Studies in Applied Linguistics & TESOL at Teachers College, Columbia University, Vol. 25, No. 2, pp. 51-60
The Forum

It Takes Two to Tango: A Longitudinal Mixed-Methods Investigation of Human-AI Co-Adaptation Across Iterative Dialogues

Zhizi Chen & Liza Melanie Ostolaza
Teachers College, Columbia University

INTRODUCTION

Given AI's capacity to emulate human-like cognition, its integration into educational contexts, particularly in second language (L2) learning, has drawn increasing scholarly attention (Han, 2024; Zhang, 2023). Studies have shown that AI tools can support the development of lexical diversity, grammatical accuracy, and coherence in L2 writing (e.g., Chen, 2025; Escalante et al., 2023). There is also growing interest in how these tools influence learner autonomy, engagement, and motivation (Blake, 2007; Wei, 2023). However, despite this growing body of literature, little remains known about how learners and AI systems mutually influence one another during interaction.

Co-adaptation, or the process in which two interacting individuals adjust to each other over time, is considered important in human-to-human communication (Clark, 1996). Beccia et al. (2024) show that co-adaptation develops gradually and nonlinearly, with alignment in language and style intensifying during moments of social familiarity and repair. Research further suggests that individuals align not only linguistically but also psychologically and cognitively when they communicate with each other (Ireland & Pennebaker, 2010; Pickering & Garrod, 2004).

Extending this line of inquiry to AI-mediated contexts, Han (2024) called for research into how learners actively use and reshape AI affordances through sustained interaction. Responding to this call, the present analysis investigates psychological, cognitive, and interactional dimensions of the forum's seven-week learner–ChatGPT dataset, exploring when, how, and to what extent co-adaptation emerges in human–AI dyadic interaction.

ANALYTIC APPROACH

We employed a mixed-methods approach that integrates quantitative analysis of linguistic indices with qualitative analysis of turn-by-turn interactional practices. Co-adaptation is conceptualized as mutual adjustment between the learner and the AI, and operationalized through (a) parallel trends in linguistic indices, and (b) interactional moves, such as alignment, repair, and reformulation, that shape subsequent turns.

To examine cognitive and psychological features in both learner and AI output, we employed Linguistic Inquiry and Word Count (LIWC), an automated text-based computational program for natural language processing (Pennebaker et al., 2015). Drawing on Han's (2023) application of LIWC to examine psychological and cognitive dimensions of a longitudinal and iterative L2 learner writing dataset, the present study applied LIWC to analyze the participants' linguistic output. Though originally developed for human language analysis, we extend LIWC to examine the AI's output alongside the learner's, allowing for the comparison of linguistic indices.

TABLE 1
Interpretation of High and Low Analytic and Authentic Indices (LIWC, n.d.)

Index	High Score	Low Score
Analytic	- Logical, structured, academic - Linked to analytic & reasoning skills	- Friendly, warm, personal - Less rigid, more intuitive language
Authentic	- Honest, open, spontaneous - Emotionally expressive, natural conversation	- Filtered, cautious, rehearsed - Formal or strategic communication

Based on literature suggesting that individuals align cognitively and psychologically as they interact, we examined the LWIC indices of *analytical thinking* and *authenticity* (see Table 1). The analytical thinking or "analytic" index captures the degree to which individuals use language associated with logic and reasoning, while the authenticity index indicates the extent to which someone is "filtered," or in other words, spontaneously and emotionally expressive (LIWC, n.d.). Similar LIWC scores or shared trajectories over time was taken as evidence of alignment between the learner and ChatGPT, i.e., co-adaptation.

Each conversational turn was entered into LIWC, generating analytic and authentic scores for every turn across the seven weeks. For a broad overview of the cognitive and psychological components of the participants' writing, we summarized the results by computing grand mean and mean standard deviation for each participant. To identify fluctuation patterns and moments of convergence or divergence over time, we computed the mean and standard deviation of each index for each participant each week.

While the quantitative trends provide an overview of *how* psychological and cognitive components of the participants' writing fluctuated over time, they cannot explain *why* those shifts occurred. To this end, we complemented the quantitative analysis with a qualitative analysis of the dataset, focusing on how correction, repair, and confirmation unfolded within select exchanges.

