 unsure and 12 not likely.

Excited vs Worried ~ Years of Experience

The teaching experience variable has four different groups of participants represented: those with less than 5 years of experience (n = 9), those with 5–10 years of experience (n = 11), those with 10–20 years (n = 23), and those with 20+ years of experience (n = 15). When responding to the question, "How excited are you about the spread of ChatGPT?" participants gave answers on a 5-point Likert scale from "Not at all [excited]" (dummy coded as 1) to "A lot [excited]," (dummy coded as 5). The <5 years of experience group had an average response of 2.11 (σ = 1.17). The group with 5–10 years of experience gave an average response of 1.82 (σ = 1.40). The group with 10–20 years of experience gave an average response of 2.39 (σ = 1.31). And the group with 20+ years of experience gave an average response of 1.73 (σ = 1.53).

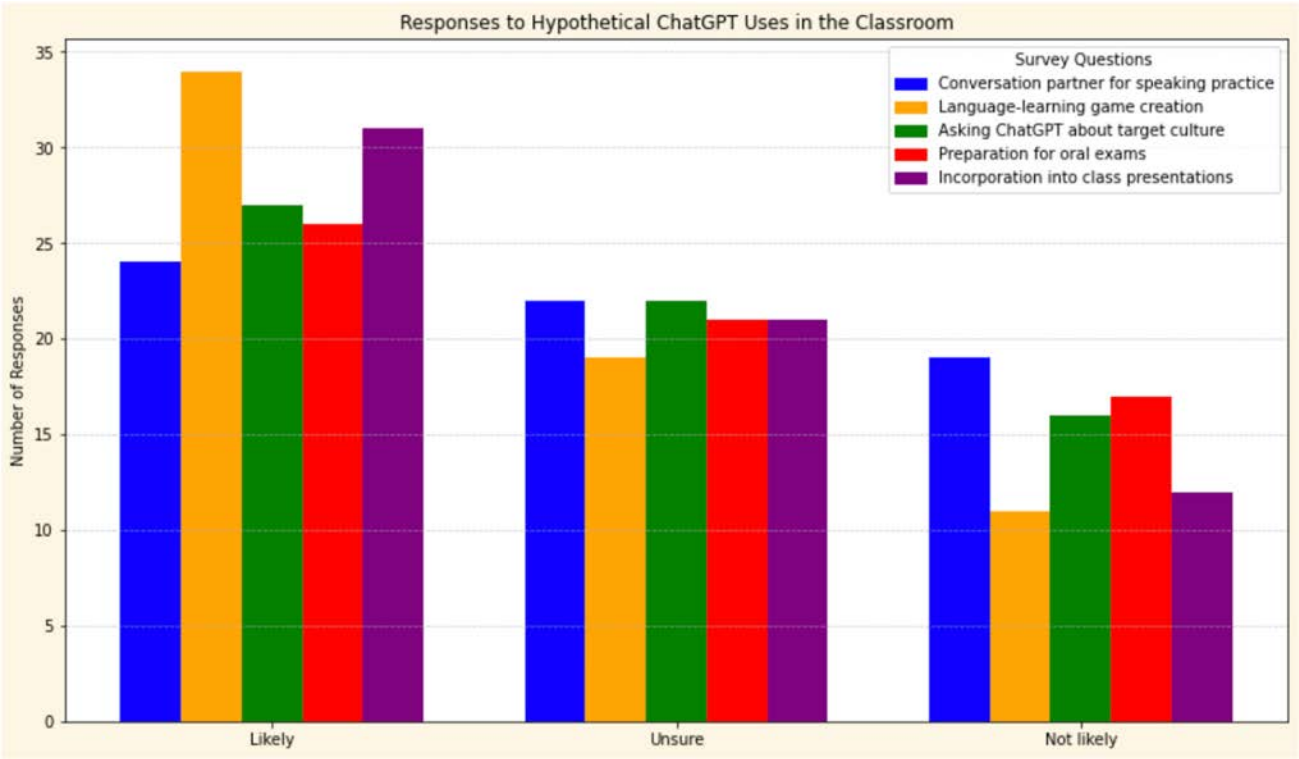

Figure 2 *Responses to Hypothetical ChatGPT Uses in the Classroom.*

When responding to the question, "How worried are you about the spread of ChatGPT?" the <5 years of experience group had an average response of 1.44 ($\sigma = 1.01$). The group with 5–10 years of experience gave an average response of 1.82 ($\sigma = 1.66$). The group with 10–20 years of experience gave an average response of 2.13 ($\sigma = 1.32$). And the group with 20+ years of experience gave an average response of 2.00 ($\sigma = 1.73$).

A linear regression model was fit to the data to find if there were any significant differences in responses to either of these questions between the participants with different years of experience. A mixed-effects model was run to see if there were any significant interactions between years of experience, age, and/or having used ChatGPT before, and no significant mixed effects were present. The only results discussed here are main effects. The linear model results looking for a main effect of years of experience indicated that, for both questions, the y-intercept coefficient (i.e., the <5 years of experience group) were significantly different from zero (worried: $\beta = 1.44$, $p = 0.005$, $R^2 = 0.045$; excited: $\beta = 2.11$, $p < 0.05$, $R^2 = 0.027$). However, none of the other categorical variable levels are significantly different from the comparison group. Therefore, the null-hypothesis cannot be rejected.

Excited vs Worried ~ Age

The age variable also has four different groups of participants represented: those who are 18–26 years old (n = 4), those who are 27–42 years old (n = 26), those who are 43–58 years old (n = 26), and those who are 59–77 years old (n = 2). When responding to the question, "How excited are you about the spread of ChatGPT?" the 18–26-year-old group had an average response of 2.75 ($\sigma = 1.26$). The group of 27–42-year-olds gave an average response of 1.96 ($\sigma = 1.28$). The group of 43–58-year-olds gave an average response of 2.08 ($\sigma = 1.41$). And the group of 59–77-year-olds gave an average response of 2.00 ($\sigma = 2.83$).

When responding to the question, "How worried are you about the spread of ChatGPT?" the 18–26-year-olds gave an average response of 1.75 ($\sigma = 1.26$). The group of 27–42-year-olds gave an average response of 1.96 ($\sigma = 1.48$). The group of 43–58-year-olds gave an average response of 2.00 ($\sigma = 1.50$). And the group of 59–77-year-olds gave an average response of 1.00 ($\sigma = 1.41$).

The results from the linear regression model showed that the responses from the 18–26-year-olds for both questions were statistically significant from zero (worried: $\beta = 1.75$, $p = 0.021$, $R^2 = 0.017$; excited: $\beta = 2.75$, $p < 0.05$, $R^2 = 0.02$). However, none of the other categorical variable levels are statistically significantly different from the comparison group. So, again, the null-hypothesis cannot be rejected.

