

# *Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis*

HUIFEN LIN

*Foreign Languages and Literature Department, National Tsing Hua University, No. 101,  
Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C.  
(Email: huifen@mx.nthu.edu.tw)*

---

## Abstract

The ever growing interest in the development of foreign or second (L2) oral proficiency in a computer-mediated communication (CMC) classroom has resulted in a large body of studies looking at both the direct and indirect effects of CMC interventions on the acquisition of oral competences. The present study employed a quantitative meta-analytic approach to investigate such effects by synthesizing (quasi)experimental studies that provide empirical quantitative data for effect size calculation. A literature search located 25 relevant studies for the final analysis. Each study was independently coded for learner, design and publication characteristics. The averaged effect size was estimated from the included studies. The results of the meta-analysis reveal that communication mediated by computer/technologies produced a moderate positive effect on L2 learners' oral proficiency compared to face-to-face (F2F) communication or no interaction. Furthermore, CMC has roughly similar effect on pronunciation, lexical and syntactic level of oral production; however, it might have a negative impact on fluency and accuracy. This meta-analysis also found that the effect of CMC on oral proficiency depends on several methodological factors such as task type, outcome measurement, treatment length, and assessment task. Major findings of the current meta-analysis include: (1) studies relying on elicited data are superior to those utilizing naturalistic data; (2) reading aloud seems to be the task that could elicit the best oral performance from students; (3) surprisingly, CMC appeared to be harmful for accuracy and fluency; (4) studies that employed decision-making generated the largest effect size, followed by studies that used more than one task type; (5) among the four tasks, jigsaw actually generated a negative effect on oral performance; and (6) as the most popular task employed by primary researchers, opinion-exchange studies produced the smallest effect size. These findings need to be interpreted as exploratory rather than confirmatory since each of them became less trustworthy after taking into consideration numerous other factors such as CMC task and the particular CMC tool used, etc. Future research suggestions are provided and the limitations of this meta-analysis are addressed.

Keywords: computer-mediated communication, meta-analysis, oral proficiency, second language acquisition, research synthesis, effect size

---

## 1 Introduction

CALL researchers have proposed that the affordances and features of CMC technologies provide mediating tools that can create or support a social context in virtual space that may be beneficial for the cognitive development needed for L2 learning (Meskill, 1999;

Peterson, 2009). A number of justifications for CMC to be used for language development in accordance with the social-cultural framework have been made. For example, interpersonal and intrapersonal interactions mediated by CMC tools such as chat rooms, emails or discussion bulletin boards offer opportunities for learners to tackle communication problems collaboratively, which is believed to facilitate L2 learning. Furthermore, some features of CMC tools, such as the modality of information presentation during interaction, might be effective in terms of enhancing depth of processing. For example, some technologies such as email support only written text, while others such as audio/video-conferencing include audio, video and visuals. Dual or multiple coding of the same information through a number of sensory channels is believed to enhance the comprehension and retention of incoming knowledge (Smith, Alvarez-Torres & Zhao, 2003). Another feature of CMC, temporality, which differentiates between communication taking place in real time or involving some delay in message transfer, also significantly impacts the discourse and interaction patterns of learners (Smith *et al.*, 2003). When there is some delay in responding to a message, more time is given for contemplating the content of the message and drafting a response; in contrast, in real-time communication, contemplation of the received message and thoughtful responses are less likely. These features of CMC seem to provide opportunities to create a social interaction context with more flexibility that cannot be afforded in a traditional face-to-face environment. When CMC is found to have a greater effect than face-to-face interactions, factors such as task type, treatment length and the way oral proficiency is measured might also mediate such superiority. Given that numerous empirical studies have accumulated, and since we are not informed by these studies of a definite direction to follow when choosing to use CMC in our L2 classrooms, the time seems ripe for a retrospective review and systematic synthesis of the accumulated primary studies which might assist us in “discovering new developments in the research fields and identifying unsolved problems or needs ... provid[ing] directions for future research” (Dinsmore, 2006: 59–60).

### 1.1 Theoretical framework

Advances in communication technologies have prompted researchers in second language acquisition (SLA) to investigate the potential of using CMC in a language classroom to foster language development. Most CMC research has been conducted through the interactionist theory of language learning (Satar & Özdener, 2008). Interaction, broadly speaking, can refer to both interpersonal and intrapersonal activities that arise when people engage in face-to-face communication. Intrapersonal activity is a required mental processing of language to prepare for interpersonal interaction, which can in fact be triggered by interpersonal interaction (Ellis, 1999). The relationship between interpersonal and intrapersonal relationships can be manifested by examining three theories, the Interaction Hypothesis, Socio-cultural Theory and the Depth of Processing Theory. Originally used to address communication breakdowns experienced by native speakers in first language (L1) settings (Ellis, 1999), the Interaction Hypothesis has also provided insights into examining interaction behaviors and features of second language learning/acquisition. Providing a key insight into the Interaction Hypothesis, Hatch (1978) as cited in Ellis (1999) has argued that what is directly relevant in second language learning is that, rather than simply serving as a product of what learners have already learned, the process of interaction can actually be used as an L2 learning process if the learners engage in “negotiation of meaning”. Language

acquisition can be achieved, or at least facilitated, if L2 learners collaboratively tackle communication problems arising in interaction via negotiating or modifying the input so that it becomes comprehensible (Krashen, 1985). Comprehensible input is a necessary condition for SLA to take place because it activates internal acquisitional mechanisms. If the input is not made comprehensible, not only will the “affective filter” be high, making learners less motivated to learn, but it may also be difficult for their acquisitional mechanism to operate.

