Effects of Dynamically Computer-Mediated Communication and 'Pseudo' Communication on L2 Learning

Leander S. Hughes (hughes@mail.saitama-u.ac.jp)
Saitama University, Japan

Abstract

This study introduces an online text-based language learning application offering a novel kind of computer mediation which is 'dynamic' in that learner output is monitored and revised to optimize learning. Two experimental versions of this application are compared: a 'chat' version which administers a text-based synchronous computer-mediated communication task and a non-communicative 'note' version through which learners take notes individually on target content that are subsequently integrated into questions on that content. The latter version is referred to as 'pseudo' communicative because it requires learners to essentially send messages to themselves through taking notes that they must later read and comprehend in order to successfully complete the task. The heightened communicative accountability of the pseudo communicative task is hypothesized to induce a level of need to engage the target language that is more comparable to true communicative tasks than conventional output-only tasks. Findings reveal a significant advantage in learning efficiency for the communicative task, but no significant difference in absolute gains between tasks. Additionally, a significant relationship was found between delayed gains in knowledge of lexical meaning and the amount of output produced by learners in both groups. It is posited based on the Involvement Load Hypothesis that the lack of a significant difference in absolute gains was due to an increased sense of need to engage the target language induced by the pseudo communicative task compared to conventional non-communicative tasks.

Keywords: Interaction; Output; Synchronous computer-mediated communication (SCMC); Second language acquisition (SLA); Involvement Load Hypothesis (ILH)

Introduction

Research in second language acquisition (SLA) has investigated the effect of interaction on language learning (Loewen & Sato, 2018), including a substantial number of empirical studies attempting to disentangle the effects of negotiated input and output (e.g., de la Fuente, 2002, 2003; Ellis & He, 1999; Ellis *et al.*, 1994; Golonka *et al.*, 2017; Loschky, 1994; Mackey, 1999; Tare *et al.*, 2014). Meanwhile, advances in technology along with the onset of the COVID-19 pandemic have led to widespread adoption of synchronous computer-mediated communication (SCMC) for language teaching and learning (Junn, 2021). Text-based SCMC (TSCMC) in particular, has emerged as a convenient, yet innovative way to investigate the role of interaction in SLA (Hughes,

2022; Lin *et al.*, 2013). The present study introduces an online application which provides computer mediation that is 'dynamic' in that it actively monitors and revises student messages to encourage full task engagement. Specifically, it modifies learner messages by redacting (replacing with blanks) content words and phrases re-used from target sections of text while also rewarding the use of synonyms in order to promote paraphrasing and attention to meaning.

Two versions of the application are compared: a communicative version involving TSCMC between learners and an output-only 'pseudo' communicative version in which the learner must convey messages to their near-future self through taking notes on the target text that they must read and understand later to successfully complete the task. This comparison allows a new perspective on the effect of communicative versus output-only individual study when input, aims, and importantly, the need to engage the target language are controlled for.

Theoretical Accounts of Output and Interaction in SLA

Various theoretical accounts of SLA attempt to discern the roles of input, output, and interaction in language acquisition. One strand of interactionism hypothesizes the importance of negotiation of (or for) meaning (NoM) in the language learning process (Gass & Mackey, 2015; Long, 1996; Loewen & Sato, 2018). NoM begins when the learner indicates a possible lack of understanding with regard to a message (Gass & Mackey, 2015). If the lack of understanding is due to a gap in the learner's interlanguage, then the process by which the learner works with their interlocutor to resolve non-understanding may enable the learner to notice the gap and comprehend the language, thereby facilitating its acquisition (Long, 1996). The grammatical segmentation of messages during NoM as well as direct partner feedback may also facilitate focus on form and consequent adjustments to interlanguage (Long, 1991, 1996). Swain's (1995, 2005) Output Hypothesis is often included in discussions of the Interaction Hypothesis and posits that output improves fluency, assists recognition of gaps in language ability, allows hypothesis testing about linguistic forms, and develops metalinguistic knowledge.

The Involvement Load Hypothesis (ILH) was developed by Laufer and Hulstijn (2001) and builds on Craik and Lockhart's (1972) assertion that the retention of information is determined by the 'depth of processing' it undergoes. The ILH posits that the depth of processing or 'involvement load' induced for a linguistic item during a task depends on the 1) *need* to understand or use the item, 2) *search* required to find the meaning of that item, and 3) *evaluation* of that item or its comparison with other items and their meanings. A meta-analysis by Yanagisawa and Webb (2021) suggests that adding two further factors improves the predictive power of the ILH: 4) *sentence-level varied use* of the target item and 5) *composition-level varied use* of that item. While the ILH does not specifically pertain to tasks involving interaction, it is included here because well-designed language learning tasks, be they communicative or not, tend to induce involvement load.

Willingness to communicate (WTC) or "the willingness to seek out communication opportunities and the willingness actually to communicate in them" in

the L2 (MacIntyre *et al.*, 1998, p. 547) is an important concept when considering the impact of interaction on language learning. Studies have revealed relationships between WTC and various factors including the ideal L2 self, growth mindset, classroom enjoyment, and grit (Lan *et al.*, 2021; Lee, 2022; Lee & Lee, 2020). WTC has been investigated in various CALL contexts, including extramural digital settings with findings revealing that it may be influenced by interpersonal, social, and affective factors as well as educational practices (Soyoof, 2022). We may also posit a relationship between WTC and the factor of need to engage the target language in the ILH. Specifically, the perceived need to engage the language during language learning tasks that require communication may be moderated by learners' WTC, whereas non-communicative language learning tasks should be less affected in comparison.

Research within the framework of socio-cultural theory or social constructivism (Lantolf & Thorne, 2006) is largely founded upon the work of Vygotsky (1978) and views SLA as the process by which language is gradually internalized by the learner through mediation by other people as well as other elements of the social environment including technology. As such, social constructivist approaches have been usefully applied to CALL and its various subfields including mobile assisted language learning (MALL) (Hellermann & Thorne, 2022; Lei *et al.*, 2022), informal digital learning of English (IDLE) (Soyoof, Reynolds, Vazquez-Calvo, & McLay, 2021; Zhang & Liu, 2022), and digital game-based language learning (DGBLL) (Reinhardt & Thorne, 2020; Soyoof, Reynolds, Shadiev, & Vazquez-Calvo, 2021) with results indicating that both communicative and non-communicative technology-mediated learning can promote language acquisition. Of particular note, given the focus of the present study, are findings suggesting that technology-mediated learning may enhance the development of productive vocabulary knowledge compared to conventional approaches (Lei *et al.*, 2022; Soyoof, Reynolds, Shadiev, & Vazquez-Calvo, 2021).

The Effect of Interaction on SLA

After Ellis *et al.* (1994) and Mackey (1999) demonstrated that allowing negotiation of input increased language acquisition, researchers turned their attention to examining the effect of output during interaction.

Ellis and He (1999) conducted a study involving 50 participants that compared the effect on comprehension and vocabulary acquisition of pre-modified input, interactionally modified input with a native speaker (NS), and negotiated output in learner-learner pairs during an object placement information gap task. Results showed that the negotiated output group significantly outperformed the other groups both in comprehension and vocabulary recognition. Ellis and He observed that interaction with the NS in the interactionally modified input group was rather stilted, and the teacher employed low frequency words to explain the target items, whereas student pairs in the negotiated output group communicated with each other more actively employing high frequency vocabulary and NoM.

De la Fuente (2002) conducted a similar study with 32 participants that involved more control for time on task. Negotiated output led to significantly higher

comprehension and productive acquisition of vocabulary. De la Fuente speculates that negotiated output focuses attention and promotes noticing of new vocabulary items as gaps in interlanguage, thereby increasing acquisition.

A number of more recent studies have investigated vocabulary learning during collaborative versus individually completed tasks, revealing positive results in favor of interaction (Kim, 2008a; Niu & Helms-Park, 2014; Rabie-Ahmed & Mohamed, 2022). Rabie-Ahmed and Mohamed (2022) involved 52 learners of Arabic at an American university who performed one of two different vocabulary-oriented tasks created from the same short story—a fill-in-the blank task or a multiple-choice comprehension check task—either in groups of three or individually. Results showed a significant difference in pre- to delayed post-test gains in favor of the groups that interacted during the tasks. The researchers hypothesize that negotiation occurring during the interactive versions of the tasks together with the production of target vocabulary in output contributed to the advantage in language gains.