Excited vs Worried ~ +/-used ChatGPT

The variable indicating whether participants have/have not used ChatGPT in the past is a binary variable. Participants either said, "Yes," indicating they have used ChatGPT, or "No," indicating that they have not used it before. For those participants who said they have used ChatGPT before, their average response to the question, "How excited are you about the spread of ChatGPT?" is 2.85 ($\sigma = 1.18$). In response to how worried they are, the average response of this group is 1.64 ($\sigma = 1.57$). For those participants who said they have *not* used ChatGPT before, their average response to the question, "How excited are you about the spread of ChatGPT?" is 1.66 ($\sigma = 1.28$). In response to how worried they are, the average response of the "No" group is 2.08 ($\sigma = 1.38$).

The linear regression model used the "No" category as the reference group in the analysis (excited: $\beta = 1.66$, $p < 0.05$, $R^2 = 0.18$; worried: $\beta = 2.08$, $p < 0.05$, $R^2 = 0.02$). While there was no significant difference between the two groups' responses to the question, "How worried are you about the spread of ChatGPT?" ($p = 0.288$), there was a significant difference between the two groups regarding how excited they are for the spread of ChatGPT ($\beta = 1.19$, $p = 0.001$, $R^2 = 0.18$). The coefficient here indicates that the average response related to excitement for the spread of ChatGPT from those who have used ChatGPT in the past is +1.19 compared to those who have not used ChatGPT. This indicates that those who have used ChatGPT are on average more excited about its spread than those who have not used ChatGPT.

Qualitative Results

Faculty Perception of Student Usage

This section illustrates teachers' perceptions of whether students use ChatGPT and the level of instruction at which students are more likely to use this tool. In response to the question regarding the probability of incorporating these tools into their language learning process, most survey respondents expressed a belief that students are likely or very likely to use LLMs. As **Figure 3** shows, among the respondents, ten instructors indicated that students are "very likely" to adopt ChatGPT, while thirty-five instructors believed that students are "likely" to do so. Only a small number of instructors (three) expressed a contrary view, that students would be unlikely to use it, with one attributing that unlikely to students' unfamiliarity with accessing this tool. Some respondents highlighted the prevalence of students already misusing electronic translators, leading them to predict that students would be tempted to take the path of least resistance and use LLMs like ChatGPT to complete assignments.

Survey respondents acknowledged that while students at any language proficiency level may be inclined to use LLMs, the nature of their usage might differ based on their instructional level. Lower-level

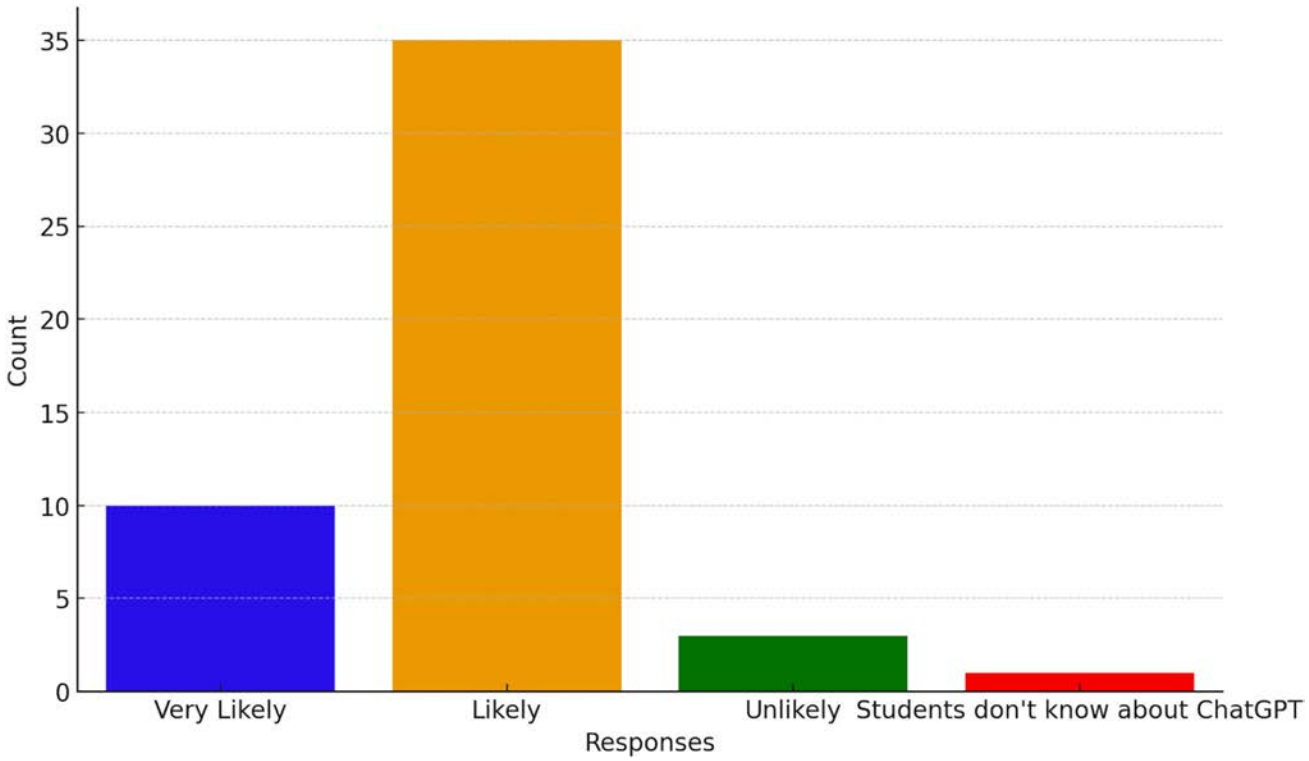

Figure 3 *How Likely do you Think Students are to Use ChatGPT?*

students might use the generated text as-is, without substantial editing, whereas more advanced students would likely use it as a foundation and then make substantial edits to the generated content. Faculty respondents see Novice and Intermediate students as more likely to embrace these tools compared to Advanced students. Instructors attributed this disparity to various motivational and situational factors, such as language requirements, lack of interest or preparation, busy schedules, and academic pressure. As one noted, "ChatGPT is a great tool for students who are not interested in actually learning the language and are looking for the fastest means of completing their work. I suspect that students who are taking literature courses are not likely to use ChatGPT because these students have a more genuine interest in proficiency and are minors or majors in the language."