While the Interaction Hypothesis stresses the important contribution of the interaction process itself to second language learning, the social context in which such interaction takes place plays an indispensable role. Learning takes place when the learner and other peers engage in collaborated tasks afforded by the social context (Simpson, 2005; Peterson, 2009). The sole provision of social context might not be sufficient, though, for learning to take place. According to Vygotsky’s Sociocultural Theory (1978), appropriate use of tools can facilitate cognitive development. One example of such mediating tools is language. Through the use of language, learners are able to transform “lower mental functions such as memory, conceptual thought and problem solving ... to higher-level functions” (Peterson, 2009: 304). Furthermore, “peer interaction, scaffolding and modeling” are activities that can assist individual learners in advancing their cognitive development with the help of more capable peers according to Vygotsky’s Zone of Proximal Development (ZPD) theory (Lin, 2009: 16). Within this zone, teachers or peers with different levels of expertise and a variety of tools can be utilized in combination to support intentional learning (Lin, 2009).

### ***1.2 CMC and oral proficiency development***

The theoretical discussion of Interactionist, or more generally, SLA theories above seems to endorse the use of CMC in the language classroom given that it creates a social context that is similar to face-to-face in which most features of authentic interaction can be replicated. How might CMC provide an ideal venue, in particular, for oral proficiency development? Payne and Whitney (2002) conducted an early empirical study to test the hypothesis that there is a link between CMC and oral proficiency. They used Levelt’s model of language production with the Working Memory theory to examine whether oral proficiency can be indirectly improved for students exposed to synchronous computer-mediated communication. Levelt’s model was originally developed to explain first language production (Levelt, 1989) but was later widely employed to explain L2 production (Payne & Whitney, 2002). It comprises three stages of language production. In the first stage, communication intentions are perceived and the semantic meaning of the intended messages is generated, e.g. preverbal messages, in the Conceptualizer. In the second stage, the preverbal message enters into the Formulator where grammatical and phonological coding is performed, surface structures are determined and finally an articulatory plan is formed, waiting to be carried out through the Articulator where an utterance is ready to be produced. It has to be noted that before entering the Articulator, speakers still have the chance to monitor their intended utterance with the support of subvocalization. The role of working memory in this model provides justifications for CMC to be an ideal venue for oral proficiency development. According to Payne and Whitney (2002), the semantic content of the intended spoken message has to be maintained in working memory before it is fed into the Formulator. Additionally, an articulatory plan of utterance has to be stored in the Articulatory Buffer

(working memory) before it is carried out by the Articulator to produce an utterance. For spontaneous conversational speech, the demand on working memory for language production is usually high. For language learners who are still in the Interlanguage development stage in which target language hypothesis testing and searching for accurate semantic, lexical and surface structures to match intended messages takes more time than it does in their L1, reducing the burden on the working memory needed for these processes seems to be a priority. Features and affordances of communication technologies have the potential to either reduce the burden on or induce a more effective use of working memory. Payne and Whitney (2002) suggest two features associated with text-based chat room discussions that might reduce the memory load: decreased speed of conversational exchange, and the non-ephemeral nature of the interaction due to the medium. It is proposed that when L2 learners engage in text-based synchronous CMC, they have more time for pre-task planning, which is likely to result in more fluent, complex and accurate output (Ortega, 1999). In addition, this reduced rate of exchange is believed to benefit L2 learners with low phonological working memory capacity. From another working memory perspective, the chat scripts available to learners allow them to re-read and refresh the conversation without having to first store and then retrieve it from their working memory. Working memory can thus be freed up for some other processing of the interaction. In other words, computer-mediated communication, by “developing the same cognitive mechanisms underlying spontaneous conversational speech” (Payne & Whitney, 2002: 7) and reducing working memory load, provides L2 learners with an environment to practice language production at a reduced rate. The relatively reduced rate of exchange and lag-time induced by the text-chat software allows L2 learners “more time to both process incoming messages and produce and monitor their output” (Sauro & Smith, 2010: 557). Additionally, text-based CMC involves written and spoken features, which provides opportunities for noticing the form while engaging in conversations in slow motion (Smith, 2003a; Beauvois, 1998). These features are attractive to both L2 learners and L2 instructors, given that it is not always possible to provide students with plentiful opportunities for speaking due to time constraints, and the high level of anxiety usually felt by students in face-to-face conversation. CMC might be regarded as an ideal alternative to develop speaking ability when face-to-face is not possible or does not work as expected.

## 2 Literature review

The large body of research on the use of CMC in language learning began in the mid-1980s (Cummins, 1986; Wang, 2010). The number of studies on CMC for oral development, in particular, has been increasing since communication technologies have become more advanced and capable of supporting opportunities for social networking and collaboration. This line of research, conducted from different perspectives, has attempted to uncover both the product and nature of communication mediated by computers/technologies in an L2 context. CMC researchers have investigated topics such as:

- a. a comparison of effect on oral development between F2F, synchronous CMC and asynchronous CMC (Beauvois, 1998; Pyun, 2003; Chang, 2007, 2008; Xiao, 2007; Satar & Özdener, 2008; Sequeira, 2009; Volle, 2005);
- b. the quantity and quality of students’ language output via CMC and via F2F (Chang, 2007, 2008);

- c. the effects of different kinds of CMC dyads on oral development (Alastuey, 2010 );
- d. the effects of different CMC tools on overall oral proficiency (Kost, 2004; Sanders, 2005; Satar & Özdener, 2008; Chang, 2007, 2008; Wang, 2010), pronunciation (Alastuey, 2010; Lord, 2008, Volle, 2005), anxiety (Satar & Özdener, 2008), accuracy (Xiao, 2007; Zheng, 2010; Sun, 2012; Volle, 2005; Pyun, 2003), fluency (Volle, 2005; Xiao, 2007; Sun 2012 ), lexical range/richness (Fitze, 2006; Huang & Hung, 2010), use of vocabulary (Fuente, 2003), speaking rate (Blake, 2009), and syntactic complexity (Abrams, 2003; Wang, 2010; Huang & Hung, 2010; Sun, 2012);
- e. transferability from CMC to oral discussion (Payne & Whitney, 2002; Wang, 2010; Zheng, 2010; Yang, 2006);
- f. different instructional strategies/task types of CMC (Li, 2008; Blake, 2000; Smith, 2004; Sauro & Smith, 2010);
- g. the relationship between working memory and oral proficiency development in a CMC context (Wang, 2010; Payne & Whitney, 2002; Payne & Ross, 2005; Zheng, 2010).