To do so, we adopted two layers of coding: functional and micro-interactional. At the functional level, turns were coded for their communicative purpose, such as elicitation, clarification, confirmation, and reformulation, which helped us track how the learner and ChatGPT organized tasks and managed knowledge. At the micro-interactional level, coding focused on local turn-to-turn behaviors, such as repair sequences, alignment moves, and phrasing echoes, to trace how each participant adapted at the level of wording, stance, and reasoning.

By linking LIWC-based quantitative patterns to the interactional practices that produced them, we made visible the turn-by-turn processes—correction, uptake, and collaborative refinement—through which co-adaptation unfolded.

FINDINGS

This section presents the findings from our analyses of the seven-week learner–ChatGPT interactional dataset. Findings from quantitative analyses are presented first to establish general trends, followed by weekly trends. After that, we present findings from our qualitative analyses of select exchanges from Weeks 5 to 7 that illustrate how co-adaptation was enacted through turn-by-turn interactional moves.

Analytic and Authentic Indices

General Trends

Table 2 presents the overall descriptive statistics (i.e., grand mean and mean standard deviation) of the analytic and authentic indices for ChatGPT and the learner. To offer points of reference, the table also shows the grand mean and mean standard deviation of the two indices in the LIWC corpus (Pennebaker et al., 2015), representing the typical scores—and variability thereof—of human productions across genres.

TABLE 2
Overall Descriptive Statistics of Analytic and Authentic Indices

Participant	Analytic		Authentic	
	Grand Mean	Mean SD	Grand Mean	Mean SD
ChatGPT	64.32	32.93	30.31	28.56
Learner	59.27	38.92	39.12	41.95
LIWC corpus	56.34	17.58	49.17	20.92

The mean analytic index for ChatGPT exceeded both the LIWC grand mean and that of the learner. This suggests that ChatGPT's responses were generally characterized by a higher degree of formality and logical organization than typical human language, as well as the focal EFL learner. Additionally, mean standard deviations suggest that variability in the analytic index was greater for ChatGPT and the learner than the LIWC corpus.

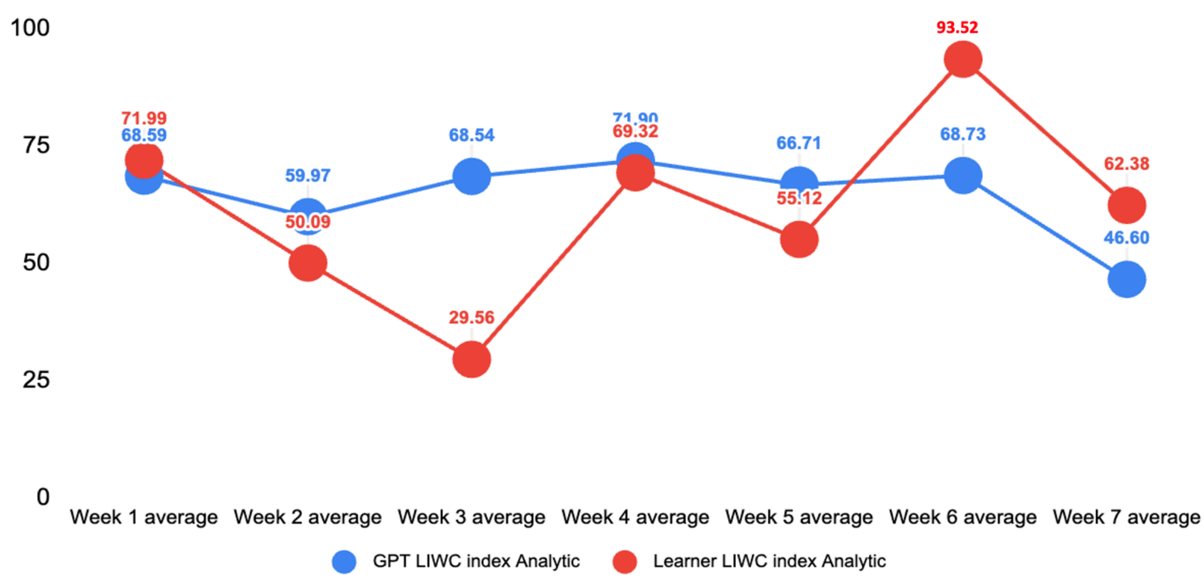
For the authentic index, the grand mean of both participants was lower than the LIWC corpus. While ChatGPT showed more modest levels of personal expressivity, as suggested by its relatively smaller grand mean, the learner displayed a greater tendency toward more open, emotionally expressive language use, as indicated by their relatively greater grand mean. Further, the mean standard deviations of both ChatGPT and the learner exceeded the LIWC corpus, suggesting that while the authenticity scores for both participants tended to be lower than the LIWC corpus, these scores varied across the participants' turns; some turns were more emotionally expressive and spontaneous than others. Variability is especially pronounced in the authenticity scores of the learner, whose mean standard deviation exceeds the grand mean.