In contrast, another respondent noted that "students at more advanced levels would be *more* likely to use Chat GPT to help them do their homework or even have the artificial intelligence do their homework for them. I think that from the intermediate levels onwards the use of GPT Chat would be more common. For example, for written projects such as essays or book summaries'' (emphasis added). Another noted that as the level of the language course increases, students are more likely to use ChatGPT as a tool rather than solely relying on it to create content for them, since upper-level courses often involve more complex tasks: "I assume it will also be more difficult for the AI tool to create content for more elaborate or specialized topics." While some respondents did not address the difference in levels, they did note that the likelihood of using LLMs depends on the type and complexity of the assignment.

Noticing Students' Use

Question 7 of the survey asked language teachers about their perceived ability to notice LLM and MT uses by students while completing an assignment. 90% of the respondents claimed that they would

be able to notice if a student had used ChatGPT or Google Translate. Many language educators stated that they would not have any problem identifying works created with the help of MT and LLMs. For example, a respondent said that they felt that "usually can tell when Google Translate is used," while another was very confident of their ability to notice LLMs improper use: "I have done quite a lot of playing around with Chat GPT and have noticed that it's not able to handle a really good prompt that can easily be checked for real outside sources."

Many respondents attributed their ability to notice the use of these tools in lower-level courses due to the presence of advanced structures, grammar, and idiomatic expressions usually not present in the writing of their students. "I think I am more likely to have suspicions in the lower levels since it's more obvious when there are uses of certain advanced structures and even nuances that the students couldn't possibly have mastered yet." Others credited their ability to identify the use of these tools to their familiarity with the students' work and levels. One instructor said that their students "make lots of mistakes that are predictable; I know what they don't know" while another felt that "it is fairly easy to spot misuse because my classes are small and I know my students' abilities." Additionally, a few were confident that answers generated by ChatGPT are easily identifiable because they feel "stiff, formulaic," with "unnatural structures," and are usually "vague and not matching the prompt."

While the general consensus among language instructors is that they would be able to detect the use of LLMs and MT, as the answers may use advanced grammar or be at a level that does not match the proficiency level of their students, a small minority is unsure whether they would be able to do so, especially for students taking a more advanced language classroom: "Chat GPT will be a little harder to catch in the upper levels." A high school language teacher further confirmed this by pointing out that LLMs have "good grammar and syntax and [they] write like a proficient high school student" and that if their fourth-year student used a LLM to complete an assignment, they "might struggle knowing whether it's their work on the AI's."

Another potential issue raised by respondents was the difficulty of distinguishing between text generated by a LLM or machine translated from a text created by or with the help of a native speaker. One respondent noted that "the use of structures not seen in class or particularly idioms could *have also been* done by a native speaker *and* it is difficult to know exactly which it is." Another expressed concerns about the ability to detect when heritage speakers use LLMs to complete their assignments, since "it would be much harder to notice." Respondents are also worried about the potential improvements of this technology which "accuracy is rapidly becoming more sophisticated."

Machine Translation and ChatGPT Policies

In this section, instructors were asked to discuss their policies for either Machine Translation (e.g., Google Translate) or ChatGPT. **Figure 4** presents the results of participants' responses to the types of policies they have implemented regarding these sorts of tools. Among the respondents, 28 of them have a strict policy forbidding the use of such tools in their language classes. In contrast, a smaller group of 12 instructors indicated having a more flexible stance, allowing limited use of these tools. Finally, 16 instructors reported having no specific policy related to ChatGPT. This distribution of responses highlights varying degrees of acceptance and integration of AI tools like ChatGPT in language education, reflecting diverse pedagogical philosophies and institutional preparedness for the tools at the time the survey was taken.

On the other end of the spectrum, of the sixteen who do not currently have a policy regarding LLMs, one respondent noted that "we cannot control students' behavior outside of the classroom." Another stated that tools like ChatGPT are "unavoidable, unless the instructor has their students write their

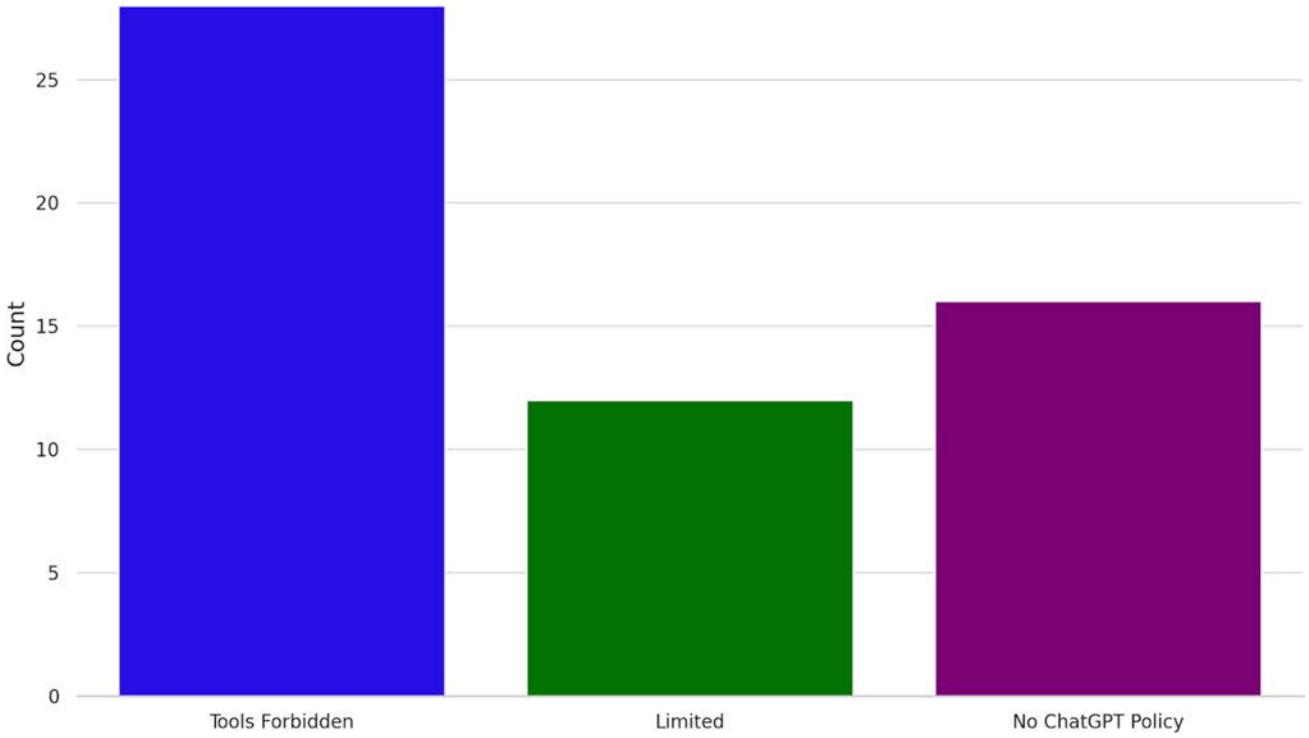

Figure 4 *Machine Translation and ChatGPT Policies Used.*

projects in class." Both respondents saw some benefit from using MT and LLM tools, when, for example, students use them to check their work and recognize or learn from their mistakes. The first of these two instructors noted that these tools can be helpful for language acquisition as long as students are learning from them; the second reflected on the fact that our students are from a "generation accustomed to technological advances and also if they want to use these tools they will do so with or without our consent."