A number of studies have also investigated students' perceptions of CMC employed to develop oral ability (Sun, 2012; Xiao, 2007). This body of research on CMC reveals the following trends:

- a. there exists a relationship between working memory and oral proficiency, which can be developed indirectly via SCMC (Payne & Whitney, 2002; Wang, 2010; Zheng, 2010; Yang, 2006);
- b. students exposed to SCMC produced significantly more language than those in a F2F context (Chang, 2007);
- c. reduced pace of interaction in the electronic discussion restrains interlocutors from expanding a topic (Pyun, 2003);
- d. speaking was improved via text chat, and oral skill transfer was possible from text chat to F2F spoken language (Satar & Özdener, 2008; Sequeira, 2009);
- e. blogging with its personal and authentic nature might encourage students to focus more on meaning than accuracy (Sun, 2012);
- f. generally students hold a positive attitude toward the use of CMC in the L2 classroom for language practice (Lord, 2008; Wang, 2010; Xiao, 2007; Kost, 2004; Sun, 2012).

The ever growing body of research on CMC for speaking purposes demonstrates the burgeoning interest of both L2 researchers and practitioners in examining the feasibility and superiority of CMC over face-to-face interaction in an L2 context. However, previous research has demonstrated mixed and sometimes contradictory findings. For example, while Blake and his colleagues (2008) reported significantly better oral performance of a CMC group over a face-to-face group, such a finding was not found in a later study by Blake (2009). Also, the two studies conducted by Chang (2007, 2008) revealed contradictory findings. A brief count reveals that some studies found significantly better oral performance of the CMC group than the control (F2F) group (e.g. Abrams, 2003; AbuSeileek, 2007; Ahn, 2006; Chang, 2008; Satar & Özdener, 2008; Sequeira, 2009; Volle, 2005; Wang, 2010; Xiao, 2007; Blake, 2009; Chen, 2008; Huang & Hung, 2010; Kost, 2004; Li, 2008; Lord, 2008; Payne & Whitney, 2002), while other studies did not support such findings

(e.g. Blake *et al.*, 2008; Chang, 2007; Loewen & Erlam, 2006; Pyun, 2003; Sanders, 2005; Sun, 2012; Zheng, 2010).

The present meta-analysis investigates whether there is a causal relationship between spoken CMC and L2 oral proficiency development. To include as large as possible a body of empirical studies that address the use of CMC in L2 speaking, oral proficiency is defined as language learners' competence as demonstrated in key traits of oral interactions such as pronunciation, syntactic complexity, lexical complexity, density, richness, overall accuracy and fluency, while accepting that the measures of such traits may vary from study to study (Iwashita, Brown, McNamara & O'Hagan, 2008; Iwashita, 2010; Norris & Ortega, 2000; Ortega, 2003) Specifically, the meta-analysis addresses the following questions:

1. Compared to face-to-face interaction or no interaction at all, how effective is CMC in promoting L2 oral proficiency?
2. Is the effectiveness of CMC related to the type of data collected (e.g. naturalistic vs. elicited data) and assessment task (e.g. oral interview, response to a topic)?
3. What components of oral competences (fluency, lexical, accuracy, etc.) is CMC most likely to facilitate?
4. Are certain task types (e.g. jigsaw, information exchange, etc.) more effective than others in promoting oral proficiency in a CMC environment?
5. Is there a relationship between treatment duration and CMC effectiveness?

### 3 Method

To include as much literature as possible that investigates the use of CMC in an L2 classroom for speaking/oral purposes, the analyst first browsed key words used in primary studies in major journals that publish studies related to language and technology. The major key words identified include: (synchronous/asynchronous) CMC, (learner) interaction, CALL, communication strategies, and SLA. These key words were then used in combination with oral/speaking and tools/techniques/interface features of CMC to form a list of search keywords. ESL, EFL, language learning, and L2 were used interchangeably to limit the search outcomes. A series of keyword searches was then used to search for eligible studies published in (1) journals that publish SLA studies; (2) journals that publish studies that use technology in language learning/teaching; (3) dissertations and theses; (4) major educational/linguistic databases; (5) conference proceedings, working papers and technical reports; and (6) Google Scholar. To minimize the "file drawer" problem addressed by previous meta-analyses, the review pool of the current meta-analysis included both published and unpublished studies (Rosenthal, 1979; Hunter & Schmidt, 2004; Konstantopoulos & Hedges, 2004; Lipsey & Wilson, 2001; Li, 2010). The journals that were searched include *Language Learning & Technology*, *Computer Assisted Language Learning*, *ReCALL*, *System*, *CALICO*, *JALT CALL Journal*, *Language Learning*, *The Modern Language Journal*, *TESOL Quarterly*, *Canadian Modern Language Review*, *Foreign Language Annals*, *Second Language Research*, and *Studies in Second Language Acquisition*. The major databases searched include *Education Abstracts Full Text* (Wilson), *Education Resources Information Center* (ERIC), *ProQuest Psychology Journals*, *Springer Online Journal Archives*, *JSTOR – Arts & Sciences III Collection*, *EBSCOhost*, *Linguistics and Language Behavior Abstracts* (LLBA), and the *Social Science Citation Index*.