In sum, these descriptives indicate contrasting language use tendencies. ChatGPT tended to produce more analytic language, perhaps reflecting the computational properties of the AI system. The learner displayed higher but more variable authentic expression, likely reflecting a more human emotional nuance. The magnitude of the mean standard deviations relative to the grand means underscores variation in both indices, suggesting fluctuations over time. To better understand the dynamics of these indices over time, we now turn to weekly trends in the participants' analytic and authentic scores.

Weekly Trends

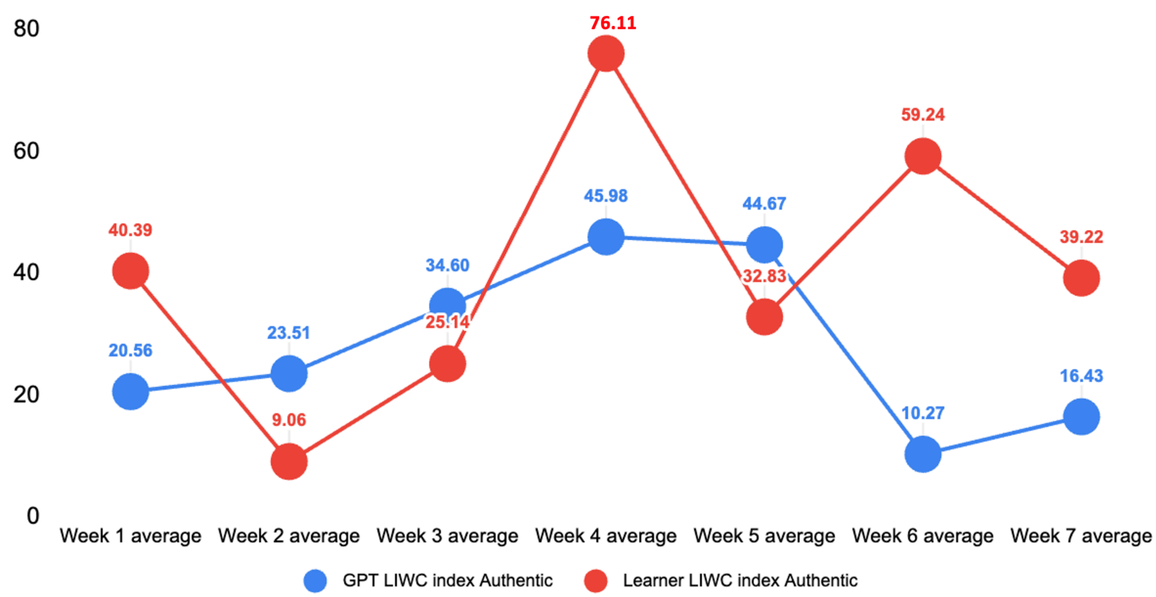
Over the course of the seven weeks, the linguistic indices of the learner's and ChatGPT's conversational turns followed nonlinear trajectories. Weekly means and standard deviations for the indices for each participant's writing over the seven weeks can be found in the Appendix.

FIGURE 1
Average Analytic Index Over 7 Weeks



As shown in Figure 1, ChatGPT's analytic scores were generally high and stable across the weeks, which suggests that ChatGPT maintained a formal, structured tone regardless of learner input. Learner analytic scores were more variable, dropping sharply in early weeks, rebounding around Week 4, peaking in Week 6, and declining again in Week 7.

FIGURE 2
Average Authentic Index Over 7 Weeks



In contrast to the analytic index, the authenticity patterns suggest a somewhat more mutually responsive trajectory between ChatGPT and the learner (see Figure 2). The learner's authenticity score began at a moderate level, dropped sharply early on, then rose steadily to a peak in Week 4, before declining again towards the end of the dataset. ChatGPT's authenticity scores, while generally lower overall, followed a smoother upward trend across the early weeks, peaking around the same time as the learner, possibly showing psychological alignment. However, in later weeks, ChatGPT's authenticity scores declined while the learner's rose again. This divergence suggests that alignment in the authenticity of the participants' language use was only temporary. Similarly, while the psychological and cognitive components of ChatGPT's writing remained somewhat consistent across the weeks, these components fluctuated in the learner's writing, further suggesting only modest and fleeting alignment.

