TESL-EJ Teaching English as a Second or Foreign Language

The Electronic Journal for English as a Second Language

Using GPT-3 as a More Knowledgeable Other to Provide Feedback on English Language Learners' Spoken Transcripts

August 2025 – Volume 29, Number 2
https://doi.org/10.55593/ej.29114a7

Paul Raine
Ritsumeikan University, Osaka, Japan
<paul.raine@gmail.com>

Abstract

Generative Pre-trained Transformers (GPT) have shown promise as More Knowledgeable Others (MKO) for English Language Learners (ELLs) by providing on-demand, contextualized feedback within the Zone of Proximal Development (ZPD). In this study, a web app collected spoken responses from 680 adult ELLs, automatically transcribed them and sent them to GPT-3 for corrective and elaborative improvements. Of the 428 instances that participants rated, approximately 74% were judged "very helpful," and perceived helpfulness rose with learner proficiency. Semi-structured interviews with six users offered illustrative insights into how feedback was interpreted and applied, particularly by intermediate-level learners. Although this research shows promise for utilizing GPT-3 as an MKO for ELLs, the implementation needs to be further refined and developed, and we need to continue to bear in mind the limitations inherent in GPT-based systems.

Keywords: Generative Pretrained Transformer (GPT), Zone of Proximal Development (ZPD), More Knowledgeable Other (MKO)

GPT in Context: A Technological Breakthrough

True technological breakthroughs are not as frequent as they may seem. In the realm of digital technology, we would almost certainly include the introduction of the personal computer in the 1970s, the spread of the Internet in the 1990s, and the proliferation of the smartphone in the early 2000s. Each of these developments has profoundly changed how we interact with technology and how we interact with each other through technology. The latest technological breakthrough to have such an impact has been the rapid improvement in Generative Pre-trained Transformers (GPT), particularly those developed by the Silicon Valley-based company OpenAI (Elsen-Rooney, 2023; Greteman, 2022; Hern, 2022; Karpf, 2022; Marche, 2022). This paper will focus on GPT version 3 (GPT-3), which was the latest version available at the time this research project commenced, although later versions are even more powerful.

What is GPT-3?

GPT-3 is a Large Language Model (LLM) that has 175 billion parameters and was trained on 45 terabytes of textual data, which represents a large proportion of all the text on the open Internet (Aaronson, 2022; Zhang & Li, 2021). The basic function of an LLM is prediction, i.e., "[it] must estimate the probability of a passage, or the probability of occurrence of a certain language segment" (Zhang & Li, 2021, p. 831). The ability to accurately predict the next word or segment in a textual sequence might not sound very impressive on the face of it, but it makes GPT-3 an immensely powerful and versatile tool, and it is easily able to perform a wide variety of language tasks previously considered to be unique to humans. A short list of examples includes "question answering, reading comprehension, summary generation, automatic chat, search matching, code generation, and article generation" (Zhang & Li, 2021, p.833).

GPT-3's utility for English learning

This paper will focus on the capabilities of GPT-3 that are of relevance to English language learners (ELLs). Specifically, it will describe a research project where the power of GPT-3 was embedded via its Application Programming Interface (API) in a web application (web app). The web app was developed by the researcher as both a way for ELLs to orally interact with GPT-3 and simultaneously to collect data from users about their experiences. The web app utilized GPT-3 to both correct and improve the spoken English produced by its users. In this sense, it assumes that "correction is helpful" (Dodigovic, 2007), although there has been some debate about this in the field of Second Language Acquisition (SLA), particularly when it comes to feedback on writing (e.g., Mohebbi, 2021; Truscott, 1996).

Feedback on speaking: a gap in the literature

Current research on the use of AI systems to give feedback to language learners has focused on using Natural Language Processing (NLP) tools to provide Automated Written Corrective Feedback (AWCF) on the proper use of grammatical structures (e.g., Dodigovic, 2007; Huang et al., 2020; Karlström & Lundin, 2013). The arguments in the literature are largely that AWCF is effective, notwithstanding some claims to the contrary (e.g., Truscott, 1996). There are two shortcomings in the current literature. Firstly, corrective feedback has tended to focus mainly on written tasks and non-spoken ones. Secondly, there are no known studies that investigate the potential of GPT as a More Knowledgeable Other (MKO) for giving feedback on speech produced by ELLs. This paper investigates the extent to which GPT-3 can act as an MKO for ELLs by correcting and improving their spoken transcripts, and in doing so, tries to address these shortcomings.