With the exception of Mackey (1999), all of the studies discussed so far have investigated lexical acquisition, but research has also revealed significant effects of interaction on grammar development (Iwashita, 2003; Leeman, 2003; Mackey & Philip, 1998; Takashima & Ellis, 1999). These studies all involved form-focused interventions on the part of the teacher or researcher, including recasts, negative feedback, and requests for clarification, leaving open the question of whether learner-learner interaction can promote grammar development. Thus, Sato and Lyster (2012) conducted a 10-week quasi-experimental study with 167 learners of English at a university in Japan. Four intact classes were assigned to different conditions: peer-interaction (PI) with training on how to provide CF to peers in the form of recasts, PI with training in giving CF as prompts, PI only, and a control condition. Pre-post measures of grammatical accuracy showed significant differences in favor of the PI-recast and PI-prompt conditions over the PI-only and control conditions, while there was no significant difference between the PI-only and control conditions. The results suggest that learner-learner interaction can indeed promote grammar acquisition but not without training learners on how to provide each other CF.

The positive findings described in this section have motivated CALL research, with particular focus on TSCMC, into the possible benefits of interaction that is computer mediated as discussed below.

TSCMC

One of the most popular research areas in CALL currently is SCMC, which ranked fourth out of 15 topics in a recent meta-analysis of trends by Chen *et al.* (2021). Within this domain lies TSCMC or real-time chat via text messages, an ideal modality for comparing communicative to non-communicative tasks as text is a familiar mode of expression for both communicative purposes (e.g., instant messaging) and non-communicative purposes (e.g., note-taking).

TSCMC has emerged as effective for language learning and research with the meta-analysis by Lin *et al.* (2013) showing a significant effect on language acquisition for TSCMC compared to other modalities and a narrative review by Hughes (2022)

providing further support. The textual and permanent nature of TSCMC may enhance noticing (Smith, 2012; Yuksel & Inan, 2014) and increase the frequency of language related episodes (Zeng, 2017) as well as the quality of attention to form (Hsu, 2022). Also, compared to face-to-face communication (FTFC), it may provide a sense of anonymity that mitigates learner anxiety and shyness (Van der Zwaard & Bannink, 2014). Further, given the finding by Lan *et al.* (2021) that shyness negatively moderates WTC, it is unsurprising that Lee *et al.* (2022) found learners experienced higher WTC in digital settings involving TSCMC (e.g., chat via Facebook) than during in-class FTFC.

A number of studies have investigated the effect of output on SLA during interaction through TSCMC. De la Fuente (2003) compared a vocabulary information gap task completed through FTFC versus TSCMC. Whereas, FTFC resulted in significant differences in immediate post-test scores over the TSCMC group, particularly for oral productive knowledge, delayed post-test scores 3 weeks later showed no significant differences between groups. The researcher concludes that TSCMC may equal FTFC in effectiveness, except regarding oral vocabulary acquisition.

Smith (2012) investigated the effect of negotiation during a jigsaw and decision-making task involving unfamiliar target vocabulary carried out through TSCMC. The researcher found that negotiation was relatively frequent accounting for 34% of the turns taken and that immediate and delayed vocabulary acquisition was significantly higher for items that were negotiated versus those that were explained preemptively by an interlocutor and those that were not engaged, indicating support for the Interaction Hypothesis in online contexts.

Tare *et al.* (2014) investigated the effect of interaction versus output alone on SLA, comparing communicative tasks completed through TSCMC to individual writing assignments involving the same input and similar aims. For example, during one of the interactive tasks, learners received information on various apartments and had to agree on which apartment to rent based on the information. Meanwhile learners in the individual study group received the same information and were asked to choose an apartment and write their reasons for the choice. The interaction group achieved significantly greater gains in target vocabulary knowledge as well as oral production, while no significant difference in gains emerged for writing complexity or accuracy. The researchers considered these findings as supportive of the Interaction Hypothesis, and in a follow-up study analyzing the chat logs (Golonka *et al.*, 2017), identified instances of NoM and CF which lend additional support to the efficacy of interaction.

More recent studies have been conducted investigating the effect of TSCMC versus FTFC (e.g., Kessler *et al.*, 2020; Kim *et al.*, 2020; Kourtali, 2022) and other SCMC modalities (e.g., Dao *et al.*, 2021; Torres & Yanguas, 2021). However, no other studies before or since Tare *et al.* (2014) have attempted to compare the effect of TSCMC tasks on language development to non-communicative tasks that require a similar amount of output production and involve the same input.

Dynamically computer mediated communication and 'pseudo' communication

The author has developed an online application called Text Detective (TD) which can generate and mediate a communicative information gap task from virtually any text of sufficient length. TD (https://su-apps.org/td) is designed to facilitate the learning of low frequency vocabulary in a text and comprehension of the contexts in which they appear. During a communicative task in TD, one learner plays the 'detective' and sees four highlighted low-frequency 'keywords' within the text. Their partner plays the role of the 'witness' and sees one section of the text highlighted. This 'key section' contains one of the keywords seen by the detective. The partners must exchange information to help the detective discover which of their four highlighted keywords is contained within the witness's key section, after which they switch roles and go through the same process with a new keyword and key section. TD ensures that learners engage the meaning of the keywords and key sections by dynamically mediating exchanges. Specifically, keywords, content words, and wordings that are overly similar to segments of key sections are automatically redacted from learners' messages and replaced with blanks ("____"), thereby necessitating the paraphrasing of target content rather than simple reproduction. For example, given the following target section:

The employee proposed several approaches to resolving this issue.

If the learner attempted to paraphrase the section in the following way:

The employee proposed some ways to resolve the issue.

Their message would appear in the chat window as:

The ____ ____ ____ some ways to ____ the ____.

with all of the re-used content words and words with the same roots as content words redacted. To express the meaning of the target section clearly, the learner must paraphrase more of those content words, for example as follows:

The worker proposed some ways to solve the problem.

which would appear in the chat window as:

The worker ____ some ways to solve the problem.

with only the repeated content word "proposed" redacted.

To further encourage paraphrasing, TD rewards learners for using synonyms of keywords and other content words. Thus, TD encourages engagement with the meanings of keywords and contexts from virtually any English text, making it useful both for teaching and research.

After developing the communicative version of TD described above (*TD Chat*), a second version was created called *TD Note*. In TD Note, the learner works individually and is shown three of the four key sections of the text (each containing one of the four keywords), one at a time and asked to type a note on each. Notes typed by learners are dynamically mediated in the same way as messages in TD Chat (i.e., content words and verbatim wordings from key sections are redacted from the note and replaced with blanks while the use of synonyms in the note is rewarded). Each note is saved and when notes have been taken for three of the key sections, two questions appear, one at a time, for the learner to answer: TD displays one of the three notes that the learner has taken and the four keywords and asks the learner which one of the four keywords was in the section of text for which the learner had taken that note. To successfully answer the question, the learner must be able to understand the note well enough to recall which key section the note refers to and thereby choose the correct keyword (the keyword featured in that key section). Since TD prevents the re-use of keywords and other content words from key sections in notes, the learner cannot identify the correct key section by simply finding the part of the text that has the same words as their note. Instead, they must engage the meaning of their note with reference to the meanings of the key sections. The process of typing notes that paraphrase key sections and then later on having to read and reconnect those notes with those sections can be seen as a sort of communication through time between the learner when they were taking the notes to themselves during the question-answering stage. However, since the learner is essentially sending messages (in the form of notes) to their near-future self, rather than another person, the task cannot be called 'communicative,' and is instead referred to hereafter as 'pseudo' communicative.

This pseudo communication arguably provides a sense of communicative accountability and immediacy lacked by conventional non-communicative tasks, such as those used by Tare *et al.* (2014), and increases the need to engage the target language. It is hypothesized that this increased need will raise the task-induced involvement posited by the ILH to lead to language acquisition. Unlike actual communication though, pseudo communication does not provide opportunities for NoM, CF, and other interaction moves that are hypothesized to promote language development.

The Present Study

TD Chat and TD Note offer an opportunity for a highly controlled comparison of the effect of interaction versus output-only individual study in that the input is held constant and the questions engaged are as similar as possible, given the fundamental differences between the two approaches. Most importantly, the need to engage the target language is more comparable between TD Chat and TD Note than for conventional output-only tasks, due to TD Note's encouraging higher communicative accountability.

Therefore, while there still may be an advantage for TD Chat which offers opportunities for interaction moves such as NoM and CF, that advantage should be somewhat diminished due to TD Note's more comparable level of need to engage the target language relative to conventional output-only tasks. In this context, the present research investigates the following questions:

1. Do communicative and pseudo communicative tasks lead to significant language development?
2. Does communicative learning lead to significantly more language development than pseudo communicative learning?
3. Is there a significant relationship between the amount of communication or pseudo communication during the tasks and the amount of language development?