Additional steps were performed with an aim to uncovering studies that were eligible but had not been identified through the above major searches. These steps included expanding the scope of journals to those that published studies on educational technology not necessarily in language learning contexts. This step thus further searched the *Journal of Computer-assisted Learning*, the *British Journal of Educational Technology*, *Educational Technology Research & Development (ETR&D)*, and *Computers & Education*. A further step was to manually search the references of the primary studies to identify potentially eligible studies.

### 3.1 Inclusion/exclusion criteria

Once the potential primary studies had been identified through the steps discussed above, they were carefully read and evaluated according to the inclusion and exclusion criteria. These criteria were constructed based on the research questions guiding the present meta-analysis. In order to be included, the study had to:

1. be published between 2000 and 2012;
2. compare a treatment that used some form of CMC (e.g. email, chat, video/audio conferencing, discussion forums, CMS, Moodle, etc.) with face-to-face communication or no communication;
3. administer a measurement of participants' oral skills performance;
4. involve instructional effects of CMC on any oral proficiency feature (e.g. fluency, accuracy, speech rate, etc.);
5. use an experimental or quasi-experimental design;
6. recruit participants who were L2 learners;
7. include quantitative data suitable for a meta-analysis.

A considerable number of studies were excluded from the current meta-analysis since most CMC studies adopt qualitative data interpretation, e.g. with discourse analysis of interactional features in CMC environments, or a comparison of these features between the CMC and face-to-face environments relying on chat scripts, audio/video transcripts, postings or other sources of data. Furthermore, a substantial number of studies (e.g. Smith, 2003a, 2003b) compare the difference in interactional features between different tasks/communication strategies adopted in CMC environments. These studies, although providing very rich description of what happens in the interaction process in the CMC environment by manipulating different tasks (decision-making, jigsaw, etc.) and communication strategies, were excluded since they do not single out the effect of CMC by comparing it with face-to-face interaction. As such, the studies included in the current meta-analysis might not be a representation of the entire domain of CMC research, and subsequently the findings can only be generalized or interpreted within the boundaries defined by the inclusion and exclusion criteria specified above.

### 3.2 Coding scheme

A coding scheme was developed to describe and record the characteristics of each primary study. Coding schemes developed in previous SLA meta-analyses were used as the primary reference for the draft coding scheme which included three parts: publication features,

design features, and learner features. The scheme was first piloted with ten studies and then refined to incorporate features that were not captured. The items included in the final coding scheme related to methodological features are summarized in Table 1. To enhance coding reliability, three independent coders coded each of the studies (the meta-analyst and two MA TESOL students). The results of the coding were compared item by item; for items that received consistent coding by at least two coders, the code was kept as it was; for items for which discrepancy was found among the three coders, disagreements were discussed and a consensus was reached.

### 3.3 Effect size calculation

In this meta-analysis, effect size based on Hedge's  $g$  is computed to represent the effectiveness of CMC on oral proficiency development. The calculation of  $g$  is based on the following formula which can be interpreted as the magnitude of an observed difference between the control (no communication) or comparable (face-to-face communication) situations and CMC as expressed by the number of standard deviations (Norris & Ortega, 2000; Lipsey & Wilson, 2001). In practice, the mean difference in a study is divided by its pooled standard deviation to calculate a standardized mean difference, which is used to represent each study and is then compared with other studies. A weighted mean is computed by the inverse of its variance to adjust for the difference in sample size. The weighted mean is believed to be able to estimate the summary mean (see Formula 1).

$$\hat{g} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{(N_{\text{total}}-2)}}} \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

#### Formula 1 Weighted effect size

The effect sizes thus obtained can be interpreted according to standard deviation or average percentile gain or loss. For an effect size of 0.50, this means that the experimental group students scored 0.50 of a standard deviation above the scores of the control group students on average, or the experimental group students scored in the 69th percentile of the control group, or students scoring in the 50th percentile on achievement tests before intervention (control group) would be predicted to score in the 69th percentile after the intervention (experimental group).

### 3.4 Effect size calculation for studies with different designs

Effect size calculation is rather straightforward if it represents mean difference between a control and experimental group on one dependent variable. However, this between-group contrast in only one dependent variable is not the norm in SLA studies. A significant number of studies included in this analysis adopted between-group comparisons with more than two groups for more than one outcome. Some outcomes were also measured at more than one level. A single study may thus produce several effect sizes. A variety of ways to address the problems in which multiple effect sizes are generated for studies employing different research designs have been discussed in the literature (see Norris and Ortega for a review, 2006). The following procedures describe how effect sizes were handled in this analysis following Norris and Ortega (2000).



Table 1 *Coding scheme: methodological variables*

Variables	Description
<b>Measurement</b>	
Standardized	Well-established national or international standardized test such as TOEFL, GEPT (Taiwan's official General English Proficiency Test), or ACTFL
In-house	Developed by the researcher for the purpose of the study
Institutional	Developed and designed by an institution to screen or place students into different proficiency levels, for example, BYU Web CAPE
Impressionist	No formal measurement was developed or provided and performance was based on the researcher's subjective judgment
Mixed	More than one measurement was adopted
<b>Task type</b>	
Opinion exchange	Can take many different forms of which the major one will engage students in open discussions or exchange of ideas, opinions or information
Jigsaw	The flow of information is two-way in that each participant plays the role of both information holder and supplier, and they engage in communication to complete a task with one single outcome
Information gap	One participant has to hold part of the information that the other interlocutor does not have, and there is a need for communication to put together the information to complete a task
Decision-making	Students are asked to work together toward a single resolution of a problem for which a number of outcomes may be possible
Mixed	More than one kind of task was used
<b>Temporality</b>	
ACMC	There was a delay in message transformation
SCMC	Communication occurred in real time
Both	Both modes of communication were used interchangeably
<b>Modality</b>	
Text	Written communication
Voice	Oral communication
Both	Both written and oral communication
<b>Interlocutor</b>	
Peers	Classmates
Teacher	Instructors
Mixed	Native speakers recruited for the purpose of the study and peers