Scope and aim of this paper

The hypothesis and core argument here is that GPT-3 can act as an effective MKO if certain criteria are met. The argument is important because of the significant number of English learners worldwide, estimated to range from 1.5 to 2 billion (Bentley, 2022). However, the relative scarcity of qualified English teachers poses a significant challenge in educating these learners. Furthermore, the limitations of human English teachers make it incredibly challenging to provide highly personalized feedback to each student individually. To set out the core argument of this study, the researcher first presents a literature review of the relevant theoretical frameworks and research done in similar studies. Next, the research design of the current study is described, followed by the presentation of the results. Finally, there is a discussion of the results and a tentative conclusion. It is hoped that this paper and its conclusions will be of interest to Second Language Acquisition (SLA) and FLT (Foreign Language Teaching) scholars, especially those with an interest in Computer Assisted Language Teaching (CALL) or Technology Enhanced Language Learning (TELL).

Literature Review

This study relies on elements of the theoretical framework known as Sociocultural Theory (SCT), which was first developed by Lev Vygotsky in the 1930s, although it has been significantly adapted and updated since then. Vygotsky (1978) was not directly concerned with how adults learn second or foreign languages. Rather, his focus was on how children learn and how languages (and other systems of signs) are used to mediate the learning process (McLeod, 2023; Vygotsky, 1978).

This study relies primarily on two specific concepts associated with SCT, namely the Zone of Proximal Development (ZPD) and the More Knowledgeable Other (MKO). Although SCT is a much deeper and broader theory than the ZPD and the MKO alone, these two concepts have survived almost a century and have been applied repeatedly across multiple domains.

The seminal passage from Vygotsky (1978) explaining what the ZPD is and how it works goes as follows:

> "[The ZPD] is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers." (Vygotsky, 1978, p.86)

Vygotsky (1978) employs a "flower bud" metaphor to analogize the ZPD: "[the ZPD] defines those functions that have not yet matured but are in the process of maturation, functions that will mature tomorrow but are currently in an embryonic state." (Vygotsky, 1978, p.86)

When we think of scaffolding, we might think of the steel poles used by builders when constructing a new development. With Vygotsky's theories, scaffolding is anything that assists the learner to advance to the next level of their ZPD, or as Donato (1994) puts it:

> "[a] situation where a knowledgeable participant can create supportive conditions in which the novice can participate, and extend his or her current skills and knowledge to higher levels of competence" (p.40).

It is the responsibility of the MKO to guide the learner to the next stage of the ZPD. In SLA, this is achieved in a variety of ways. For example, in research conducted by de Almeida Mattos (2000), it was shown how a more competent student guided a less competent student to the correct answer through collaborative dialogue. Even something as subtle as the MKO's "raised eyebrow" can be interpreted as a signpost by the learner that a mistake has been made and a correction should be attempted (Kinginger, 2002). Other kinds of "scaffolding" traditionally adopted in SLA contexts include a variety of oral prompting, "gap fill" activities, L1 glosses for L2 materials, and subtitles on videos (Hosogoshi, 2016).

Several neo-Vygotskian scholars have stretched the boundaries of Vygotsky's theories, with some claiming that the idea that the MKO must be a human adult or peer who is physically present in the classroom is a "stagnant view" of the concept (Cicconi, 2014, p.58) while others point out that the MKO is often "wrongly believed to be only a teacher" (Baleghizadeh et al., 2011, p.43).

A more inclusive conception of the ZPD, the MKO, and the "scaffolding" metaphor might be:

> "[the use of] supportive templates by which guidance is offered to the students through a semiotically mediated situation in order to achieve higher level competence and regulation." (Baleghizadeh et al., 2011, p.44).

In other words, textual materials that have been "designed by an expert" (Baleghizadeh et al., 2011, p.45) can fulfill the role of the MKO and help guide a learner through their ZPD. There

are several advantages to conceptualizing the MKO in this way. The most obvious of these would be using artificially intelligent MKOs to provide personalized feedback to learners on a scale and at a speed unattainable by human teachers.