This study also provides an opportunity to investigate the effectiveness of dynamic computer mediation defined as the direct mediation of learner messages or notes based on the target content to which those messages/notes refer. While there have been studies on similar sounding computer-mediated learning and teaching approaches such as dynamic glosses (e.g., Rassaei, 2020), auto-regulated learning (e.g., Wang *et al.*, 2009) and tasks involving augmented reality (e.g., Sydorenko *et al.*, 2021), none of these involved the direct and automatic manipulation of learners' messages by the computer based on target content. To the author's knowledge, therefore, the present study is the first to investigate dynamic computer mediation as defined above.

Furthermore, results showing an equivalence of effectiveness between the dynamically computer mediated communicative and non-communicative (pseudo-communicative) tasks would indicate the possibility that learners might be able to reap some of the benefits associated with communication during individual study via computer, opening up possibilities for more effective learning when an interlocutor is not immediately available. This would carry important implications for second and foreign language teaching stakeholders by, for example, expanding effective options for out-of-class study when synchronous communication between learners may be difficult to arrange.

Finally, although this study does not directly investigate the role of WTC in language learning, results that favor the communicative task over the pseudo-communicative one, might suggest the added impact of WTC on perceived need to engage the language (although this could also indicate the effect of other features unique to interaction such as NoM). The effect of differing levels of WTC might also reveal itself in a wider distribution of output wordcounts for participants who complete the communicative task, as higher or lower levels of individual WTC should be expected to stimulate or impede output production, whereas WTC should have little or no effect on output production during the pseudo-communicative task as it does not involve true communication.

Method

Participants

This study involved 47 intermediate and upper-intermediate learners of English aged 18 to 22 ($M = 19.30$, $SD = 0.76$) at a public university in Japan from a first-year ($n = 26$) and second-year general English course ($n = 22$). Learners were randomly assigned to use TD Note (the *note group*) or TD Chat (the *chat group*). Table 1 shows the group assignment numbers per class. Note that the original number of students assigned to each group within a class was equal but changed due to absences on the day of the experiment.

Table 1
Group Assignment per Class

	Note	Chat
First-year class	12	14
Second-year class	11	10
Total	23	24

Materials

This study employed TD which facilitates the study of virtually any target text. For this experiment, a text from the previous week's homework for each class was chosen for participants from the first-year and second-year courses of 372 and 733 words in length respectively. The experiment employed two different versions of TD: *Chat* and *Note.*

TD Chat

TD Chat is the communicative version of TD used by the chat group which provided a two-way information gap task to facilitate a deeper understanding of the keywords and key sections of the target text. Each text featured four key sections. Each key section was composed of 1-3 sentences and was approximately 30 to 40 words in length. Each key section also featured one 'keyword' upon which the task focused. There was thus a total of four keywords. The task began with partners assigned to either the role of *detective* or *witness*. As shown in Figure 1, both partners were able to see the target text above the chat window, but different parts of the text were highlighted depending on their role: The detective saw the four keywords highlighted in the text and also as choices below the text window, whereas the witness saw a highlighted key section.

Figure 1

TD Chat Task: The Detective (Left) Chats With the Witness (Right) to Determine Which of Their Highlighted Keywords is in the Witness's Highlighted Key Section
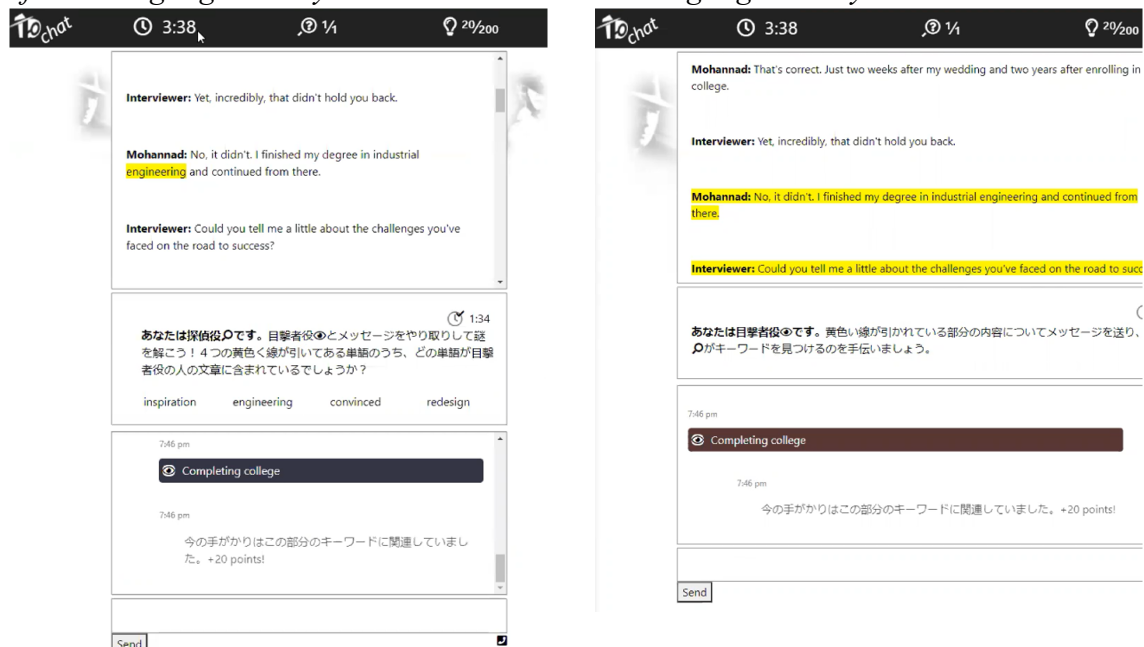
Note. The text displayed in TD in this and all other images is the transcript from a listening activity in the textbook used by the second-year course participants in this study (Scanlon & Vargo, 2015, p. 183) and was the actual text used with those learners in the experiment.

The detective's challenge was to determine which keyword was in the witness's key section through TSCMC in the chat window.

To familiarize learners with the keywords and section, TD's built-in on-click dictionary function (Figure 2) displayed the translation and an image for any word clicked.

Figure 2

TD's Onclick English-Japanese Dictionary Function

However, learners were not allowed to use the keywords and other content words in the sections nor sentences that were overly similar in wording to the key sections, as that would nullify the communicative value of the task. Instead, learners had to indicate which word or section they were referring to by paraphrasing. Attempts to use a keyword or overly similar phrase would trigger an error message explaining why those words were not allowed. The learner's message would then appear in the chat window with those words redacted and replaced with blanks (Figure 3).

Figure 3
TD's Error System: TD Redacts Keywords or Copied Segments of Key Sections From Learners' Communications, Providing an Error Message (Left) and Displaying the Redacted Version of the Message (Right) in the Chat Window

In the same way, TD also redacted non-English words to prevent communication in the L1 as well as redacting numbers and other vocabulary that would allow learners to easily refer to sections and words by location. To discourage guessing, TD also required at least one chat message from each partner before the detective could make their choice. Finally, TD helped maintain the anonymity of partners by redacting names, emails, and other identifying information from messages.

TD also positively reinforced paraphrasing by awarding bonus points for the use of synonyms for keywords and other content words from key sections (Figure 4), which were given both to the learner who typed the synonym and to their partner.

Figure 4
TD's Synonym Bonus System
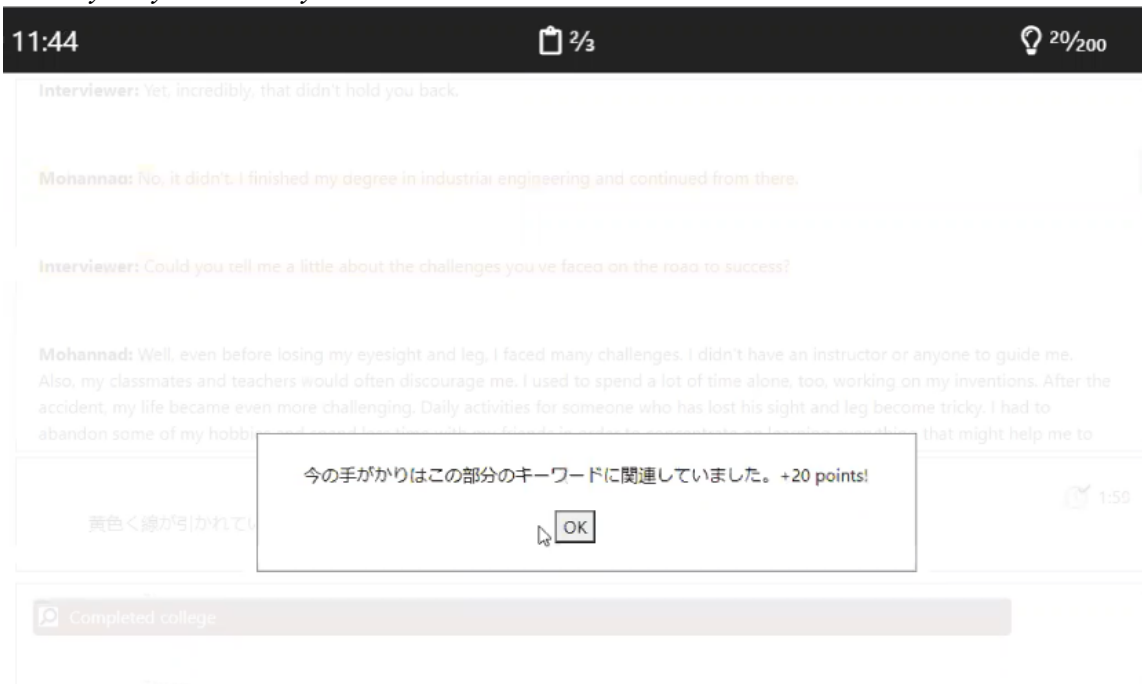
The detective had two chances to choose the correct keyword. Choosing correctly on the first try earned both partners 100 points, while a correct response on the second try earned them 50 (Figure 5).