Table 1 *Continued*

Variables	Description
<b>Grouping</b>	
Pair	2 persons
Small	3–6 persons
Large	>6 persons
Class	no grouping
Mixed	more than one grouping
<b>Assessment task</b>	
Interaction transcripts	Transcripts from online or offline interactions
Speech-giving	Give a prepared speech related to the main issues of the units students studied
Oral presentation	Oral PowerPoint presentation
Oral interview	Consisted of a warm-up phase, level checks and probes, and a final wind-down phase or in the form of a conference
Read aloud	Reading aloud passages
Response to topics	Recording a response to a written prompt/topic
Information exchange	Students share information with each other about topics in a questionnaire
Multiple tasks	More than one task was employed in the assessment; examples included read aloud, listen and repeat, say the opposite, answer short questions, build sentences from jumbled-up word combinations, answer open-ended questions, and retell stories
<b>Treatment length</b>	
Short	≤15 weeks
Long	>15 weeks

1. When estimating the overall effectiveness of CMC, the principle of one effect size for one study is adhered to as closely as possible. So for studies which compare two CMC groups (e.g. ACMC and SCMC0) with a face-to-face or control group, the effect size was calculated respectively for the difference between ACMC and Control/F2F and for the difference between SCMC and Control/F2F. The two calculated effect sizes were then combined and averaged to represent the overall effect of CMC for that study without differentiating temporality.
2. For studies adopting a between-group design with repeated measures (i.e. pretest and posttest), the effect size was calculated based on (1) the difference between the two groups in the posttest if no significant difference was reported in the pretest, or (2) the gain scores between the pretest and posttest.
3. For studies using a within-group pretest/posttest design, the effect size was calculated based on the difference between the pretest and posttest.
4. For studies that administered both an immediate and a delayed posttest, only the immediate posttest was used since the aim was not to measure the long-term effect of CMC.
5. For studies which administered more than one measurement, the effect size was calculated for each measurement and was then averaged and treated as the representative ES for that study (Li, 2010). However, when different variables were targeted for comparison, individual effect size was treated as the unit of analysis. For example, studies might report separate scores for syntactic complexity, number of words, syntactic richness, etc. to reflect overall oral proficiency. When the question was what aspect of oral proficiency is CMC most effective at facilitating, these separate scores were not combined but were compared across studies.

### 3.5 Confidence intervals

In the analysis, 95% confidence intervals (CI) were calculated to test the statistical trustworthiness of the individual and averaged effect sizes. If the resulting confidence intervals include zero, it means that the effect size calculated might be due to chance, and it is possible that the true effect size is zero and thus not trustworthy. The confidence intervals were calculated following procedures suggested by Lipsey and Wilson (2001).

## 4 Results

In total, 25 primary studies were included in this meta-analysis. In order to provide an overall profile of the included studies, substantive features across studies were tallied. The total sample size for the included studies is 1,712. A wide range of N-sizes across studies was found, with a minimum of 16 and a maximum of 334 participants in a single study. Most of the included studies were conducted in English as a foreign language settings, and peers are almost the exclusive type of interlocutor when engaging students in communication activities. The majority of studies were published in journals (N = 14); there are also seven dissertations and three theses. German, English, and Spanish are the target languages investigated. Regarding participants' L1 across studies, it was found that nine studies dealt with participants with mixed L1s and another eleven recruited participants

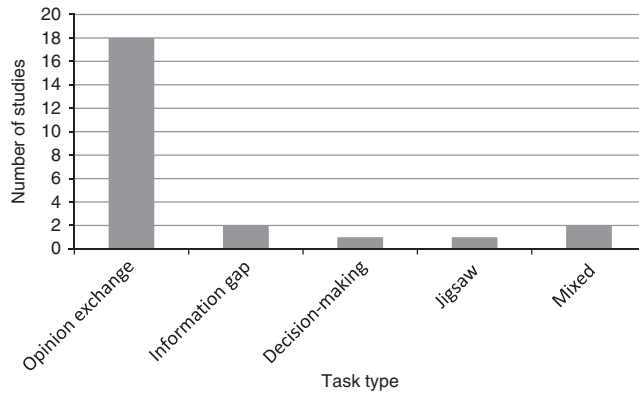


Fig. 1. Frequency of task type

whose L1 was an Asian language; three studies involved participants whose native languages are less commonly studied, i.e. Turkish and Arabic. This pattern indicated a prevalent research interest in EFL Asian learners. Furthermore, within this group, Chinese L1 speakers were the most often studied. The studies were categorized depending on the number of students per group when engaged in CMC activities. The result showed that pairs or small groups of three to five students are the groupings adopted by the majority of studies; only one engaged students in a large group of more than six learners, and another used no grouping at all. In terms of communication mode, more than half of the studies (56%) used voice chat, while approximately one third employed text-chat (36%); only two studies used both modes. Regarding treatment duration, 17 of the 25 studies lasted between 11 and 24 weeks; only one lasted for more than 24 weeks. With regard to temporality, seventeen studies (68 %) employed real-time synchronous communication, while only three (12%) adopted delayed asynchronous communication tasks, and five (20%) adopted both modes of communication. A wide variety of tools were used to explore the effect of CMC on oral proficiency, ranging from researcher-developed platforms specifically for the study purpose, to free chat room facilities provided by Skype, to discussion forums and class management systems such as Moodle, to name just a few.