In the field of SLA, Dodigovic (2007) showed how an artificially intelligent tutor was "an efficient instrument of error remediation" for learners of English as a Second Language (ESL) of multiple nationalities. While Dodigovic (2007) does not explicitly acknowledge Vygotsky's theories, they can be inferred in her theoretical framework and research design. Additionally, in their 2013 study, Karlström and Lundin found that "NLP [Natural Language Processing] in the form of syntax highlighting seems to be a relevant tool for language learners, provided that they are given tasks corresponding to their current proficiency level" (Karlström & Lundin, 2013, p.426). In a more recent study, Ekstrand et al. (2023) highlight the different roles that can be played by a human MKO in collaboration with artificially intelligent information systems. Even though the information system itself is not considered an MKO, their study shows how AI can support the MKO in facilitating learning.

Studies such as these provide a basis for treating AI as an MKO that can help the learner reach the next level of their ZPD. Research relating specifically to GPT is rare, but Rudolph et al. (2023) support the idea that GPT can be treated as an MKO in academic fields such as medicine, law, and aviation. They also highlight how Intelligent Tutoring Systems (ITS) can "simulate the assistance provided by a tutor" (p.9). Rudolph et al. (2023) do not look specifically at the use of GPT in SLA, and so that is a gap the current paper attempts to address.

Research Design

Research Question

The research question addressed in the current report is:

> To what extent can a Generative Pre-trained Transformer (GPT)-based system act as a More Knowledgeable Other (MKO) for English language learners (ELLs)?

Research site

This study was administered entirely online with a combination of a GPT-enabled web application, which is described in more detail below, and interviews conducted using a synchronous online video conferencing tool (Zoom).

Instrument design

To allow users to interact orally with GPT-3 in a way that is compatible with Vygotsky's theories, and at the same time to collect data for this study, the researcher developed a web app (Figure 1). The process of using the web app was straightforward, and the interface was designed to be as simple as possible to allow low proficiency English speakers to use it with ease.

Figure 1. The web app *Welcome Screen*

The web app *Welcome Screen* (Figure 1) displayed a randomly selected question to the user and allows them to change it by clicking a button if they desire. The user was then required to answer the question in less than 30 seconds by clicking the picture of the microphone. Their response to the question was recorded, and after recording, they could listen to their attempt before submitting it for evaluation (Figure 2).

Figure 2. The web app *Audio Review Screen*

After reviewing their recording, users clicked on either the "Try Again" button or the "Continue" button. If they clicked on "Try Again" they were taken back to the *Welcome Screen* to re-attempt the question or choose a different question. If they clicked on "Continue" their audio recording was uploaded for transcription and evaluation.

Once a user's audio recording had been uploaded to the server, several processes were triggered. In the first step, the audio recording was sent to a cloud-based Automatic Speech Recognition (ASR) service for transcription. The app made use of AssemblyAI's cloud-based "Universal" speech-to-text model with the "Global English" language setting. This model has a 93.4% word-level accuracy rate for English recognition. When considering non-native English, with its higher proportion of challenging features for standard ASR, a recent study conducted by McGuire (2025) showed that AssemblyAI's Universal model performs very well, with a word-level accuracy rate of 90.4% for a corpus consisting entirely of spontaneous non-native English speech.

Once the transcription was obtained, it was sent to OpenAI's GPT-3 API endpoint for analysis. The following prompt was used to retrieve an "improved" version of the audio transcript:

> "An English learner attempted to answer the following question: <QUESTION>. This was their response: <ANSWER>. In less than 100 words, re-write their response in standard academic English, expanding it where necessary with additional details, examples, or reasons."

The question that the user selected was inserted into the <QUESTION> section, and their transcribed answer was inserted into the <ANSWER> section. Providing the question to the GPT-3 endpoint gave it more context and helped generate more relevant responses.

By including expansion alongside correction, the wording of the prompt was intended to scaffold richer language to show the learner how to make more elaborate arguments and provide supporting evidence. Finally, a synthesized speech recording of the improved answer was also retrieved from OpenAI's cloud-based text-to-speech (TTS) service. An email was then sent to the user to inform them that their feedback was ready, and the feedback was presented to them as shown in Figure 3.