Figure 5
TD Reward System for Correct Answers
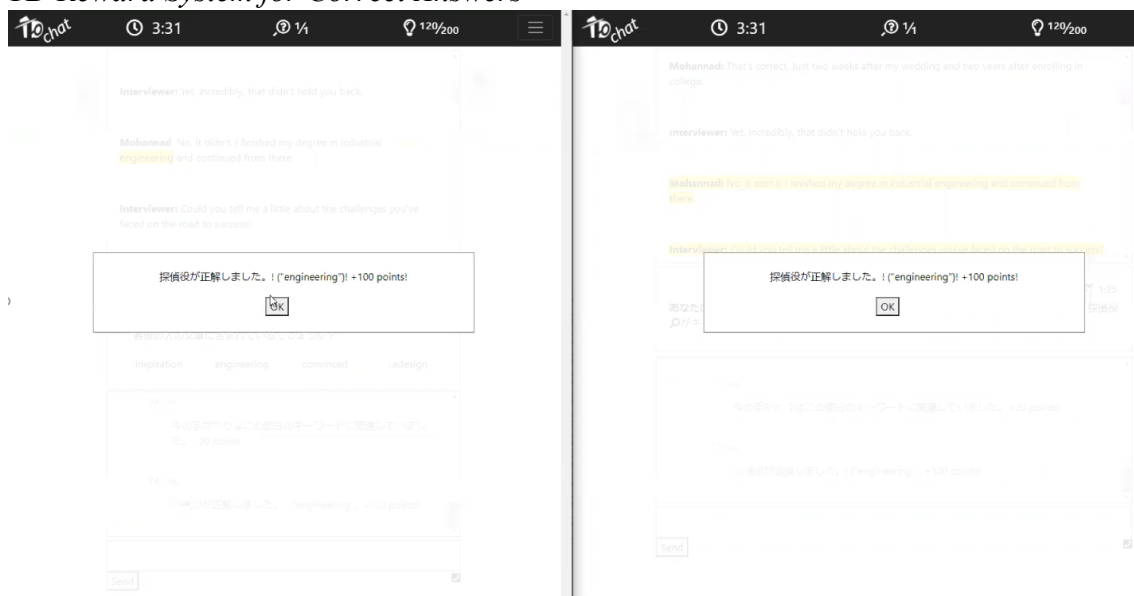
Points had no bearing on participants' grades in the course.

After completing one question, TD switched partners' roles and highlighted a new key section for the new witness and the four keywords for the new detective. Learners had 6 minutes to complete each question, resulting in a total maximum time for the task of 12 minutes.

TD Note

TD Note provided a task that was similar to TD Chat's but done individually. TD first showed learners three key sections of the text, one at a time, and asked them to take notes on each (Figure 6), saving the note in the chat window where it could be reviewed at any time throughout the task.

Figure 6
TD Note Task

TD Note also included the same on-click dictionary function and error and bonus system used by TD Chat to encourage paraphrasing.

Upon completing notes for three of the four key sections in the target text, TD presented the learner with a question which showed a blurred out key section along with the note the learner had taken for that section and, below the note, four keyword choices (Figure 7), one from each of the four key sections.

Figure 7

TD Note Question Featuring the Learner's Note for the Blurred Out Section ("completed college") and Four Keyword Choices Below It
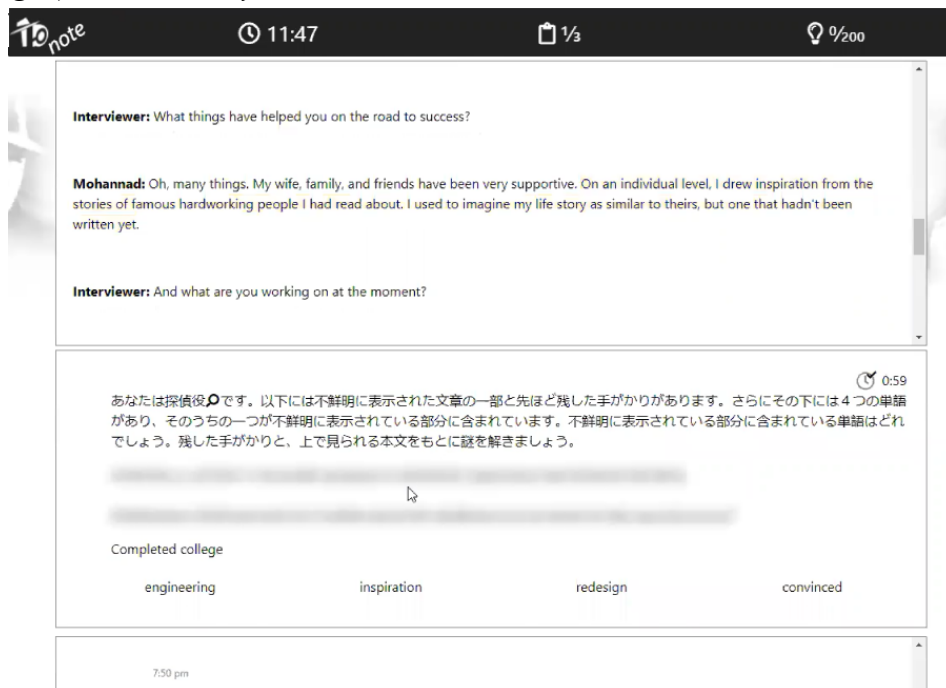
The learner then had to choose the keyword that was in the blurred-out section on which they had taken the note. As with TD Chat, the target text was still visible during the questions, so it was possible for learners to scroll to where they thought the key section was in the text and identify the keyword within it. However, participants had to provide enough information in their note about that section to enable its location when the note appeared in the question. After answering the question, learners were presented with one more in the same format about another section they had taken notes on, making the total number of questions they answered 2—the same as the total number responded to in TD Chat. TD Note was thus designed to provide a task that was as similar as possible to the task in TD Chat but without true communication.

Measures

As the main aim of the TD task was to familiarize learners with the meanings of keywords and the key sections of the target text in which they appeared, an instrument that measured linguistic knowledge of those words and sections was required. In order to sample linguistic knowledge of target forms and meanings and any corresponding syntax or morphology as widely and as sensitively as possible within the shortest amount of time, a modified C-test (Chapelle, 1994) format was developed for this experiment (Figure 8).

Figure 8

The Modified C-test for Measuring Pre and Post Knowledge of Form and Meaning

C-tests have been found to be reliable and time-saving general measures of language proficiency (Chapelle, 1994; Babaii & Ansary, 2001; McKay, 2019; Wolter, 2002) and the cloze test, a predecessor of the C-test, has been employed in the study of productive vocabulary acquisition (Zaki & Ellis, 1999). Modified C-tests for this experiment were created from the target texts used in the tasks and administered through TD. Texts were displayed with the last half of five words in each key section (including the keyword) blanked out requiring students to complete the word, resulting in five completion items per section. These items occurred at one-word or two-word intervals within the key sections. To save time, the latter halves of other words in the texts that would normally have also been blanks were simply blurred out. Since there are multiple levels to 'knowing' a word (Nation, 2001) and learners were expected to have at least some partial knowledge of the target vocabulary, an additional item type was added to the test. Following the keyword completion input in each key section was an additional input for the Japanese translation of that word. Note that entering translations for incomplete items was acceptable, as this indicated that the meaning of the word was understood despite the inability to produce its form. Also, because the test was not meant to measure spelling ability, misspelled words were colored red (though corrections were not suggested).