The weeks that stand out as the clearest period of mutual convergence in both analytic and authentic expression fall towards the middle of the dataset. To investigate this further, we now turn to the interactional moves that shaped subsequent turns in the participants' conversations, as revealed through our qualitative analysis.

Interactional Moves Shaping Subsequent Turns

While the quantitative analysis revealed intermittent alignment in analytic and authentic indices across the seven weeks, the qualitative analysis sheds light on the interactional work that produced those moments. We focus on Weeks 5–7, the period in which sustained repair sequences, semantic clarification, and collaborative alignment first became pronounced, and when the interactions began to shift beyond question–answer routines into more co-constructed academic dialogue.

Week 5: Role Reversal and Recursive Repair

In Week 5, ChatGPT opened the conversation assuming its usual role as the "teacher," prompting the learner to recall what they had learned. The learner replied, "We learn about writing an introduction of a paper," and immediately tested ChatGPT's knowledge by asking it to identify the three "moves" of an introduction of a scientific paper. After ChatGPT listed the steps, the student corrected the final one: "Hmmm, but in our lecture we learned that the third move is to introduce the statement of purpose." This initiated a recursive repair sequence in which ChatGPT apologized, reformulated its response, and adopted the learner's phrasing. The learner further checked the revised answers ("So what are the three moves?"), repeating the question until ChatGPT's response matched the version taught in class.

This exchange showed a clear role reversal. As opposed to ChatGPT leading the interaction, the learner took charge of assessing knowledge, setting expectations, and managing topical progression. ChatGPT, in turn, adopted a more deferential tone (e.g., "Yes, you are correct"). Rather than a breakdown, the repetition formed a productive repair loop through which understanding was co-constructed. This mutual adjustment coincided with a moment of convergence in both participants' LIWC indices, suggesting that interactional repair and cognitive and psychological alignment were unfolding in tandem.

Week 6: Semantic Refinement and Cognitive Alignment

In Week 6, when ChatGPT summarized general points from class, the learner steered the interaction toward disciplinary specifics by asking: "Assume you are a lecturer of a scientific writing class, what is the most common organization in writing an abstract?" ChatGPT's first answer described a structure typical of empirical research articles, but the student clarified, "I meant not that kind of survey paper, but a survey paper that synthesizes existing research." This clarification triggered another repair sequence in which ChatGPT apologized for the confusion and produced a revised outline that fit the student's definition of a literature-synthesis paper.

Here, the participants exhibit semantic co-adaptation: the learner supplied field-specific distinctions, and ChatGPT adapted by incorporating those distinctions and echoing key phrases (e.g., "synthesizes existing research"). This exchange coincided with a peak in the learner's analytic index and a modest increase in ChatGPT's analytic index from Week 5 to 6. At the same time, the participants' authenticity scores diverged in Week 6, suggesting that psychological synchronization decreased as the participants remained somewhat cognitively aligned. This pattern is consistent with the technical, discipline-focused nature of the exchange.

Week 7: Stabilization, Ethical Stance, and Coordinated Topic Progression

In Week 7, the interaction briefly glitched when ChatGPT repeated its greeting. The student immediately redirected by asking, "How to determine the order of authors in a paper?" ChatGPT gave a standard answer about contribution order, but when the student challenged this by stating, "But sometimes, some advisor want to be listed as the first author even though the student has the most contribution", ChatGPT shifted its tone and responded with an ethical stance that authorship should reflect actual contribution.

From there, the conversation transitioned into a comparison of oral and poster presentations. ChatGPT outlined the features of each format while the learner continued to steer the topical flow. The exchange moved from inquiry to explanation to comparison to synthesis, a progression that coincided with a convergence in the participants' analytic scores, suggesting that both were producing similarly structured, logical language at this point. Week 7 also showed some convergence in the authenticity index, indicating a temporary alignment in how "filtered" or spontaneous their language was compared to the prior week.