Figure 3. The web app *Feedback Screen*

On the *Feedback Screen*, the user could listen to their own answer and then read and listen to the model answer. In order to further illustrate a typical user response and GPT-improved version of a transcript, an example response and improvement are provided below. This example was not included in the dataset and was recorded by the researcher himself:

Question:

> *What games are popular in your country?*

User Response:

> *In my country, soccer is most popular support. Soccer is good. Everybody wants to be good. Soccer Player.*

GPT-Improved Version:

> *In my country, soccer is the most popular game. It is very popular amongst all ages, and everyone has a favorite team they support. Soccer is a great sport that promotes team building and fair competition. It also encourages physical activity and provides an opportunity for people to come together and form friendships. Additionally, people idolize soccer players and aspire to become like them. It is no surprise that soccer is the most popular game in my country.*

In the final part of the process, users were encouraged (but not required) to rate the suggestion by using a Likert scale at the bottom of the *Feedback Screen,* which allowed them to answer the question "Was this suggestion helpful" with options ranging from 1 (No, not at all) to 5 (Yes, very helpful). This data is collated and further analyzed below.

Participant selection and recruitment

At the time this paper was written, 680 users had signed up for the web app. Users were invited to register for the app by sharing the link through social media platforms such as Facebook and LinkedIn. The only requirements were for respondents to be English language teachers or learners and at least 16 years of age. A convenience sampling method was adopted, i.e., "nonrandom sampling where members of the target population … meet certain practical criteria, such as easy accessibility, geographical proximity, availability at a given time, or the willingness to participate" (Etikan et al., 2016). In the current research project, geographical proximity was not a limiting factor because all the data was remotely collected online.

The users came from a variety of national and cultural backgrounds and shared 47 different first languages between them. When signing up for the web app, users were required to provide their email, their name, their gender, their age, their English proficiency level, and their first language. In terms of proficiency level, users self-reported by selecting one of the following levels (from high to low): Proficient (4%), Advanced (6%), Upper -intermediate (16%),

Intermediate (31%), Pre-intermediate (25%), and Beginner (18%). As for the first languages of users, the vast majority of users (62%) came from a Turkish language background. Other numerically significant first languages included Azerbaijani (10%), Russian (5%), English (3%), Vietnamese (2%), Arabic (2%), and Spanish (1%), with other languages comprising the remaining 15%.

Follow-up interviews

As well as interacting with the app and responding to the Likert scale question each time feedback was generated, the respondents were also approached to participate in follow-up interviews. Of the six respondents who accepted this invitation, four were teachers, and two were learners of English. Further generic details of these respondents are summarized in Table 1, below:

Table 1. Details of follow-up interview participants

Respondent Number	Role	Age Group	Gender	English Proficiency
1	Student	16-20	Female	Upper-intermediate
2	Teacher	56-60	Female	Proficient
3	Student	41-45	Female	Intermediate
4	Teacher	61-65	Male	Proficient
5	Teacher	41-45	Female	Advanced
6	Teacher	56-60	Female	Proficient

The interviews were semi-structured, and the core questions used are provided in Appendix 1. There are two sets of questions that vary slightly depending on whether the interviewee was a teacher or learner of English. The set of questions for teachers had one extra question that probes respondents on their knowledge of Vygotsky's theories and how they might apply to the web app. The interviews were recorded, and the recordings were automatically transcribed with another cloud-based ASR service (Otter.AI). The transcripts were then analyzed and thematically coded according to the "theoretical" thematic analysis procedure laid out by Braun and Clarke (2006), which is "driven by the researcher's theoretical or analytic interest in the area" (Braun & Clarke, 2006, p.84).

All the interviews were conducted online using a synchronous online video conferencing tool (Zoom). The time and word length details of each of the interviews is provided in Table 2, below.

Table 2. Time and word length of follow-up interviews

Respondent Number	Interview length	Transcript length
1	14 mins	1326 words
2	23 mins	2844 words
3	14 mins	1466 words
4	17 mins	2638 words
5	22 mins	3666 words
6	10 mins	1752 words

Ethics

For the quantitative section of this research, which comprised minimally invasive Likert-scale ratings for AI feedback, and anonymized demographic data, participants provided consent via a standard English-language form embedded in the web app. In addition, a standard "terms and conditions" page describing data collection and retention policies was made available. Because modern browsers offer automatic page translation (for example, Chrome's "Translate this page" feature), it was assumed that non-native speakers could easily translate this information if they struggled to understand it, although this was not verified. For the qualitative section, participants were provided with a more detailed consent document before participating in the interviews, which outlined the study's aims, interview procedures, and data storage policies.