There were 20 completion and four translation items (24 items in total). Completion items were scored by TD automatically from 0 to 1 point each, and partial scores were possible: .5 for either a minor spelling mistake or inflection error, and .25 for both errors. Translation items were scored from 0 (incorrect) to 2 (correct), with 1 indicating partial correctness (total points possible: 8) by a native Japanese speaker who was highly proficient at English. Though C-tests measure a range of linguistic knowledge

(Babaii & Ansary, 2001), the completion and translation items were posited to primarily measure productive knowledge of linguistic form and meaning respectively.

Cronbach's alpha for pre-test data for the first- and second-year tests was found acceptable at .71 and .73 respectively. Pre- and post-tests were identical, with linguistic gains determined by subtracting pre-test scores from post-test scores. The post-test was administered immediately after the task and once again 1 week later.

Normally, the use of the target text for pre- and post C-tests might be ill-advised, as learners might simply practice typing out the tested content verbatim during the task, thereby achieving gains due to rote memorization rather than actual acquisition. However, TD discouraged this approach by automatically redacting content words and verbatim segments from messages and notes. The 4-minute time limit in which to complete the entire 24 items also discouraged relying on rote recall to respond correctly.

The amount of communication and pseudo communication was measured as the number of words typed during the task, including those redacted by TD. For the chat group, three different measurements were made: words typed by the learner (output), words typed by the interlocutor (input), and the combined total words typed (interaction).

Procedure

This experiment was reviewed and approved by the ethics committee of the university at which it was conducted and informed consent obtained from the participants to use the data collected while preserving anonymity. Two months before the experiment, participants watched a video orientation on TD in class and read a written explanation in Japanese of the research to be conducted. For the next two months, learners completed practice activities in class using the different versions of TD with texts from their textbook and other course materials. All participants gained experience using both experimental versions of TD during this time. They also completed practice tests on the texts in the same format as the tests used in the experiment. This reduced the possibility of inexperience with the application or test negatively affecting performance or a novelty effect positively inflating results. Neither experimental version of TD was purported to be more effective than the other, and while work in TD was graded as part of class participation, all learners received full credit for simply completing the tasks and tests regardless of their performance.

The experiment was conducted during class. The target texts for each class had been part of the previous week's homework assignment in which learners had read and/or listened to the texts and answered comprehension questions on them. The homework had emphasized general comprehension and had not encouraged detailed study of the target content.

Participants first took the 4-minute pre-test. Afterward, both groups completed a 4-minute pre-task warm-up to study the target texts (See Appendix). Next, they began the main task, further studying the target texts through either TD Chat or TD Note. Both groups had 12 minutes to complete the task, after which they took the immediate post-test (identical to the pre-test). Participants were not told that they would do the test again later, nor were they directed to study the target texts or vocabulary further in the meantime.

One week later, the post-test was administered once again thereby obtaining the delayed learning scores.

Results

Shapiro-Wilk tests ($p < .05$) demonstrated the non-normality of test score and gain distributions necessitating non-parametric statistical analyses. For difference tests, effect sizes were considered small when r was near .25, medium when close to .40 and large when approaching or beyond .60 (Plonsky & Oswald, 2014). To answer the three research questions, test scores were first compared within-group and then between groups. Finally, the relationship between communication or pseudo communication and language gains was examined.

1. Do communicative and pseudo communicative tasks lead to significant language development?

Wilcoxon signed rank tests were conducted to investigate within-group for significant ($\alpha < .05$) immediate and delayed pre-to-post total gains as well as gains in knowledge of form and meaning (Tables 2 and 3).

Table 2

Pre- and Immediate Post-Test Meaning, Form, and Total Scores with Comparisons for All Participants and Each Group

	n	Pre-test median (IQR)	Immediate post-test median (IQR)	Median difference	Z^*	p^*	r^*
All participants	47						
Form		13.50 (3.50)	17.50 (2.50)	4.00	-5.85	< .001	.85
Meaning		3.00 (2.00)	6.00 (3.00)	3.00	-5.25	< .001	.77
Total		16.50 (5.88)	22.50 (3.75)	6.00	-5.91	< .001	.86
Note group	23						
Form		13.50 (3.00)	16.50 (2.50)	3.00	-4.11	< .001	.86
Meaning		3.00 (2.00)	6.00 (2.00)	3.00	-3.58	< .001	.75
Total		16.00 (4.25)	22.00 (2.50)	6.00	-4.11	< .001	.86
Chat group	24						
Form		13.50 (4.75)	17.50 (2.50)	4.00	-3.74	< .001	.76
Meaning		3.00 (2.25)	6.00 (3.25)	3.00	-3.88	< .001	.79
Total		16.50 (5.81)	23.25 (4.63)	6.75	-4.29	< .001	.88

*Statistic obtained from a Wilcoxon signed-rank test applied to pre- and post-test comparison

Table 3

Pre- and Delayed Post-Test Meaning, Form, and Total Scores with Comparisons for All Participants and Each Group

	n	Pre-test median (IQR)	Delayed post-test median (IQR)	Median difference	Z*	p*	r*
All participants	44						
Form		13.50 (3.50)	16.75 (3.50)	3.25	-4.97	< .001	.75
Meaning		3.00 (2.00)	5.00 (3.00)	2.00	-3.60	< .001	.54
Total		16.25 (5.81)	21.00 (5.63)	4.75	-4.75	< .001	.72
Note group	21						
Form		13.50 (3.00)	16.50 (3.00)	3.00	-3.33	.001	.73
Meaning		3.00 (2.00)	5.00 (2.00)	2.00	-2.61	.009	.57
Total		16.00 (4.00)	21.00 (3.50)	5.00	-3.06	.002	.67
Chat group	23						
Form		13.50 (4.50)	17.00 (3.25)	3.50	-3.74	< .001	.78
Meaning		3.00 (2.50)	5.00 (3.50)	2.00	-2.49	.013	.52
Total		16.50 (5.88)	20.00 (7.00)	3.50	-3.61	< .001	.75

*Statistic obtained from a Wilcoxon signed-rank test applied to pre- and post-test comparison

As shown, all tests revealed significant pre-post language gains both in total and for knowledge of form and meaning respectively with moderate to large effect sizes.

2. Does communicative learning lead to significantly more language development than pseudo communicative learning?

To investigate for a difference in gains between the two groups, Mann-Whitney U tests were employed. First, pre-test scores were compared to determine the base equivalency of the groups, revealing no significant difference in pre-test total or sub-scores (Table 4).

Table 4

Comparison of Pre-Test Scores Between Groups

	Chat (*n* = 24) median (IQR)	Note (*n* = 23) median (IQR)	Median difference	U^*	p^*	r^*
Form	13.50 (4.75)	13.50 (3.00)	0.00	253.00	.624	.07
Meaning	3.00 (2.25)	3.00 (2.00)	0.00	230.00	.325	.14
Total	16.50 (5.81)	16.00 (4.25)	-0.50	258.00	.701	.08

*Statistic obtained from a Mann-Whitney U test applied to a comparison between groups

Next immediate and delayed language gains were compared between groups, revealing no significant differences (Tables 5 and 6).

Table 5

Comparison of Immediate Score Gains Between Groups

	Chat (*n* = 24) median (IQR)	Note (*n* = 23) median (IQR)	Median difference	U^*	p^*	r^*
Form	3.75 (3.75)	4.00 (4.00)	-0.25	264.50	.806	.04
Meaning	3.00 (1.25)	2.00 (3.50)	1.00	210.50	.153	.21
Total	6.00 (4.13)	5.50 (4.50)	0.50	244.50	.501	.14

*Statistic obtained from a Mann-Whitney U test applied to a comparison between groups

Table 6

Comparison of Delayed Score Gains Between Groups

	Chat (*n* = 23) median (IQR)	Note (*n* = 21) median (IQR)	Median difference	U^*	p^*	r^*
Form	2.50 (2.00)	3.00 (3.50)	-0.50	217.50	.570	.08
Meaning	2.00 (3.00)	2.00 (2.00)	0.00	237.00	.914	.02
Total	5.00 (3.75)	5.00 (4.50)	0.00	218.50	.588	.11

*Statistic obtained from a Mann-Whitney U test applied to a comparison between groups

Finally, groups were compared on other factors related to their task performance including output, time spent on the task, scrolling instances, and words looked up via clicking. Also compared were errors triggered (due, for example, to using a target word in a message), synonyms used for keywords and content words, and percentage of TD questions answered correctly. Table 7 displays each comparison.