*4.1.1 Type of task.* The instructional treatments in the 25 studies were categorized under one umbrella term as type of task, and were further classified into opinion (information) exchange, information gap, decision-making, jigsaw and mixed. This categorization is based on the typology established by Pica, Kanagy, and Falodun (1993), who incorporated two features (goal and activity) of discussion tasks to examine the effect of task type on the nature of interaction and language acquisition in SLA (Smith, 2003a). Focusing on the wide variety of details provided in each primary study in terms of the treatment design, each study was carefully reviewed and categorized into one of the five commonly used task types. As shown in Figure 1, opinion exchange is the dominant task employed by 18 of the 25 studies when engaging students in CMC. No more than two studies employed information gap, decision-making or jigsaw. This finding indicates a concentration of research interest on investigating the effect of CMC in the potentially least facilitative task type

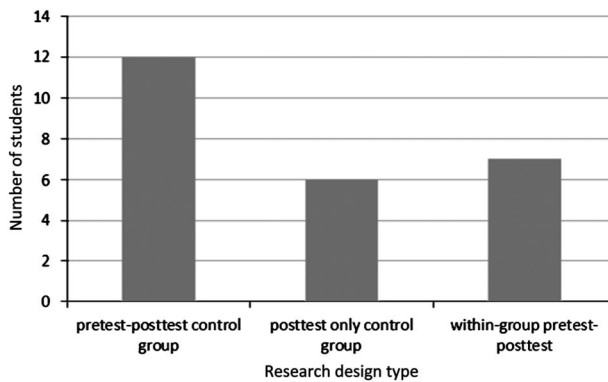


Fig. 2. Frequency of research design

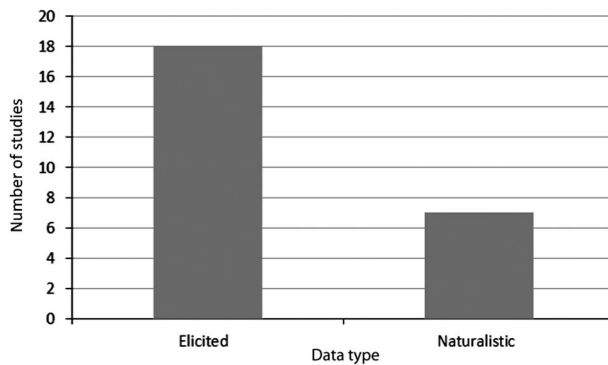


Fig. 3. Type of data provided in primary studies

(i.e. opinion exchange) rather than those deemed to be most facilitative for SLA (i.e. jigsaw, information gap, and problem solving).

**4.1.2 Research design and type of data.** Figure 2 reports the research design and the type of data provided for analysis across study reports. Among the most notable findings is the inclusion of a control group in the studies. Of the 25 studies, eighteen included a control group to contrast with the treatment group, with fourteen of the control groups being face-to-face and another four having no interaction. A second point is that the majority of the included studies ( $N=19$ ) administered a pretest to establish the threshold ability of participants. In terms of the research design, nearly half of the studies ( $N=12$ ) adopted a pretest/posttest design that included control and experimental groups, six used a posttest-only design with both control and experimental groups, and the remaining seven adopted a within-group pretest/posttest design that included only experimental groups. This pattern signals a widely received recognition of the importance of including a control or comparison group in a cause-and-effect study to measure the absolute effectiveness of the instruction under investigation.

In Figure 3 the type of data collected from different types of outcome measures across studies is shown. Of the 25 studies, eighteen used some form of assessment task to elicit oral

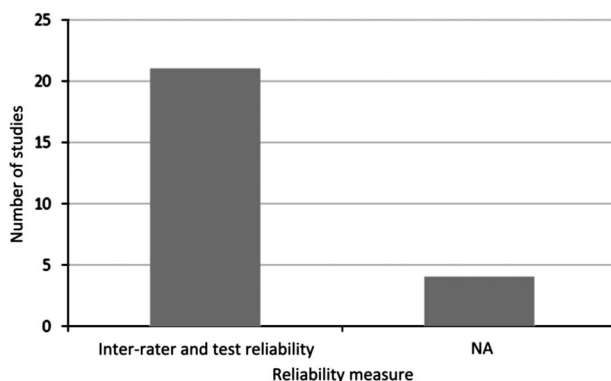


Fig. 4. Reliability reports in primary studies

performance from participants after the instruction, while only seven analyzed transcripts derived from naturalistic discussions without administering an oral test of any kind.

*4.1.3 Measure, assessment tasks and reported reliability.* Major trends in the assessment measures and tasks across the 25 studies can be found in the supplementary material. Of the eighteen studies that used some kind of assessment procedure to elicit students' oral performance, five adopted language tests. The remaining thirteen employed a wide range of performance-based assessment tasks. Oral interviews in the form of open-ended two-way information exchange or teacher-student conference was the most commonly adopted assessment task, followed by audio-recordings (hence one-way communication) of responses to assigned topics. Prepared speeches or presentations were also used in two studies. The five that assessed students' performance in a natural setting collected data from open discussion mediated by computer/technology.

Given that the reliability and validity of instruments are features associated with the quality of empirical studies (Cooper, 2009), the availability of validity and reliability in the primary studies was examined. Figure 4 shows that 21 studies reported either inter-rater reliability or test reliability; only four studies failed to provide any type of validation procedure for their measurements.

*4.1.4 Target oral components.* Figure 5 presents the frequency counts of the oral components measured in the primary studies. As shown, most studies evaluated students' overall performance by examining accuracy, fluency or holistic performance in oral production, while some studies specifically examined oral performance at the phonetic, lexical and syntactic level. Specifically, 16 of the 25 studies employed a holistic oral proficiency measure. Approximately one third measured syntactic complexity, followed by accuracy and fluency. Only two studies were interested in assessing the effect of CMC at the lexical level.