Results

Quantitative evaluation

In total, 680 respondents submitted 939 recordings via the web app, of which 428 feedback instances (45.6%) received a rating. Among those 428 rated responses, 316 (73.8%) were marked "very helpful," indicating a broadly positive reception among raters. Note that the question "Was this suggestion helpful?" was kept intentionally simple to accommodate users with varied English proficiency.

The researcher next examined whether perceived helpfulness varied across self-reported English levels. After merging ratings (n = 428) with proficiency data, participants were assigned to three groups:

- Beginner (Beginner + Pre-intermediate, n = 151)
- Intermediate (Intermediate + Upper-intermediate, n = 173)
- Advanced (Advanced + Proficient, n = 104)

The researcher calculated the proportion of "very helpful" (rating = 5) in each. Beginners rated 67.5% of feedback instances as very helpful, Intermediates 73.4%, and Advanced learners 85.3%. A chi-square test of independence confirmed a significant association between proficiency level and helpfulness ratings ($\chi^2 = 33.72$, $p = .028$), indicating that more proficient users generally found the feedback more helpful than less proficient users.

Qualitative evaluation

The following major themes were derived from a combined analysis of all six interview transcripts.

Advantages and disadvantages of AI

Given the human propensity to weigh up the pros and cons of any new technology (especially in comparison to human teachers), it comes as no surprise to see that all the respondents touched on at least once on what they perceived to be the good and bad points of the app.

One significant advantage of the app, for example, was seen to be its ability to offer steady, day by day study opportunities for English learners:

> "[The students'] motivation increases, step by step. And I like it very much that it this progress is very steady every day." (Respondent 2)
>
> "[It provides] better scaffolding for more repetitive daily type of stuff, which I can't do when I'm only seeing people for an hour a week." (Respondent 3)

Another advantage that a couple of respondents touched on was the feeling of the app acting as a "personal coach" (Respondent 2) and offering more variation in response at a faster rate than humans could hope to achieve (Respondent 4).

In addition, some of the respondents touched on the convenience of being able to use the system in a time and place that suits them, for example:

> "I use this application [on] my mobile phone. And it's available everywhere. And when I have just five minutes or 10 minutes, time I use it" (Respondent 3).

Of course, some disadvantages of the app were also mentioned. For example, the problem of using such a system with elderly or computer illiterate users who may not understand how to use the system (Respondent 4) or the inability of AI to have a genuine emotional reaction to an interaction (Respondent 5) or indeed the temptation for learners to use such systems to plagiarize their homework assignments (Respondent 6).

The ZPD and the MKO

Since respondents were encouraged – either explicitly or implicitly – to express their views on the viability of GPT-3 as a More Knowledgeable Other, this was also a significant theme in the transcripts. Overall, the respondents seemed to be open to the idea that Vygotsky's ideas were still relevant and applicable to modern methods of technology-enhanced learning, with Respondent 2 stating that she believed "Vygotsky would be astonished by how his ideas can fit the present-day situation."

Respondent 4 touched on the idea that AI is potentially better at creating scaffolding for students than teachers are, because it can automatically create textual passages that are slightly different every time, and unique for each student: "That's going to outdo what I can do and do it faster as well" (Respondent 4).

Regarding whether the revised transcripts provided by GPT-3 were inside or outside the learner's ZPD, most respondents seemed to believe that they were inside the ZPD:

> "I think the responses are right above the student's level, so they give her some room for development." (Respondent 2)
>
> "Sometimes I feel it's [at my level] but sometimes it's just above my level." (Respondent 3)

In addition, one respondent pointed out that the system is better for higher-level students:

> "I thought it was better suited to higher level adult students. I didn't try it with any of my child students, for example." (Respondent 4).

Language learning strategies

The final numerically significant theme encompasses the variety of ways that respondents felt that English language learners might utilize the feedback provided by GPT-3 to improve their English.

Respondent 1, for example, mentioned her use of an English dictionary to study the words in the improved version of the transcript, and claimed that she likes to "read aloud while listening [to the transcript]". Reading the improved transcript aloud was also a technique favored by Respondent 2, who stated that she "make[s] [her students] read aloud, imitating the intonation of the speaker."