Table 7

Comparison of Output, Time on Task, Scrolling, Words Looked Up, Percent Correct on Task Questions, Synonyms Used, and Errors Triggered per Group

	Chat (*n* = 24) median (IQR)	Note (*n* = 23) median (IQR)	Median difference	*U**	*p**	*r**
Output (words typed)	17.00 (14.50)	53.00 (27.50)	-36.00	61.00	< .001	.67
Time on task (minutes)**	5:00 (3:15)	6:00 (5:30)	-1:00	169.00	.021	.34
Scrolling	11.50 (8.25)	4.00 (3.50)	7.50	108.00	< .001	.52
Words looked up before task***	1.50 (4.25)	0.00 (1.00)	1.50	176.50	.019	.34
Words looked up during task	0.00 (0.25)	0.00 (0.00)	0.00	207.00	.011	.37
Errors triggered	2.00 (2.25)	2.00 (3.00)	0.00	178.00	.033	.31
Synonyms used	0.00 (1.25)	0.00 (1.00)	0.00	234.00	.297	.15
Percent correct on task questions	100.00 (6.25)	100.00 (0.00)	0.00	0.00	.654	.07

*Statistic obtained from a Mann-Whitney U test applied to a comparison between groups
**Due to a programming error, time on task could only be measured rounded to the nearest minute
***Words looked up during the warm-up task described in the Appendix

As shown, there were significant differences between groups on every factor except for synonyms used and percent correct on task questions. Of particular interest is the note group's significantly greater amount of output (*r* = .67) and time spent on the task *(r* = .34) compared to the chat group.

3. Is there a significant relationship between the amount of communication or pseudo communication during the tasks and the amount of language development?

No significant relationship was found between output and immediate gains, but this may have been due to a strong ceiling effect evident in the negatively skewed score distribution for the immediate post-test. With regard to delayed gains, while no significant relationships with total gains or gains in knowledge of form emerged per group, one-tailed rank correlation analyses revealed a significant relationship between output and delayed gains in knowledge of meaning for both the note group, $r(19) = .43$, $p = .027$, and the chat group, $r(21) = .39$, $p = .034$. As a moderate ceiling effect likely imposed an artificial limit on gains for participants who scored highly on pre-test meaning items, the data for those scoring higher than the third quartile (4 points) on the meaning section were omitted (5 participants from each group) revealing stronger significant rank correlations for both groups: $r(14) = .53$, $p = .018$ for the note group and $r(16) = .46$, $p = .028$ for the chat group. Since both groups exhibited a similar relationship between output and delayed gains in knowledge of meaning with no significant differences in gains between them, the two groups were combined for further analysis. Because the groups differed significantly

in their output amount, output data was first transformed into within-group percentile ranks to enable comparability across groups. The combined groups exhibited a significant moderate rank correlation between output and delayed gains in knowledge of meaning, $r(32) = .50$, $p = .001$.

Synonyms used and words looked up did not significantly correlate with language gains. Also, for the chat group, there was no significant rank correlation between *partner* output (input) or combined output (interaction) and delayed gains in knowledge of meaning. Finally, rank correlation analyses of pre-test scores (total and partial) and both output production and delayed gains in knowledge of meaning, revealed no significant rank correlation between base proficiency and output or gains.

Discussion

The results show that both the chat and note group achieved significant immediate and delayed gains in knowledge of form and meaning with no significant differences in gains between them, indicating that the tasks were equally effective. However, the chat group was able to achieve the same amount of learning as the note group despite taking significantly less time and producing significantly less output, indicating an advantage for communication in learning efficiency.

One noteworthy finding was that very few words were looked up via TD's dictionary function, with the majority of learners in the note group having not used the function at all. This suggests that learners possessed some partial knowledge of the vocabulary used in key sections including the keywords, but that this knowledge was not sufficient to enable correct responses on some pre-test items within the time limit. To some extent, partial knowledge of the target content was assumed, as the target texts had been assigned as homework the previous week. The gains on the post-tests should therefore be interpreted as primarily gains in partial knowledge of forms, meanings, and associated syntax and morphology, rather than the acquisition of completely new linguistic items.

The effect of communication and pseudo communication on language learning is evidenced by the significant relationship between output and delayed gains in productive knowledge of meaning of the keywords which were the focus of the TD task questions. These results align with previous studies that have found technology-mediated approaches to enhance the development of productive vocabulary knowledge (e.g., Lei *et al*., 2022; Soyoof, Reynolds, Shadiev, & Vazquez-Calvo, 2021). As for overall gains in knowledge of form measured by the C-test completion items, no significant correlation with output emerged suggesting that the gains were not dependent on interaction or note-taking during the task.

The significant correlation between output and delayed gains in knowledge of lexical meaning for the chat group might appear at first glance to support the Interaction Hypothesis. However, the lack of a significant correlation between partners' or total combined output and those gains, suggests that learning occurred primarily through individual production of output rather than overall interaction. This view is further

supported by the similar correlation between output amount and delayed gains in knowledge of meaning for the note group whose gains were no different statistically from those of the chat group despite not having been able to engage in any NoM or other interaction moves hypothesized to promote language development.

Results indicate that output-only individual study can be as effective as communicative study, contrasting with Tare's *et al.* (2014) findings that interaction was superior for developing lexical knowledge. The key difference in this study was that the output-only task was pseudo communicative in that task questions were constructed by TD from the notes that the learner produced such that understanding the notes and how they related to the target content was essential for successful task completion. TD Note thus likely induced a level of need as conceptualized by the ILH which was more comparable to that of communicative tasks than conventional output-only tasks, such as those employed in Tare *et al.* (2014). These findings support the ILH while providing preliminary evidence that dynamically computer-mediated individual tasks can emulate communicative tasks and match them in effectiveness. This has important pedagogical implications for language education stakeholders suggesting learners may experience some of the benefits normally associated with communication through completing pseudo communicative tasks on their own. Since no partner is required, such tasks can be done without having to make prior arrangements with an interlocutor. This makes pseudo communitive tasks ideal for out-of-class learning, particularly informal digital learning situations (Lee, 2017) where it may be difficult to find an interlocutor who is willing and able to communicate with the learner about a particular text they wish to study.

While the results appear to support the ILH in their implying the importance of need to engage the target language, this does not necessarily cast doubt upon the Interaction Hypothesis. Although neither task took very long to complete, the chat group was significantly more efficient in their learning compared to the note group. This significant difference in efficiency in favor of communication implies an advantage for interaction that may stem from features intrinsic to the communicative process such as NoM. However, the lack of a significant difference in absolute gains, does imply the possibility that previous studies comparing interactive to non-interactive tasks (e.g., de la Fuente, 2002; Ellis *et al.*, 1994; Ellis & He, 1999; Mackey, 1999; Tare *et al*, 2014) may have neglected the implications of the ILH to their detriment. Future studies would do well to acknowledge the possibility that the advantages of communication may not exclusively stem from the exchange of information between interlocutors, but also from the need to engage the target language that communication naturally induces.

Although WTC was not the primary focus of this study, it is interesting to observe that the chat group experienced greater variation in the amount of output they produced (as measured by the IQR for output wordcount) if taken as a percentage of the median (85%) than the note group (52%). This may indicate a moderating effect due to individual differences in participants' WTC on the output of the chat group which would not have been present for the note group. Given the significant correlation between output and language gains, this may suggest a moderating role for WTC on learning during communicative tasks, amplifying the effect of the communicative task for those with a high WTC while dampening it for those with a low WTC.

It is important to note that while the significant moderate correlations between output amount and delayed gains in knowledge of lexical meaning for both groups suggest that the production of output led to the gains, there may be other explanations for these correlations. For example, perhaps initial differences in proficiency led to both the differences in output and the differences in gains. This seems unlikely though, given the lack of significant correlations between pre-test total or sub-scores, and output or delayed gains in meaning. Future studies might include an input-only control group to more clearly distinguish the effect of output from other factors.

For now, we may get a sense of the differences and similarities between successful and unsuccessful learners in both groups by comparing excerpts from the chat/note logs of students from each category (Table 8).

The students in the chat condition were playing the role of the witness and had to describe the following key section to their partner, while those in the note condition took notes on the same section:

> *Mohannad: No, it didn't. I finished my degree in industrial engineering and continued from there.*
>
> *Interviewer: Could you tell me a little about the challenges you've faced on the road to success?"* (transcribed from the listening activity in Scanlon & Vargo, 2015, on p. 183)

The keyword choices for the question on this section were: *engineering, redesign, inspiration,* and *convinced* (Correct choice: *engineering*).