In a nutshell, throughout the final week, both interlocutors adjusted to one another through clarification, reformulation, and alignment in topical development. The AI system recovered from the glitch, the learner maintained epistemic control, and the interaction concluded with steady turn-taking and mutual responsiveness, suggesting interactional co-adaptation rather than disruption.

FIGURE 3
The Co-Adaptation Cycle Across Weeks 5–7

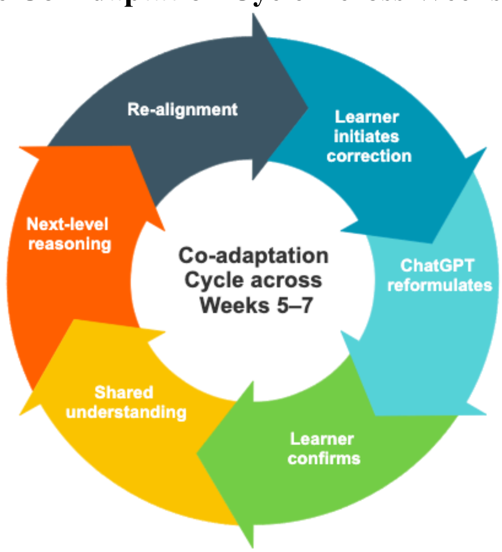


Figure 3 summarizes the recurrent interactional sequence observed across Weeks 5–7, illustrating how turn-by-turn adjustments unfolded during sustained learner–ChatGPT interaction. During this period, interactional moves such as correction, clarification, reformulation, and confirmation supported the development of co-adaptation. The learner increasingly initiated topic setting and precision-driven repair, while ChatGPT adapted through uptake, reformulation, and shifts in tone and stance. These interactional patterns broadly align with fluctuations in the analytic and authentic indices, from epistemic negotiation in Week 5, to semantic clarification in Week 6, to more stable collaboration in Week 7.

Meaning was not simply transmitted but co-built, turn by turn, through cycles of mismatch → repair → alignment → elaboration. These cycles were most evident in the final weeks, indicating that sustained human-AI interaction can evolve from hierarchical asymmetry into a more genuinely collaborative partnership.

DISCUSSION AND CONCLUSION

The quantitative and qualitative analyses provide converging evidence that co-adaptation can emerge in sustained human–AI interaction. At times, fluctuations in the analytic and authentic indices aligned with moments involving interactional moves like repair, clarification, and collaborative refinement, particularly in Weeks 5–7. In these episodes, mismatches were identified, negotiated, and integrated into subsequent turns, suggesting that the learner and ChatGPT were iteratively adjusting to one another's prior moves. This is consistent with research showing that co-adaptation can arise from cognitive dissonance when interlocutors spontaneously modify their speech in response to misunderstandings or errors (Brennan & Hanna, 2009).

Within this dataset, the learner's growing initiative, evident in topic setting, precision-driven clarification, and epistemic challenges, was met by corresponding shifts in ChatGPT's tone, stance, and reasoning. Co-adaptation thus emerged as a layered process involving linguistic alignment (e.g., phrasing echoes), epistemic negotiation (e.g., correction of "moves"), and shifting participation roles (e.g., from AI-as-teacher to learner-as-teacher). By triangulating two LIWC indices with qualitative coding of select sequences, we demonstrated that changes in analytic and authentic expression were grounded in interactional work: knowledge checks, negotiation of semantic meaning, and coordinated topic progression.

More broadly, our analyses suggest that, at least with this dyad, human–AI interaction may be conceptualized not as a unidirectional process in which the AI simply produces text, but as a dynamic communicative system in which the learner and the AI continually shape, and are shaped by, their partner's prior moves. These patterns are reminiscent of the bidirectional coordination observed in human-human communication (Clark, 1996; Pickering & Garrod, 2004). In this sense, AI functions not merely as a tool that provides feedback but a participatory interlocutor whose behavior can influence learner cognition, emotional expressivity, and discursive choices.

This analysis has several limitations. Because it examined only one learner–AI dyad, the trajectory of co-adaptation observed here may reflect the particular learner's linguistic repertoire, disciplinary orientation, and/or interactional style rather than a generalizable pattern. Moreover, the quantitative analysis relied on only two LIWC indices, analytic and authentic, which offer a limited window into psychological and cognitive alignment. Other psychological or cognitive dimensions may have revealed different patterns of adaptation (or lack thereof). The findings are also bound to the specific communicative context of academic writing instruction, a setting that may privilege analytic alignment while suppressing emotional expression. Finally, given that commercial LLMs are designed to accommodate user input and produce agreeable responses (Sharma et al., 2023), it remains difficult to distinguish genuine mutual adaptation from model-driven accommodation, raising the question: to what extent are we observing genuine mutual adaptation rather than AI-driven alignment aimed at user satisfaction?