Respondent 3 talked about her attempts to integrate the phrases or pronunciation of the improved transcript into her daily life: "I try using these sentences or pronunciations in my daily routine and make sentences like that."

Memorization of the improved transcript was another strategy mentioned by the respondents, with some encouraging this technique and others discouraging it:

> "If they want to memorize it, I think it's probably good" (Respondent 5).
>
> "I'd be afraid for them to sort of memorize… I always highly discourage that" (Respondent 6).

Most of the respondents, however, agreed on the need to "absorb", "digest", or "reproduce" the improved transcript in some way, with one respondent stating that "just listening" to the transcript one time would not be an effective strategy (Respondent 2).

Discussion

Quantitative evaluation

Quantitatively, 316 out of 428 feedback ratings (73.8%) were "very helpful," showing that those who chose to respond found the suggestions beneficial. However, because fewer than half of all sessions (428 / 939 = 45.6%) yielded a rating, this figure applies only to respondents and cannot be assumed to generalize across the entire user base. Prior work indicates that optional, in-app surveys often achieve response rates of 10–30% and are prone to self-selection bias (Fan & Yan, 2010; Groves & Peytcheva, 2008). Thus, our "73.8% very helpful" reflects only those who elected to rate; the true helpfulness distribution among all users may differ. Future work should also explore methods to increase rating participation or compare observable behaviors of responders versus non-responders to assess and mitigate potential bias.

The stratified analysis revealed that more proficient users reported the highest helpfulness rates (85.3%), with lower rates for Intermediate (73.4%) and Beginner (67.5%) groups. One possible explanation is that advanced learners, who are already more familiar with the language, were better able to interpret and apply nuanced suggestions, thereby perceiving them as highly useful, while beginners may have struggled to understand or fully implement the feedback.

Qualitative evaluation

The qualitative analysis of the interview transcripts revealed three major themes: the advantages and disadvantages of AI (compared to human teachers), the ZPD and the MKO (our theoretical framework), and language learning strategies.

It is very natural and understandable that humans want to weigh up the pros and cons of any new system and compare artificially intelligent technologies to their own intelligence. We can see that some of the respondents touched on the same advantages of AI systems as were mentioned in some of the studies from the literature review, in particular the key point that AI applications have "exceptional potential for improving intelligent student support systems and scaffolding student learning in adaptive and personalized ways" (Rudolph et al., 2023, p.350 citing Zawacki-Richter et al., 2019). Giving personalized feedback to more than a handful of students at a time is a practical impossibility for all but the most zealous teachers. But even for those teachers who manage this considerable task, it is not the end of the story. Not only do they need to provide personalized feedback, but they also need to provide it in a way that helps the learner to advance within their ZPD.

This leads us directly onto the second theme to emerge from the analysis of the transcripts, which centers on the respondents' thoughts and opinions about the ZPD and the MKO. As Respondent 4 points out, AI applications can potentially provide level-appropriate feedback to learners faster and better than a human teacher could. But even for AI, this is a tricky process. The web app in this study provided quite high-level feedback. Respondent 4 picked up on this point: "I thought it was better suited to higher level adult students."

It is deemed likely that some of the improved transcripts introduced vocabulary or grammatical structures that fell too far outside the learner's current ZPD, thereby overshooting and becoming less accessible, especially for beginners. The feedback process could be improved by giving lower proficiency level learners more narrowly focused corrections. There is some evidence from the literature review that "less is more" when it comes to ZPD scaffolding, and we know that even a "raised eyebrow" can constitute effective scaffolding in some cases. Although the respondents in the current study ostensibly felt that the feedback was within their ZPD (evidenced by their ratings of helpfulness), it could be made to align more closely with the user's current developmental level. Future versions of the system could pass the user's proficiency rating through to the prompt so the AI would have a better understanding of how to more accurately tailor the feedback to the user's level. Providing feedback in the user's L1 would also be a possibility given the multi-lingual capabilities of the latest GPT models.

Finally, we come to learning strategies. The use of digital technologies such as the web app could itself be considered a learning strategy. But more frequently than not, the use of such technologies still needs to be combined with the use of more traditional language learning strategies in order to be fully effective. Some of the strategies mentioned by the respondents include checking new words in a dictionary, reading aloud or memorizing the improved version of the transcript, and trying to integrate the suggested phrases into their daily lives.