Table 8

Comparison of Excerpts From Students (S1 - S4) With Low and High Output/Gains (and Their Partners: P1 - P2)

	Chat	Note
Low output/gain	**S1:** he completed his study and didn't stop *S1 receives a synonym bonus* **P1:** is he talk about his specializing? **S1**: yes!* **S1:** he is talking about it **P1:** ok *P1 chooses the correct keyword*	**S3:** started from speaker's denying, taiking about study. ended in I's asking *Later, on the relevant question…* *S3 chooses the incorrect keyword on the first try, but chooses correctly on the second try*

| High output/gain | *S2 looks up 'continued'*
S2: he still continue after end of <u>degree</u>
S2: <u>interviewer</u> ask the difficult to best <u>carrer</u>
S2: at <u>first</u> he <u>denys</u>
S2 looks up 'degree'
S2 looks up 'industrial'
S2 looks up 'engineering'
S2: he graduate the science type of major
P2: like he continued after the end of his study in school
P2 receives a synonym bonus
S2: <u>yes</u>
P2 chooses the correct keyword | **S4:** that did nothing. I completed <u>the</u> grade of factory and keep doing from here. <u>Can you say to me a few things on the difficulties you had on</u> your way <u>to</u> successful life
S4: teach me difficult things on your way to successful life
Later, on the relevant question…
S4 chooses the correct keyword |

*Underlined words were redacted by TD and visible only as blanks (___) in the chat/note window

The students in the chat condition (S1 and S2) both engaged in the NoM with the confirmation request of P1 ("is he talking about specializing?") and prompt for confirmation of P2 ("like he continued after the end of his study in school"), but this negotiation did not result in substantial elaboration for either group. On the other hand, S2 provides unprompted elaboration on the key section, including an attempt to paraphrase "engineering" as a "science type of major," whereas, S1 provides only minimal description of the key section and does not attempt to paraphrase the keyword. Also, S2 looks up four content words from the key section, including the keyword just before paraphrasing it.

In the note condition, S4 provides more elaboration on the key section, including re-paraphrasing the part of their initial note that was redacted due to its similarity to the original. This elaboration includes the phrase "grade of factory" which may be an attempt to paraphrase "degree in industrial engineering." S3, on the other hand, avoids the need to paraphrase the keyword and surrounding context by simply indicating the topics touched on at the beginning and end of the key section.

From this albeit cursory comparison, we may hypothesize that the relationship between output and delayed gains in meaning may have been due in part to the fact that those who attempted to thoroughly paraphrase key sections tended to process, write, and retain more with regard to the meanings of the key sections and keywords they were writing about. Paraphrasing itself is a form of evaluation, one of the three factors contributing to involvement load in the ILH, and is also considered a kind of 'generative processing,' both of which have been found to promote vocabulary acquisition (Joe, 1998; Keating, 2008; Kim, 2008b; Yanagisawa & Webb, 2021; Zaki & Ellis, 1999).

Conclusion

This study introduced TD, a new kind of online application for text-based language learning through dynamically computer-mediated communication and pseudo

communication. Results showed that both the communicative TD Chat and pseudo communicative TD Note led to significant immediate and delayed gains in knowledge of linguistic form and meaning, with no significant difference in gains between them. Also, the significant relationships between output and delayed gains in knowledge of lexical meaning found for both groups indicate the central role of output production in learning regardless of the TD version used.

The lack of a significant difference in gains between groups, suggests that non-communicative study requiring output production may be as effective as communicative study when the non-communicative task has a comparable level of need to engage the target language. Not only does this finding support the ILH by implying the importance of need to engage the target language, it also implies that previous studies comparing interactive to conventional non-interactive tasks may have overlooked the effect of the differing amount of need associated with each. This difference in need alone is likely to have led to deeper processing of input in comparison with conventional non-interactive tasks. At the same time, it could be argued that true communication still provided an advantage over pseudo communication in that the communicative task required less time and output to achieve the same amount of learning. Still, the results carry preliminary pedagogical implications suggesting that dynamically mediated pseudo communicative tasks might provide learners with a viable alternative to communicative tasks when an interlocutor is not readily available.

In addition to the main findings, this study may have also obtained indirect evidence for a moderating effect of differing individual levels of WTC on output production. This effect may have revealed itself in the greater variation in output wordcounts relative to the median for the chat task which was communicative compared to the note task which was not. Given the significant correlation found between output and gains in knowledge of lexical meaning, this suggests an important moderating role for WTC in language learning through communicative tasks. It may also imply a possible advantage of employing pseudo communicative tasks with learners who have low WTC.

The results of this study suggest a number of directions for future research. One of the most interesting findings of the current study was the advantage in learning efficiency for the communicative task over the pseudo communicative task. While previous research suggests that features of interaction such as NoM and CF are mainly responsible for language development during communication, the preliminary examination of the chat script in the discussion failed to find support for this hypothesis. Future studies comparing communicative and non-communicative computer-mediated tasks requiring output might attempt to investigate for correlations between observed interaction moves and pre-post measures of language development. Further research might also investigate whether pseudo communicative tasks are more effective for language learning than communicative tasks for learners with low WTC as postulated above. Such studies might obtain measurements of learners' WTC prior to the experiment to more directly examine the extent to which it moderates performance on the two types of task.

This study has several limitations, including the lack of an input-only control group as well as the arguably unnatural style of communication mediated by TD in which

certain words were redacted from messages in order to encourage paraphrasing and the engagement of meaning. Also, the sample size was limited and the duration of the experimental tasks was short with no further post-tests administered beyond the 1-week mark. Further studies are therefore necessary before any firm conclusions can be drawn. Still, in addition to trialing a novel approach to text-based language learning, this study provides a first glimpse into the issue of need imbalance between interactive and conventional non-interactive tasks which may explain some of the relative gains attributed to NoM and other surface-level features of interaction in previous studies on the effect of communication.

Acknowledgement

References

Babaii, E., & Ansary, H. (2001). The C-test: a valid operationalization of reduced redundancy principle? *System, 29*(2), 209-219. https://doi.org/10.1016/s0346-251x(01)00012-4

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second language research*, *10*(2), 157-187. https://doi.org/10.1177/026765839401000203

Chen, X., Zou, D., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning & Technology, 25*(3), 151-185. http://hdl.handle.net/10125/73454

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dao, P., Duong, P. T., & Nguyen, M. X. N. C. (2021). Effects of SCMC mode and learner familiarity on peer feedback in L2 interaction. *Computer Assisted Language Learning*, 1-29. https://doi.org/10.1080/09588221.2021.1976212

de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition, 24*(1), 81-112. https://doi.org/10.1017/S0272263102001043

de la Fuente, M. J. (2003). Is SLA interactionist theory relevant to CALL? A study on the effects of computer-mediated interaction in L2 vocabulary acquisition. *Computer Assisted Language Learning, 16*(1), 47-81. https://doi.org/10.1076/ call.16.1.47.15526

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*(2), 285-301. https://doi.org/10.1017/S0272263199002077

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom Interaction, Comprehension, and the Acquisition of L2 Word Meanings. *Language Learning, 44*(3), 449-491. https://doi.org/10.1111/j.1467-1770.1994.tb01114.x

Gass, S. M., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 180–206). Lawrence Erlbaum.

Golonka, E. M., Tare, M., & Bonilla, C. (2017). Peer interaction in text chat: Qualitative analysis of chat transcripts. Language Learning & Technology, 21(2), 157–178. https://dx.doi.org/10125/44616

Hashimoto, K., & Glasgow, G. P. (2019). CLIL for who? Commodification of English-medium courses in Japan's higher education. In I. Liyanage & T. Walker (Eds.), *Multilingual education yearbook 2019* (pp. 103-119). Springer, Cham. https://doi.org/10.1007/978-3-030-14386-2_6

Hellermann, J., & Thorne, S. L. (2022). Collaborative mobilizations of interbodied communication for cooperative action. *The Modern Language Journal, 106*(S1), 89-112. https://doi.org/10.1111/modl.12754

Hughes, L. S. (2022). Text-based SCMC for SLA: A narrative review. *Technology in Language Teaching & Learning, 4*(1), 1-17. https://doi.org/10.29140/tltl.v4n1.712

Hsu, H. C. (2022). Peer interaction and attention to form in web-based synchronous and asynchronous L2 collaborative writing. *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2022.2095405

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in second language acquisition, 25*(1), 1-36. https://doi.org/10.1017/S0272263103000019

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*(3), 357-377. https://doi.org/10.1093/applin/19.3.357

Junn, H. (2021). L2 communicative competence analysis via synchronous computer-mediated communication (SCMC) as an alternative to formal classrooms. *Innovation in Language Learning and Teaching,* 1-17. https://doi.org/10.1080/17501229.2021.1895802

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*(3), 365-386. https://doi.org/10.1177/1362168808089922

Kessler, M., Polio, C., Xu, C., & Hao, X. (2020). The effects of oral discussion and text chat on L2 Chinese writing. *Foreign Language Annals, 53*(4), 666-685. https://doi.org/10.1111/flan.12491

Kim, Y. (2008a). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *The Modern Language Journal, 92*(1), 114-130. https://doi.org/10.1111/j.1540-4781.2008.00690.x