AI Proofing

This section highlights the various ideas and approaches suggested by respondents to prevent the use of AI in writing assignments. 22 respondents professed having no idea how to AI-proof assignments or have not yet thought about how to do so, while 15 focused on plagiarism and AI-detection tools like Turnitin. Many respondents suggested more than one pedagogical approach, including allocating more class time to writing, incorporating more scaffolding in writing assignments, requiring multiple drafts and/or peer review, evaluating MT and LLMs within a writing assignment, personalizing assignments, making them more complex and analytical, or connecting to class content (see **Figure 5**). And of course, many of these strategies overlap, as illustrated by these two comments: "I need to design activities that build on each other, with some parts completed in the classroom and others at home. I think it would be hard for AI tools to manage personal reactions or substantive opinions." And: "I would be more likely to give students AI generated essays and have them annotate/analyze/improve/personalize them. I am trying to tailor assignments to topics that involve students' preferences, topics discussed in class, and add specific requirements (using specific grammar or vocabulary items)." Finally, four respondents remarked that language educators must get used to AI, since it is here to stay: "Instructors should be more familiar with the tools and teach students the right way to use them."

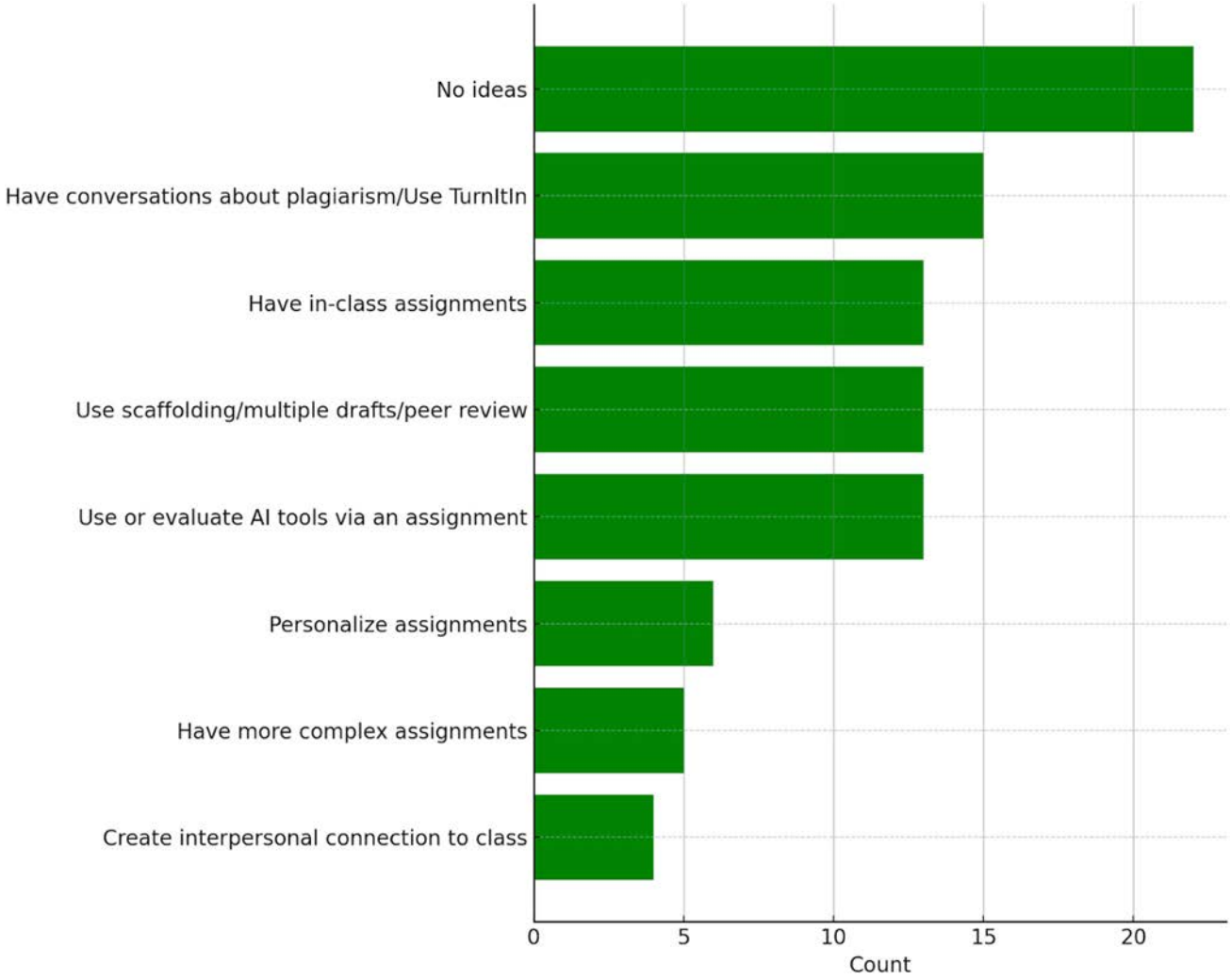

Figure 5 *AI-proofing Assignment Ideas.*

Note: Figure 5 above is an illustration of the data collected from the qualitative responses that participants gave to the question, "What ideas do you have for AI-proofing assignments?" To succinctly capture all ideas represented in the qualitative responses, two bars on the graph combine multiple responses that were similar in nature: 1) "Have conversations about plagiarism/Use TurnItIn," which combines those responses addressing plagiarism with plagiarism-detection software; and 2) "Use scaffolding/multiple drafts/peer review," connecting responses focusing on the implementation of some sort of step-by-step writing process.

Using AI

The quantitative result section, ChatGPT Experience & Hypotheticals, showed the likelihood of language instructors to use an LLM in certain predetermined scenarios. The results illustrated in **Figure 2** from the five questions (10–14) that participants responded to showed that even though there is some hesitation and uncertainty regarding implementing ChatGPT in the classroom, most participants were more likely to use some sort of classroom implementation idea than not. Q19 asked the respondents to talk about the ways they are already using LLMs in their language classrooms. Thirty-six of the respondents had already used a LLM in their classroom, both as a tool to help students learn and as a tool to help instructors get ready for teaching.