While the interviews conducted for this study highlight how teachers and learners perceived and leveraged the feedback, the sample size (n = 6, four of whom are teachers) is too limited and self-selected to support generalizable conclusions about overall benefits or ZPD effects. Instead, these findings should be viewed as preliminary observations that generate hypotheses for future work.

Conclusion

This report is a preliminary attempt to evaluate the extent to which a Generative Pre-trained Transformer (GPT)-based system can act as a More Knowledgeable Other (MKO) for English language learners (ELLs) by providing feedback on spoken transcripts. If we accept the neo-Vygotskian view that an MKO does not have to be a teacher, or even a human present in the classroom, and that Vygotsky's theories can be applied in ways he would not necessarily have envisaged, then it seems that GPT-based systems such as the app developed for this study can indeed act as an MKO for ELLs to a significant extent. However, certain criteria must be met for these systems to be effective.

Firstly, the feedback from GPT must be far enough but not too far above the learner's current level. Including the learner's spoken transcript in the prompt can provide GPT with an indication of their current level, but it might benefit from further context and biographical data about the learner in order to better tailor its responses. Secondly, having GPT provide the learner with a revised and improved version of their spoken transcript might not in itself be sufficient to bring about noticeable improvements in learners' English-speaking ability. They may need to apply one or more additional language learning strategies to more fully benefit from that feedback. Finally, we should remain conscious of the potential shortcomings of artificially intelligent systems, including the inability of AI to form genuine long-term relationships and emotional connections with humans. In addition, learners will need to continue to resist the temptation to use GPT-based systems to avoid learning, while using it instead to assist or enhance that process.

About the Author

Paul Raine is an English language teacher and web-app developer. After earning a CELTA at York English Language Center in 2006, he relocated to Japan to begin teaching English. He

then completed an MA in TEFL at the University of Birmingham in 2012 and an M,Res in Technology-Enhanced Learning at Lancaster University in 2024. He currently teaches at Japanese universities while researching and developing CALL (Computer Assisted Language Learning) and TELL (Technology Enhanced Language Learning) tools. **ORCID ID: 0000-0003-3763-0981**

To Cite this Article

Raine, P. (2025). Using GPT-3 as a more knowledgeable other to provide feedback on English language learners' spoken transcripts. *Teaching English as a Second Language Electronic Journal (TESL-EJ), 2*9(2). https://doi.org/10.55593/ej.29114a7

References

Aaronson, S. (2022). My AI Safety Lecture for UT Effective Altruism. Retrieved from https://scottaaronson.blog/?p=6823

Baleghizadeh, S., Timcheh Memar, A., & Timcheh Memar, H. (2011). A sociocultural perspective on second language acquisition: The effect of high-structured scaffolding versus low-structured scaffolding on the writing ability of EFL learners. Reflections on English Language Teaching, 10(1), 43-54. https://www.nus.edu.sg/celc/wp-content/uploads/2022/11/43to54-baleghizadeh.pdf

Bentley, J. (2022). Report from TESOL 2014: 1.5 Billion English Learners Worldwide. Retrieved from https://www.internationalteflacademy.com/blog/report-from-tesol-2-billion-english-learners-worldwide

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cicconi, M. (2014). Vygotsky meets technology: A reinvention of collaboration in the early childhood mathematics classroom. Early Childhood Education Journal, 42, 57-65. https://doi.org/10.1007/s10643-013-0582-9

de Almeida Mattos, A. M. (2000). A Vygotskian approach to evaluation in foreign language learning contexts. ELT journal, 54(4), 335-345. https://doi.org/10.1093/elt/54.4.335

Dodigovic, M. (2007). Artificial intelligence and second language learning: An efficient approach to error remediation. Language Awareness, 16(2), 99-113. https://doi.org/10.2167/la416.0

Donato, R. (1994). Collective scaffolding in second language learning. In J.P. Lantolf and G. Appel (eds) Vygotskian Approaches to Second Language Research (pp. 33–59). Norwood, NJ: Ablex.