Kim, Y. (2008b). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 285-325. https://doi.org/10.1111/j.1467-9922.2008.00442.x

Kim, Y., Skalicky, S., & Jung, Y. (2020). The role of linguistic alignment on question development in face-to-face and synchronous computer-mediated communication contexts: A conceptual replication study. *Language Learning, 70*(3), 643-684. https://doi.org/10.1111/lang.12393

Kourtali, N. E. (2022). The effects of face-to-face and computer-mediated recasts on L2 development. *Language Learning & Technology, 26*(1), 1–20. https://dx.doi.org/10125/73457

Lan, G., Nikitina, L., & Woo, W. S. (2021). Ideal L2 self and willingness to communicate: A moderated mediation model of shyness and grit. *System, 99*, 102503. https://doi.org/10.1016/j.system.2021.102503

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Lee, J. S. (2017). Informal digital learning of English and second language vocabulary outcomes: Can quantity conquer quality? *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12599

Lee, J. S. (2022). The role of grit and classroom enjoyment in EFL learners' willingness to communicate. *Journal of Multilingual and Multicultural Development, 43*(5), 452-468. https://doi.org/10.1080/01434632.2020.1746319

Lee, J. S., & Lee, K. (2020). Role of L2 motivational self system on willingness to communicate of Korean EFL university and secondary students. *Journal of Psycholinguistic Research, 49*(1), 147-161. https://doi.org/10.1007/s10936-019-09675-6

Lee, J. S., Lee, K., & Chen Hsieh, J. (2022). Understanding willingness to communicate in L2 between Korean and Taiwanese students. *Language Teaching Research, 26*(3), 455-476. https://doi.org/10.1177/1362168819890825

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition, 25(*1), 37-63. https://doi.org/10.1017/S0272263103000020

Lei, X., Fathi, J., Noorbakhsh, S., & Rahimi, M. (2022). The impact of mobile-assisted language learning on English as a foreign language learners' vocabulary learning attitudes and self-regulatory capacity. *Frontiers in Psychology*, *13*. https://doi.org/10.3389%2Ffpsyg.2022.872922

Lin, W. C., Huang, H. T., & Liou, H. C. (2013). The effects of text-based SCMC on SLA: A meta analysis. *Language Learning & Technology, 17*(2), 123-142. https://dx.doi.org/10125/44327

Liu, Y., & Hu, G. (2021). Mapping the field of English for specific purposes (1980–2018): A co-citation analysis. *English for Specific Purposes, 61*, 97-116. https://doi.org/10.1016/j.esp.2020.10.003

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching, 51*(3), 285-329. https://doi.org/10.1017/S026144481 8000125

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bathia (Eds.), *Handbook of research on second language* acquisition (pp. 413-468). Academic Press.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition, 16*(3), 303-323. https://doi.org/10.1017/S0272263100013103

MacIntyre, P.D., Clément, R., Dörnyei, Z., & Noels, K.A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 82*, 545–562. https://doi.org/10.1111/ j.1540-4781.1998.tb05543.x

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in second language acquisition, 21*(4), 557-587. https://doi.org/10.1017/S0272263199004027

Mackey, A., & Philip, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82*(3), 338-56. https://doi.org/10.1111/j.1540-4781.1998.tb01211.x

McKay, T. (2019). *More on the Validity and Reliability of C-test Scores: A Meta-Analysis of C-test Studies* [Doctoral dissertation, Georgetown University]. Georgetown University. http://hdl.handle.net/10822/1057315

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.

Niu, R., & Helms-Park, R. (2014). Interaction, modality, and word engagement as factors in lexical learning in a Chinese context. *Language Teaching Research, 18*(3), 345-372. https://doi.org/10.1177/1362168813510383

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. https://doi.org/10.1111/ lang.12079

Rabie-Ahmed, A., & Mohamed, A. (2022). Collaborative and individual vocabulary learning in the Arabic classroom: The role of engagement and task demands. *Foreign Language Annals*, 1-19. https://doi.org/10.1111/flan.12636

Rassaei, E. (2020). Effects of mobile-mediated dynamic and nondynamic glosses on L2 vocabulary learning: A sociocultural perspective. *The Modern Language Journal, 104*(1), 284-303. https://doi.org/10.1111/modl.12629

Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practice, and proceduralization. *Studies in Second Language Acquisition, 34*(4), 591-626. https://doi.org/10.1017/S0272263 112000356

Scanlon, J., & Vargo, M. (2015). *Q: Skills for Success. Listening and Speaking.* Oxford University Press.

Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology, 16*(3), 53-81. https://dx.doi.org/10125/44300

Soyoof, A. (2022). Iranian EFL students' perception of willingness to communicate in an extramural digital context. *Interactive Learning Environments*, 1-18. https://doi.org/10.1080/10494820.2021.2024579

Soyoof, A., Reynolds, B. L., Shadiev, R., & Vazquez-Calvo, B. (2021). A mixed-methods study of the incidental acquisition of foreign language vocabulary and healthcare knowledge through serious game play. *Computer Assisted Language Learning*, 1-34. https://doi.org/10.1080/09588221.2021.2021242

Soyoof, A., Reynolds, B. L., Vazquez-Calvo, B., & McLay, K. (2021). Informal digital learning of English (IDLE): a scoping review of what has been done and a look towards what is to come. *Computer Assisted Language Learning*, 1-27. https://doi.org/10.1080/09588221.2021.1936562

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of William E. Rutherford* (pp. 125-144). Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-483). Lawrence Erlbaum Associates.

Sydorenko, T., Thorne, S. L., Hellermann, J., Sanchez, A., & Howe, V. (2021). Localized globalization: Directives in augmented reality game interaction. *The Modern Language Journal*, 105(3), 720-739. https://doi.org./10.1111/modl.12722

Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In Rod Ellis (Ed.), *Learning a second language through interaction* (pp. 173-188). John Benjamins.

Tare, M., Golonka, E. M., Vatz, K., Bonilla, C. L., Crooks, C. & Strong, R. (2014). Effects of interactive chat versus independent writing on L2 learning. *Language Learning & Technology, 18*(3), 208-227. https://dx.doi.org/10125/44391

Torres, J., & Yanguas, Í. (2021). Levels of engagement in task-based synchronous computer mediated interaction. *Canadian Journal of Applied Linguistics, 24*(2), 234-259. https://doi.org/10.37213/cjal.2021.31319

Van der Zwaard, R., & Bannink, A. (2014). Video call or chat? Negotiation of meaning and issues of face in telecollaboration. *System, 44*, 137-148. https://doi.org/10.1016/j.system.2014.03.007

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

Wang, H., Waple, C. J., & Kawahara, T. (2009). Computer assisted language learning system based on dynamic question generation and error prediction for automatic speech recognition. *Speech Communication*, *51*(10), 995-1005. https://doi.org/10.1016/j.specom.2009.03.006

Wolter, B. (2002). Assessing proficiency through word associations: is there still hope? *System, 30*(3), 315-329. https://doi.org/10.1016/S0346-251X(02)00017-9

Yanagisawa, A., & Webb, S. (2021). Involvement Load Hypothesis Plus: Creating an improved predictive model of incidental vocabulary learning. *Studies in Second Language Acquisition*, 1-30. https://doi.org/10.1017/S0272263121000577

Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL, 26*(3), 333-354. https://doi.org/10.1017/S0958344014000147

Zaki, H., & Ellis, R. (1999). Learning vocabulary through interacting with a written text. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 153-169). John Benjamins.

Zeng, G. (2017). Collaborative dialogue in synchronous computer-mediated communication and face-to-face communication. *ReCALL, 29*(3), 257-275. https://doi.org/10.1017/S0958344017000118

Zhang, Y., & Liu, G. (2022). Revisiting informal digital learning of English (IDLE): a structural equation modeling approach in a university EFL context. *Computer Assisted Language Learning*, 1-33. https://doi.org/10.1080/09588221.2022.2134424

Appendix

TD Warm-Up Task

TD administered an input-only warm-up task to both groups prior to the experiment that featured the same on-click dictionary function implemented in TD Chat and Note. The warm-up also featured the same keywords and key sections as the experimental tasks. It displayed three of the four key sections from the target text, one at a time, and asked learners to study each, clicking 'Continue' when they had finished (Figure A1).

Figure A1
Learners Study Key Sections of the Text Displayed by the Program

Afterword, TD scrolled to a key section in the text, highlighting and blurring out a 6-word segment of that section and asked which of the four keywords was in that blurred out segment. The learner had to recall or infer which keyword was within that blurred out segment and choose it to answer correctly (Figure A2).

Figure A2
TD Warm-Up Question

Learners answered two questions in this format and received a maximum of 4 minutes to complete the task.