One way that emerged from the results was to use ChatGPT as a tool for discussing translation; one respondent said that they generated translations of "two different texts, one of them [being] culturally

specific, [and then had their] students analyze them to check meaning and grammatical errors and/or possible/better options." Other respondents are using ChatGPT to provide feedback on student work. "We will be using it to help correct our presentations this spring," or to provide "metalinguistic reflection opportunities." Some used them as starting points for writing assignments or during group work "as a type of teaching assistant to define terms and provide words." Finally, one respondent created "a game to enjoy and practice Spanish like Karaoke" combining Chat GPT with human interaction. No more details were given about this specific use.

While some respondents reported using ChatGPT with their students, others are "actively using it for material preparation" and find them to be "very useful." The most popular use is to create texts to share with students: "I write a lot of stories for my lesson plans and that can be tiring after a while. I only have so much creativity in me before I need to recharge! Chat GPT is a great way to get small stories I need for mini-lessons or for tests and quizzes without overtaxing my brain." Other teachers have used ChatGPT to "generate ideas for writing topics," "generate discussion questions about specific readings," and to write questions for their students. A few teachers also stated that they used them to "better understand grammar topics" before explaining them to their students or to review the "grammar of long form writing before giving [them] to their students." Finally, one respondent used Chat GPT "to create model podcasts on specific topics at specific levels [to share] as models for students."

Discussion

Faculty Adoption and Feelings

Our results concerning faculty attitudes toward and usage of ChatGPT are consistent with theories on early adoption. The term was coined by Rogers (1962) and adapted by Bax (2003) to the context of Computer Assisted Language Learning (CALL). Bax's seven stages of normalization include "early adoption," when a few educators adopt a technology out of curiosity; "ignorance/skepticism," which refers the majority of educators who don't engage at all with the new technology; "try once" educators, who try a technology but reject it because of early problems; "try again" educators who attempt a technology once more because they've heard that it works; and the "fear/awe" stage when most educators start to use a technology but are wary and/or alternatingly have exaggerated expectations. These stages are followed by "normalizing," when the technology is gradually seen as something normal) and "normalization," when it becomes completely integrated. (p. 24–5). Bax (2003) argued that the concept of normalization is "relevant to any kind of technological innovation and refers to when the technology becomes invisible and embedded in everyday practice" (p. 24).

Personal computing had become normalized by the late 1990s, but many educators still questioned whether computers should be used for language teaching. They not only feared the huge change that teaching with technology brought about, but also their own technological proficiency, and the pedagogical effectiveness of CALL, as well as the loss of human interaction. This skeptical stage gradually gave way to not whether computers should be used but how (Chapelle, 2001). In 2003, Bax claimed that most educators at that time were still at the fear and awe stage with respect to CALL. That is, they had an exaggerated fear of computers taking over the world or their profession, or, on the flip side, people seemed to think CALL has a "greater potential role in language learning than is in fact likely." This "fear/awe" binary has informed our professional responses to MT, mostly with an attitude of excessive fear/awe of being replace and a consequent call for policy action. As for educators' attitudes towards LLMs, we are in both the "ignorance/skepticism" and the "tentative trying" stages, with a few early adopters pushing forward to normalizing the integration of AI tech tools into CALL.

Sixty percent of our survey respondents, across all age groups and years of experience, indicated that they were worried about the incursions of LLMs into education, while 65%, likewise across all age groups and years of experience, expressed excitement. Our survey responses parallel Pokrivakova's (2022) findings in their study of pre-service Czech English teachers' perspectives on AI chatbots. Forty-one percent of Czech respondents had no previous experience with AI chatbots at all, but 35.9% said that they were willing to give them a try; likewise, 64% of our respondents have not previously used ChatGPT but 36% have already used them in some aspect of teaching. In both Pokrivakova's study and ours, those inexperienced users are in that very early stage in the normalization process; they have likely heard about ChatGPT and other AI tools, or attended a professional development workshop on the topic, and may have even "tried once," but they remain on the fence. Those who are willing to give these tools a try or who have already used them will be at the forefront of the normalization process, as we found that educators who had previously tried Chat GPT were significantly more excited than those who had not. This finding corresponds to Bax's normalization model as well as to Bunk et al. 's (2015) study of online teaching, which found that instructors with previous experience with online teaching were more excited about the expansion and promotion of online courses than their peers who had never taught online before.

Those educators who are already excited about the ways ChatGPT can re-energize lesson planning, instruction, and feedback methods will be early innovators, and will, like those early online teachers, be the ones who provide the profession with ideas for best practices, ways to use tools to better engage and motivate learners. The rapid sharing of information about how LLMs can help cut back on grading time, quickly respond to student queries, and provide personalized feedback and tutoring will likely see more educators try to see what LLMs can do, and they'll take more than a tentative foray. Furthermore, educators may likely have no choice in the implementation of AI tools in their instructional practices, as the decision to normalize them is made at a higher administrative level, much the way the move to online learning was. While many "unwilling adopters" were forced to teach online during the Covid Pandemic, current AI skeptics might likewise see that they have no choice, when they become ubiquitous or obligatory, or are integrated, for example, into Learning Management Systems. As LLMs increasingly inform our educational practices, tools, and methods, the process of normalization will rapidly occur.

Whether language instructors are in awe or fear of ChatGPT, AI tools are here to stay, and that rather than ignore them, we must quickly learn to use them, understand the challenges and benefits, and develop both instructional approaches and policies. If the Pandemic taught us anything, it is that we need to be agile and adaptable when faced with rapidly changing instructional conditions.

Student Use

One of the main reasons we need to get ahead of generative AI is of course that students will use it. Indeed, most survey respondents believed that students would use or have already started using ChatGPT. This perception is, in large part, based on the fact that students have been using MT tools for years (Paterson, 2023). As Ducar & Schoket (2018) note, students of language are "unable to resist the lure of easy technology-mediated strategies," (p. 781) and they use online translators despite classroom policies. They argued that the constant advancement of MT tools would continue to further challenge language instructors. The arrival of LLM tools like ChatGPT has accelerated that advancement, giving rise to new anxieties. A study conducted during Summer 2023 showed that 55% of language students at a large state university have used Generative AI (Author 1, forthcoming). And the rapid proliferation of articles and websites showing students how to use AI for language learning purposes suggests that student usage is destined to grow over time.