Ekstrand, M. D., Pera, M. S., & Wright, K. L. (2023). Seeking Information with a More Knowledgeable Other. Interactions, 30(1), 70-73. https://doi.org/10.1145/3573364

Elsen-Rooney, M. (2023). NYC education department blocks ChatGPT on school devices, networks. Chalkbeat. Retrieved from https://ny.chalkbeat.org/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American Journal of Theoretical and Applied Statistics, 5(1), 1-4. https://doi.org/10.11648/j.ajtas.20160501.11

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. Computers in human behavior, 26(2), 132-139. https://doi.org/10.1016/j.chb.2009.10.015

Greteman, B. (2022). ChatGPT Can Write Better Essays Than My College Students. That's a Good Thing. Newsweek. Retrieved from https://www.newsweek.com/chatgpt-can-write-better-essays-my-college-students-thats-good-thing-opinion-1769136

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public opinion quarterly, 72(2), 167-189. https://doi.org/10.1093/poq/nfn011

Hern, A. (2022). AI-assisted plagiarism? ChatGPT bot says it has an answer for that. The Guardian. Retrieved from https://www.theguardian.com/technology/2022/dec/31/ai-assisted-plagiarism-chatgpt-bot-says-it-has-an-answer-for-that

Hosogoshi, K. (2016). Effects of Captions and Subtitles on the Listening Process: Insights from EFL Learners' Listening Strategies. JALT CALL Journal, 12(3), 153-178. https://doi.org/10.29140/jaltcall.v12n3.j206

Huang, H. W., Li, Z., & Taylor, L. (2020). The Effectiveness of Using Grammarly to Improve Students' Writing Skills. In Proceedings of the 5th International Conference on Distance Education and Learning (pp. 122-127). https://doi.org/10.1145/3402569.3402594

Karlström, P., & Lundin, E. (2013). CALL in the zone of proximal development: Novelty effects and teacher guidance. Computer Assisted Language Learning, 26(5), 412-429. https://doi.org/10.1080/09588221.2012.663760

Karpf, D. (2022). Money Will Kill ChatGPT's Magic. The Atlantic. Retrieved from https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-chatbots-openai-cost-regulations/672539

Kinginger, C. (2002). Defining the zone of proximal development in US foreign language education. Applied linguistics, 23(2), 240-261. https://doi.org/10.1093/applin/23.2.240

Marche, S. (2022). The College Essay is Dead. The Atlantic. Retrieved from https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/

McGuire, M. (2025). Automatic speech recognition for non-native English: Accuracy and disfluency handling. arXiv preprint arXiv:2503.06924. https://doi.org/10.48550/arXiv.2503.06924

McLeod, S. (2023). Vygotsky's Sociocultural Theory of Cognitive Development. Retrieved from https://www.simplypsychology.org/vygotsky.html

Mohebbi, H. (2021). 25 years on, the written error correction debate continues: an interview with John Truscott. Asian-Pacific Journal of Second and Foreign Language Education, 6(1), 3. https://doi.org/10.1186/s40862-021-00110-9

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. Journal of Applied Learning and Teaching, 6(1). https://doi.org/10.37074/jalt.2023.6.1.9

Truscott, J. (1996). The case against grammar correction in L2 writing classes. Language learning, 46(2), 327-369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International Journal of Educational Technology in Higher Education, 16(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, M., & Li, J. (2021). A commentary of GPT in MIT Technology Review 2021. Fundamental Research, 1(6), 831-833. https://doi.org/10.1016/j.fmre.2021.11.011

Appendix 1: Interview Questions

For all respondents:

Are you a teacher or a learner of English?

For teachers:

1. How do you feel about your students getting advice on their English from an AI? Do you think your students prefer to get feedback from human or AI teachers?
2. Do you feel that the suggestions given by the AI are below, at, or above your students' current level?
3. Do you feel that the suggestions given by the AI are helpful to improve your students' English?
4. How do your students use the suggestions given to improve their English?
5. Have you ever heard of Vygotsky, the Zone of Proximal Development, or the More Knowledgeable Other? If so, what do you think about these concepts?
6. Is there anything else you would like to say about the web app system?

For learners:

1. How do you feel about getting advice on your English from an AI? Do you prefer to get feedback from human or AI teachers?
2. Do you feel that the suggestions given by the AI are below, at, or above your current level?
3. Do you feel that the suggestions given by the AI helpful to improve your English?
4. How do you use the suggestions given to improve your English?
5. Is there anything else you would like to say about the web app system?

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.