*4.1.5 Treatment duration.* After converting the length of the CMC-integrated instruction of all 25 studies into weeks, the study samples were further coded for short (fifteen weeks

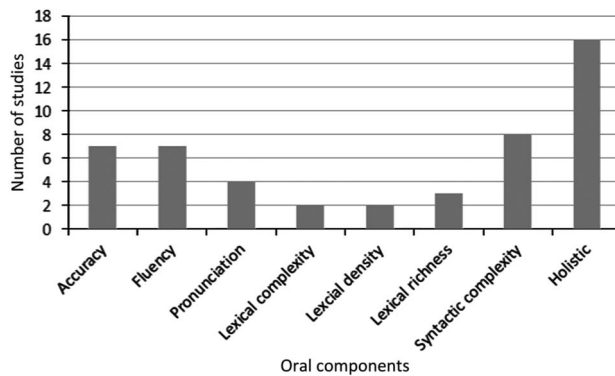


Fig. 5. Target oral components

Note: some studies provided a holistic score as well as scores for individual oral components, so the number of studies exceeds 25 in the figure.

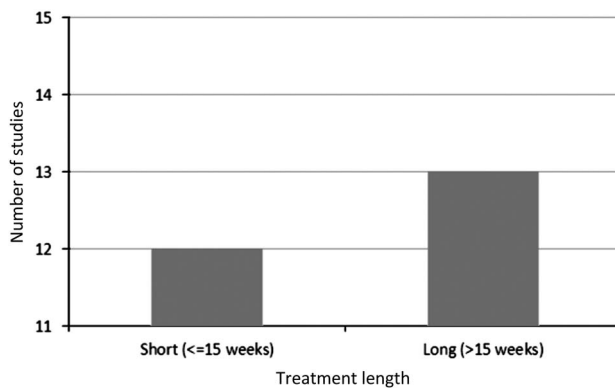


Fig. 6. Treatment duration

or less) and long (more than fifteen weeks) treatment duration categories. This cut-off point of fifteen weeks was arbitrarily determined to make the number of studies in either group within the treatment duration balanced. As Figure 6 indicates, twelve studies lasted for fifteen weeks or less, while thirteen lasted for more than fifteen weeks.

#### 4.2 The quantitative analysis

*Research question 1: Compared to face-to-face interaction or no interaction at all, how effective is CMC in promoting L2 oral proficiency?*

Our first research question concerns the overall effectiveness of CMC on L2 oral proficiency development, as indicated in the results across the 25 studies. The results showed that the overall mean effect size of the 25 primary studies was 0.40 with a standard error of 0.13, suggesting a moderate and positive effect of superiority of CMC over F2F on the immediate posttests. The lower and upper limits at the 95% confidence interval are 0.15 and 0.65, which does not include zero. The results indicate that the averaged effect size estimated

from these 25 studies can be generalized to the population at a very significant level of precision ( $p = 0.002$ ). A closer examination of the distribution of the effect sizes shows that they range between  $-0.62$  and  $1.82$ , with eight studies showing a negative effect of CMC on oral proficiency development and seventeen showing a positive effect. If we classify the effect sizes based on Cohen's (1988) suggestion that an effect size between  $0.20$  and  $0.50$  is considered small, between  $0.50$  and  $0.80$  moderate and  $0.8$  or above large, then the results indicated that of the seventeen studies that show positive effect, eight are considered to produce a small effect, three a moderate effect, and six a large effect. One striking finding is observed, though. There is a vast range in the magnitude of effects, clearly indicating a large standard deviation ( $SD = 0.65$ ) – larger in fact than the value of the average effect size. A few unusual effect sizes played a major role in this dispersion: two large negative effect sizes contributed by Zheng (2010) and Sequeira (2009); and large positive effect sizes contributed by AbuSeileek (2007), Chen (2008), Satar and Özdener (2008), and Xiao (2007). Furthermore, 17 of the 25 studies included zero in the 95% confidence interval, resulting in the average effect size having a lower confidence interval approaching zero (lower confidence is  $.15$ ). However, the number of primary studies that contributed to the average effect size ( $k = 25$ ) is large enough to produce trustworthy findings. Namely, the confidence interval for the mean effect size shows that the difference between the CMC and face-to-face groups is statistically significant, and the true effect, although small, is not zero.

To summarize the answer to the first research question, we can conclude that the effectiveness of CMC over F2F interaction is small and positive ( $g = 0.40$ ) and trustworthy (the true effect size falling anywhere between  $0.15$  and  $0.65$  standard deviation units) in terms of average change by CMC groups from pretest to posttest or when contrasted with a control/F2F group.

*Research question 2: Is the effectiveness of CMC related to the type of data collected (e.g. naturalistic vs. elicited data) and assessment task (e.g. oral interview, response to a topic)?*

To answer research question two, effect sizes were aggregated first according to the type of data so that the effects of CMC utilizing either naturalistic or elicited data could be compared. The results showed that more studies relying on naturalistic data collection yielded a negative effect ( $g = 0.15$ ) than those using elicited data ( $g = 0.50$ ). It should be noted that the confidence interval for the mean effect size aggregated from naturalistic data includes zero. On the other hand, the mean effect size calculated from the elicited studies was close to medium ( $g = 0.50$ ) and the confidence interval does not include zero. Q-test results indicate that the difference is trustworthy at a very significant level; however, such findings should not be taken at face value since the superiority of the elicited data studies may be due to the fact that the number of studies which elicited data is twice the number of naturalistic studies, and thus the result needs to be interpreted with caution. Effect sizes were aggregated again for studies that utilized elicited data according to the performance tasks that were employed to elicit oral performance. Of the performance tasks, reading aloud generated the largest effect size, and speech-giving produced the smallest. The remaining performance tasks yielded medium effects ranging from  $0.30$  to  $0.50$ . A small confidence interval between  $0.39$  and  $0.62$  was found for the eighteen observations, and the difference between the eight performance tasks was significant at the  $0.000$  level, suggesting a superiority effect of using reading aloud as a performance task to elicit the best oral performance of students to reflect the effectiveness of CMC. This pattern of results, though,



should not be taken for granted due to the uneven number of studies that contributed to the average effect size for each task type. There were three task types that were adopted in only one study (oral PowerPoint presentation, information exchange and reading aloud), and the number of studies using oral interviews and response to topics outweigh those that adopted other task types.