The fact that instructors believe that Novice and Intermediate students are more likely to use tools like ChatGPT than more Advanced students stems, perhaps, from their understanding of the way that students have been using MT. Niño credited students' dependency on MT to its wide availability, immediacy, and multilingualism, as well as its success in producing short lexical units and simply structured texts (2009). The thinking is as follows: Novice and Intermediate students often have a limited vocabulary and grasp of grammar compared to advanced students; they can benefit from the tool's ability to provide basic translations and meanings, whereas Advanced students are more likely to engage with the subtleties and nuances of a language that might challenge the capabilities of MT. However, Niño's study of advanced Spanish students revealed that 75% reported turning to MT "to obtain a quick draft to build on," and 69% declared they would use MT in the future (2009, p. 249). These numbers suggest that we are likely to see just as many Advanced students turning to ChatGPT for the same purpose; as some respondents rightly noted, easy technology-mediated strategies are appealing to students at any level.

Faculty respondents were confident in their abilities to discern MT or Chat GPT usage based on their familiarity with the students' level of instruction. Instructors are aware of what their Novice and Intermediate students can do with language, and as Frances and Zimotti (2023) showed in an experiment testing faculty ability to detect Chat GPT usage, faculty were skeptical of AI-generated text. But while it may be relatively easier for faculty to perceive whether a Novice or Intermediate student is punching above their weight, the question remains about whether they'd be able to note Advanced student usage, especially as LLMs become more sophisticated. Still, as many respondents noted, student motivation is a main factor in whether a student would choose to take shortcuts. All of this suggests, again, that we need to be proactive about developing not just policies but creative pedagogies and mechanisms to increase student motivation, especially around writing (since the main fear seems to be that students will cheat on writing assignments).

Developing Policies

Since LLMs can be misused to plagiarize work and/or cheat on assignments, it is important to develop clear policies, especially as their use becomes increasingly normalized. Many language educators have quickly adapted their policies concerning MT for LLMs. These MT policies range from banning usage outright to allowing some usage, but some instructors still don't have a policy for MT. Again, as Urlaub and Dessien (2022) argued, it is time to have a policy discussion, but it is also clear that a one-size-fits-all policy will not work for all institutions, nor is this approach advisable. Institutions or language educators should work to develop policies that fit their pedagogical needs, their students' needs, and the context of their institutions, and these policies should not only tell students what is forbidden, but also engage them in using these tools in a responsible and ethical manner. Without a doubt, policies alone will not prevent students from using LLMs in ways that are not conducive to learning. Like in the case of MT, "a clear policy with respect to authorized and unauthorized resources is absolutely necessary, but we know it is not enough to dissuade students from resorting to OTs." (Henshaw, 2020). As one survey respondent told us, these tools are unavoidable, and we can't know what students do outside of class. In fact, even IP banning these tools on the institution's network is not going to prevent students from using them, since they could access them from their phones. While having a clear policy is necessary, it is not going to be enough. The conclusion will offer some recommendations on how to teach languages at the time of Generative AI.

Limitations

While this study provides valuable insights into language instructors' perceptions of ChatGPT usage and the role of Generative AI could have in language learning, it is essential to acknowledge its

limitations. The survey was distributed through our personal networks and by using various mailing lists. While the survey was completed by more than 100 respondents, it might not have captured the sentiments and the full diversity of language instructors worldwide. This study also offers a snapshot of language instructors' perceptions at a single point in time which are based on self-reported data. Due to the rapidly evolving nature of this field, future longitudinal studies might help better explain this constantly changing field and the effects LLMs are having on language education overtime. Additionally, the survey did not include a concrete definition of AI/LLM, but we inquired about ChatGPT given the popularity of the tool at the time the survey was distributed. While piloting the study we did not receive any comments about it, however due to the growth of this technology and the proliferation of multiple LLMs (e.g., Google Bard, Facebook Llama), future studies could benefit by including clear definitions of key terms.

Empirical studies are also needed to confirm some of our findings. For example, language instructors stated that they are confident that they would be able to identify ChatGPT use in their students; these and other findings need to be confirmed by future studies testing these findings. The findings of this study might not be generalizable to instructors in other disciplines given the specific focus on language instructors. To conclude, while the findings of this study contribute to the novel and growing body of knowledge on the intersection of generative AI and language instructions, future researchers should consider these limitations while interpreting the results and plan subsequent studies.

Conclusion

The rapid and unprecedented growth of LLMs such as ChatGPT has the potential to profoundly impact language teaching and learning, whether for good or bad. It is thus important to learn how the people most affected—L2 instructors—perceive these tools.

The quantitative data collected in this study showed the instructors' levels of excitement or worry regarding the adoption of ChatGPT and similar AI tools in language education in a way that is consistent with Bax's (2003) normalization framework. The variables considered included the instructors' age, years of teaching experience, and prior experience with ChatGPT. The findings provide nuanced insights into the mindset of educators towards the integration of AI in classrooms. Again, in keeping with Bax's framework, experience with ChatGPT was found to significantly influence instructors' levels of excitement. Those who had previously engaged with the tool expressed greater enthusiasm for its wider application compared to their counterparts with no such experience. This suggests that familiarity may mitigate apprehension and foster a more optimistic outlook on the potential of ChatGPT in educational settings and will be on the forefront of moving towards normalization.

Additionally, the study found no significant evidence to suggest that age or years of teaching experience had a similar effect. Despite variations in the average levels of excitement and worry across different age and experience groups, these differences were not statistically significant. This result suggests that there is both consistent worry and excitement across all ages and years of experience for the potential of LLMs in the L2 classroom. The study also explored how educators may hypothetically use ChatGPT; there was a general openness to its application in diverse educational contexts, indicating that language educators are willing to at least try a tool. Aligning with Bax (2003), these findings suggest that while reservations exist, exposure to and interaction with ChatGPT could be a pivotal factor in shaping more positive attitudes towards the tool.

In sum, the quantitative results illuminate the complex emotional landscape that educators navigate in the face of emerging technologies. While uncertainty remains, the study offers the critical insight that

practical experience with ChatGPT could be the key to unlocking its broader acceptance, utilization, and normalization in language instruction.

Looking at the qualitative results of this study, we can conclude that L2 educators believe that most of their students will likely use ChatGPT in an educational setting, and many of them have already noticed this use in their students. However, they are also very confident (90% of the respondents) in their ability to notice ChatGPT and/or Google Translate use in their students' assignments. Their confidence in detecting potential misuse decreases with the increase of the proficiency level of the students. Moreover, a significant portion of the respondents forbids the use of LLMs like ChatGPT and of MT. In contrast, a subset of L2 educators adopts a more lenient approach allowing limited use of such tools for pedagogical purposes. Interestingly, another group of respondents lacks a formal policy.