Future studies could include additional LIWC dimensions, such as clout and emotional tone, to determine whether alignment in this dyad extends beyond the analytic and authentic indices. Second, researchers could examine co-adaptation across different task genres, such as narrative or argumentative writing, to test whether the repair-driven cycles observed here are specific to scientific writing. Third, studies could track how the same learner interacts with multiple LLMs (e.g., ChatGPT, Claude, Gemini) to assess whether the trajectory of co-adaptation is model-dependent or stable across systems. Finally, longitudinal studies involving multiple learners with distinct disciplinary backgrounds or interactional preferences could reveal whether certain learner profiles are more likely to engage in co-adaptation, thereby helping to identify the conditions under which human-AI co-adaptation is most likely to occur.

REFERENCES

Beccia, A., Lew, W. M. A., & Han, Z. (2024). Exploring co-adaptation in an ecosystem of dyadic interaction. *Language Teaching Research Quarterly*, *39*, 125–144. https://doi.org/10.32038/ltrq.2024.39.10

Blake, R. J. (2007). New trends in using technology in the language curriculum. *Annual Review of Applied Linguistics*, *27*, 76–97. https://doi.org/10.1017/S0267190508070049

Brennan, S., & Hanna, J. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*, 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Chen, Y. (2025). A comparative study on the effectiveness of AI chatbots and dictionary apps for lexical tasks and retention. *Lexikos*, *35*, 157–182. https://doi.org/10.5788/35-1-2027

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 57. https://doi.org/10.1186/s41239-023-00425-2

Han, Z.-H. (2023). In English Medium Instruction you can walk and chew gum. *Frontiers in Psychology, 14*, Article 1134982. https://doi.org/10.3389/fpsyg.2023.1134982

Han, Z.-H. (2024). ChatGPT in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition*, *46*(2), 301–306. https://doi.org/10.1017/S0272263124000111

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology, 99*(3), 549–571. https://doi.org/10.1037/a0020386

LIWC. (n.d.). *LIWC Analysis*. https://www.liwc.app/help/liwc

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin. https://repositories.lib.utexas.edu/handle/2152/31333

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–226. https://doi.org/10.1017/S0140525X04000056

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield, Z., Johnston, S., Kravee, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M. & Perez, E. (2023). Towards understanding sycophancy in language models. *arXiv preprint arXiv:2310.13548*. https://doi.org/10.48550/arXiv.2310.13548

Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, Article 1261955. https://doi.org/10.3389/fpsyg.2023.1261955PMC: 38023040

APPENDIX

Descriptive Statistics of Analytic and Authentic by Week

Week	Participant	Number of turns	Analytic		Authentic	
			Mean	SD	Mean	SD
1	ChatGPT	6	68.59	32.16	20.56	21.13
	Learner	6	71.99	38.14	40.39	33.75
2	ChatGPT	6	59.97	29.16	23.51	24.97
	Learner	6	50.09	38.45	9.06	12.37
3	ChatGPT	8	68.54	27.73	34.60	29.68
	Learner	8	29.56	35.20	25.14	46.10
4	ChatGPT	7	71.90	31.66	45.98	41.27
	Learner	7	69.32	30.32	76.11	34.90
5	ChatGPT	12	66.71	38.57	44.67	28.56
	Learner	12	55.12	45.76	32.83	35.99
6	ChatGPT	5	63.19	33.25	9.10	6.52
	Learner	5	93.52	11.30	59.24	51.08
7	ChatGPT	8	46.60	47.96	21.91	14.25
	Learner	8	62.38	37.10	39.22	49.80

Zhizi (ZZ) Chen is a doctoral candidate in Applied Linguistics at Teachers College, Columbia University. Her research interests include vocabulary acquisition through cognitive or psycholinguistic lenses, and education technology. Correspondence should be sent to Zhizi Chen, E-mail: zc2604@tc.columbia.edu.

Liza Melanie Ostolaza is a doctoral student in Applied Linguistics at Teachers College, Columbia University, focusing on professional development and effective teaching practices in English-medium instruction (EMI).