Our results suggest that while some instructors are worried about the potential for misuse, others are acknowledging the potential benefits ChatGPT can offer to language learning. Many of our respondents do not know or have not taught yet about how to AI proof their assignments while others tend to rely on using plagiarism detection software which could lead to false detection of plagiarism. Another substantial group of the respondents propose various solutions that take advantage of research-based pedagogical approaches and personalized instructions to counteract misuse of ChatGPT. The result of our study indicates that a majority of L2 instructors are open to the idea of using ChatGPT in their teaching practices, and 36% of them have already used it in their classrooms. The potential uses highlighted by the respondents are extensive and show lots of potential for this technology. Some of the uses worth mentioning are as follows: creating lesson content, providing student feedback, generating ideas, translation analysis, and explaining complicated grammar topics to refresh their own knowledge.

Our Own Takes to Chat GPT and Language Education

We would like to conclude this article with some ideas that were shaped by the results of this study, our personal experience teaching and advising language teachers, and by the countless hours of discussion we have had about LLMs in the last few months. Bax (2011) revisited his earlier normalization framework and expanded upon it with a neo-Vygotskian approach that views this process as a culturally situated social process, developed through communication and with support. Bax (2011) underscored the importance of critically evaluating the potential value of technology through a needs audit that centers learning rather than the technology and then implementing it with a judicious use of scaffolding, modeling, and critique (p. 12).

1. **Student-Centered Classroom.** In keeping with Bax's (2011) neo-Vygotskian approach, we call for a student-centered pedagogy that emphasizes the importance of social interaction in a culturally contextualized learning process. In the context of the dual threat/opportunity of LLMs, by maintaining a close understanding of learning needs, educators can have a better understanding of why and when learners use MT or LLMs to the detriment of their learning, and they can also look for ways to integrate these tools in a way that supports learning. As we've argued, LLMs are not going to disappear, so much the same way that previous generations have had to integrate technology into their teaching and learning practices, we'll need to think about innovative ways to incorporate LLMs. By foregrounding learning in a student-centered approach, educators can model for students what effective LLMs use looks like and scaffold activities that encourage students to use these tools and/or critique their shortcomings. This may entail using a combination of in-class writing activities and targeted LLM usage; maintaining reasonable class sizes is also essential to this strategy and will help

educators build strong trusted relationships between them and their students and allow for flexible instructional approaches that cater to the diverse needs of the class.

2. **Embracing Innovative Pedagogies.** Besides creating policies surrounding the use of LLMs, incorporating LLMs and other similar technologies into our practices has the potential to enhance language acquisition. For example, adopting a task-based approach would help students understand when using a LLM is beneficial for learning, and when they should steer away from it. Combining a task-based approach with LLMs has the potential to offer a dynamic and interactive learning environment. An environment that will combine real-world tasks with immediate AI-driven feedback and interactions to make sure that students are not only learning a language but also developing confidence in using it in real-world situations. Another example would be for language educators to transition from summative to formative assessments. Instead of evaluating a learner's performance at the end of the semester or of a unit, which pushes the students to cheat trying to get a good grade, assessing our students continuously and with low-stake activities will reduce the risk of students looking for an easy way out. LLMs could be integrated into specific parts of formative assessments to provide personalized feedback to each student. This would create a responsive learning environment where errors are a learning opportunity and not a punitive measure that would lower a student grade.
3. **Forward Looking Perspective.** Bax (2011) argued that normalization is not inevitable; nor does it occur to the same degree or with the same steps with each technology; nor is it even always desirable (p. 8). While normalization may not be desirable for some, certainly for those who are nostalgic for earlier teaching methods, in the case of our current AI landscape, normalization is inevitable and taking place more quickly than perhaps any other technological adaptation in the past. Given this rapid evolution, we recommend avoiding reverting to older practices like pen-and-paper in-class essays, precisely because they might not serve our students' best interest. Such an approach might fail to prepare them for the technological skills needed to succeed in our current world and might also hinder their language learning journey by keeping them away from tools that could potentially expose them to rich but artificial input. Instead, and again, with a neo-Vygotskian approach, we recommend that educators encourage students to think critically about these tools to foster responsible and effective use. We want to promote digital literacy and equip our students with the skills that will help them take advantage of LLMs and other types of technology that might arise in the future.
4. **Ethical Considerations.** Students and language instructors should be aware of several ethical considerations when engaging with LLMs. There are concerns regarding data privacy for LLMs users, the fact that information provided might not be authentic, and it is important to be aware that using this tool might pose an inclusivity challenge, since not everyone may have equal access to these resources. Additionally, the content produced by these tools might retain biases based on the content used to train them or caused by the rules and restrictions imposed by the company that owns them. As we advance in this new phase of the digital era, there's an increasing demand for more open and transparent LLMs to ensure fairness and trustworthiness.

In conclusion, the choices language educators will make now regarding LLMs, will shape the trajectory of language learning for decades to come. It is our collective responsibility to ensure that this path is marked by progress and commitment to excellent education that it is not afraid to innovate. We would like to conclude with a final remark from one of the respondents. "We can't change the fact that ChatGPT is going to change language teaching… all we can do is adjust our sails."

References

Amazon Web Services, Inc. (n.d.). What is machine translation? Neural machine translation explained. *AWS*. Retrieved March 14, 2024, from https://aws.amazon.com/what-is/machine-translation/

Anderson, D. D. (1995). Machine translation as a tool in second language learning. *CALICO Journal, 13*(1), 68–97.

Ata, M., & Debreli, E. (2021). Machine translation in the language classroom: Turkish EFL learners' and instructors' perceptions and use. *IAFOR Journal of Education, 9*(4), 103–122.

Bao, M. (2019). Can home use of speech-enabled artificial intelligence mitigate foreign language anxiety – Investigation of a concept. *Arab World English Journal (AWEJ) Special Issue on CALL Number 5*. http://dx.doi.org/10.2139/ssrn.3431734

Bax, S. (2003). CALL–past, present and future. *System, 31*(1), 13–28. https://doi.org/10.1016/S0346-251X(02)00071-4

Bax, S. (2011). Normalisation Revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 1*(2), 1–15. http://doi.org/10.4018/ijcallt.2011040101

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences, 12*(17), 8427. https://doi.org/10.3390/app12178427

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *The JALT CALL Journal, 14*(1), 3–24.

Bunk, J., Li, R., Smidt, E., Bidetti, C., & Malize, B. (2015). Understanding faculty attitudes about distance education: The importance of excitement and fear. *Online Learning, 19*(4).

Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL (Research on Youth and Language), 16*(2), 113–131.

Chapelle, C. (2001). *Computer applications in second language acquisition (Cambridge Applied Linguistics)*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Charmaz, K. (2014). *Constructing grounded theory*. Sage.

Chocarro, R., Cortiñas, M., & Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: A technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies, 49*(2), 295–313. https://doi.org/10.1080/03055698.2020.1850426

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *@tic.revista d'innovació educativa, 10*, 108–121.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2190148

Dale, R. (2016). The return of the chatbots. *Nat. Lang. Eng*. 22, 811–817.

Dijkstra, R., Genç, Z., Kayal, S., & Kamps, J. (2022). Reading Comprehension Quiz Generation Using Generative Pre-trained Transformers. In S. Sosnovsky, P. Brusilovsky, & A. Lan (Eds.), Proceedings of the Fourth International Workshop on Intelligent Textbooks 2022: co-located with 23d International Conference on Artificial Intelligence in Education (AIED 2022) (pp. 4–17). Aachen: CEUR-WS (CEUR Workshop Proceedings; Vol. 3192).

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology, 24*(1), 16–26. https://doi.org/10125/44705

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals, 51*, 779–795. https://doi.org/10.1111/flan.12366

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal of Educational Integrity, 19*(1), Article 17. https://doi.org/10.1007/s40979-023-00140-5

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a writer's Turing test? *Journal of Cultural Analytics, 5*(2). https://doi.org/10.22148/001c.17212

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38,* e12667. https://doi.org/10.1111/exsy.12667

Frances, C., & Zimotti, G. (2023). Robots vs. humans: Does ChatGPT pose a challenge to second language writing? *The FLTMAG,* (March).

García-Peñalvo, F. J. (2023). The perception of Artificial Intelligence in educational contexts after the launch of ChatGPT: Disruption or panic? *Education in the Knowledge Society,* 24, Article e31279. https://doi.org/10.14201/eks.31279

Grundy, A. F., & Grundy, J. (1996). *Women and computers*. Intellect.

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. *arXiv preprint* arXiv:2212.05856. https://doi.org/10.48550/arXiv.2212.05856

Henshaw, F. (2020). Online translators in language classes: Pedagogical and practical considerations. *FLTMAG,* 2020(July).

Hew, K., Huang, W., Du, J., & Jia, C. (2022). Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions. *Journal of Computing in Higher Education, 35,* 1–29. https://doi.org/10.1007/s12528-022-09338-x

Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments.* 1–14. https://doi.org/10.1080/10494820.2021.1952615

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: focusing on the collaboration with human teachers. *Journal of Research on Technology in Education, 55*(1), 48–63. https://doi.org/10.1080/15391523.2022.2142873

Jeon, J. (2021). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning,* 1–27. https://doi.org/10.1080/09588221.2021.1987272

Jolley, J. R., & Maimone, L. (2015). Free online machine translation: Use and perceptions by Spanish students and instructors. In A. J. Moeller (Ed.), *Learn Languages, Explore Cultures, Transform Lives* (pp. 181–200). 2015 Central States Conference on the Teaching of Foreign Languages.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education,* Volume 21, Issue 2, 2023, 100790, ISSN 1472-8117, https://doi.org/10.1016/j.ijme.2023.100790

Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications.

Nicolescu, L., & Tudorache, M.T. (2022). Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review. *Electronics 11,* no. 10: 1579. https://doi.org/10.3390/electronics11101579

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. ReCALL, *21*(2), 241–258. https://doi.org/10.1017/S0958344009000172

O'Neill, E. M. (2019). Online Translator, Dictionary, and Search Engine Use Among L2 Students. *CALL-EJ, 20*(1), 154–177.

Paterson, K. (2023). Machine translation in higher education: Perceptions, policy, and pedagogy. *TESOL Journal, 14*(2), e690. https://doi.org/10.1002/tesj.690

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2). https://doi.org/10.53761/1.20.02.07

Pokrivcakova, S. (2022). Teacher trainees attitudes towards integrating chatbots into foreign language classes. *In INTED2022 Proceedings* (pp. 8294–8302). IATED.

Rogers, E. M. (1962). *Diffusion of innovations.* The Free Press

Tai, T., & Chen, H. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments, 31*(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801

Urlaub, P., & Dessein, E. (2022). From disrupted classrooms to human-machine collaboration? The pocket calculator, Google translate, and the future of language education. *L2 Journal, 14*(1). https://doi.org/10.5070/l214151790

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention is All You Need. arXiv:1706.03762v7.* https://doi.org/10.48550/arXiv.1706.03762

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Zimotti, G., & Panday-Shukla, P. *Exploring Students' Views, and Experiences: Artificial Intelligence in Language Learning* [Unpublished manuscript].

Appendix A

1. Are you a Second Language Learning (L2) instructor/teacher?
2. How many years of experience do you have as an L2 instructor?
3. What language(s) do you teach?
4. Which levels have you taught? Multiple answers allowed.
5. What is your age?
6. How likely do you think students enrolled in a language course would be to use ChatGPT or a similar AI tool? Please give your thoughts keeping in mind the level of the course (e.g., elementary vs literature).
7. How likely do you think it is that you would be able to notice that they've used ChatGPT or Google translate?
8. What ideas do you have for AI-proofing assignments? Have you implemented any of your ideas? If so, how did it go?
9. Do you have a policy about students using Machine Translation (e.g., Google Translate) and AI Language Models (e.g., ChatGPT)? If so, for which one? Can you summarize your policy here?
10. How likely are you to use ChatGPT as a conversation partner for students to practice their speaking skills in real-time?
11. How likely are you to create a language-learning game where students have to use their language skills to communicate with ChatGPT in order to complete tasks and earn points?
12. How likely are you to encourage students to ask ChatGPT questions about the target culture and have the model provide responses based on its training data?
13. How likely are you to use ChatGPT to help students prepare for oral exams by having the student practice answering questions in the target language with the model?
14. How likely are you to incorporate ChatGPT into class presentations by having students give a presentation on a topic and have the model ask follow-up questions to encourage critical thinking and active participation?
15. How do you usually respond to students using Google Translate or ChatGPT? Can you provide some examples?
16. Do you think that the spread of AI language models (e.g., ChatGPT) is going to change the way we teach languages and assess language competences? If yes, how? If not, why not?
17. How worried are you about the spread of ChatGPT and similar AI tools?
18. How excited are you about the spread of ChatGPT and similar AI tools?
19. Can you talk about a few ways you have used ChatGPT or a similar tool in your classroom or with your students?
20. Have you asked ChatGPT to complete an assignment, produce any sort of writing, or respond at all in the language you teach?
21. Share any additional comments you might have about this topic.