To summarize the answer to the second research question, we can conclude that the effectiveness of CMC compared to face-to-face interaction depends on the type of data collected for the outcome measure. Studies relying on elicited data are superior to those using naturalistic data, and reading aloud seems to be the task that could elicit the best oral performance from students. Both conclusions, however, are tentative given the unequal number of eligible studies included in each category.

*Research question 3: What components of oral competences (fluency, lexical, accuracy, etc.) is CMC most likely to facilitate?*

Our next research question asked whether the type of oral component measured in the dependent variable is systematically related to the magnitude of effect observed across the 25 studies included in this meta-analysis. Specifically, it was possible that CMC, with its various technological affordances/features, is more likely to lend itself to improvement in some oral components (e.g. fluency) than others (e.g. accuracy). Studies targeting oral fluency may have led to results that were in some way different from studies where accuracy is the target component. To answer this research question, a series of analyses was conducted which examined the components measured for each study. Based on the data reported in the included studies, the oral components are classified into accuracy, fluency, pronunciation, lexical richness, lexical density, lexical complexity, syntactic complexity, and holistic. Among the 25 studies, five assessed oral proficiency holistically, five assessed oral performance at the lexical level, eight were at the syntactic level, seven measured either accuracy or fluency of oral performance, while only four targeted pronunciation. The magnitude of effect on oral proficiency that CMC promotes is moderate ( $ES = 0.50$ ) when no specific oral components are targeted. In terms of the specific components, CMC has a roughly similar effect on pronunciation, lexical and syntactic level of oral production. Surprisingly, CMC appeared to be harmful for accuracy and fluency. The Q test, however, indicated that the above analyses are not significant,  $Q(5) = 4.458$ ,  $p = 0.486$ .

*Research question 4: Are certain task types (e.g., jigsaw, information exchange, etc.) more effective than others in promoting oral proficiency in the CMC environment?*

Of the 25 primary studies in this meta-analysis, 18 used information exchange as the communication task, as shown in Table 2. Examples of this kind of task included having participants engage in discussions of cultural texts or video (Payne & Whitney, 2002) or an open-ended discussion of prompt topics (Pyun, 2003). All the other studies used jigsaw, information gap or decision-making, which are argued to elicit more negotiated interaction than information exchange, based on Pica *et al.*'s (1993) typology. Moreover, these three tasks accounted for no more than one-fifth of the primary studies. Two studies combined different tasks (AbuSeileek, 2007; Satar & Özdener, 2008). As shown in Table 2, there was considerable variability in the effect sizes calculated based on task type employed in the primary studies, ranging from 1.49 to 0.16. Overall, the study that used decision-making generated the largest effect size, followed by studies that used more than one task type.

Table 2 Results of moderator analysis on task type

Task	N	ES	SE	Upper CI	Lower CI
Opinion exchange	18	0.16	0.06	0.04	0.27
Jigsaw	1	-0.06	0.21	-0.47	0.35
Information gap	2	0.27	0.21	-0.14	0.68
Decision-making	1	1.49	0.44	0.62	2.36
Mixed	2	1.47	0.14	1.19	1.75

Notes.  $Q(4) = 84.442$ ,  $p = 0.000$

CI = Confidence Interval

Among the four tasks, jigsaw actually generated a negative effect on oral performance. Furthermore, as the most popular task employed by primary researchers, opinion-exchange studies produced the smallest effect size. These results, however, should again be interpreted with caution, since for all other task types other than opinion exchange, the number of studies is relatively small. Furthermore, the 95% confidence for both jigsaw and information-gap studies included zero, which indicates that the observed effect may be obtained by chance. The average effect size for both decision-making and mixed studies were similar, with the former covering a wider confidence interval (0.62 to 2.36) than the latter (1.19 to 1.75), which might hint that studies employing more than one task can produce a large effect with more precision.

*Research question 5: Is there a relationship between treatment duration and CMC effectiveness?*

Our final research question asked if the effectiveness of CMC differs depending on the duration of treatment as revealed at the point of an immediate post-test. While recognizing that it is important for the treatment duration to take into account the frequency of CMC sessions within a given time period, and hence the time-on-task, this was nevertheless not possible since researchers in the primary studies rarely provided such information. Duration of treatment instruction was converted into weeks and dichotomously coded for short (fifteen weeks or less) and long (more than fifteen weeks). For studies reporting the duration as one semester, eighteen weeks was used and for studies that reported an academic quarter, twelve weeks (equal to four quarters/per year) was used to represent the treatment length. As previously noted, the cut-off point of fifteen weeks was determined arbitrarily so as to include approximately equal numbers of studies in the categories. Results showed that the average effect size of the study samples where the CMC intervention lasted fifteen weeks or less was larger ( $ES = 0.45$ ;  $SE = 0.08$ ) than the average effect size obtained for the longer studies ( $ES = 0.25$ ;  $SE = 0.06$ ). It might be more reasonable to speculate that L2 oral proficiency requires considerably longer treatment duration than other target skills (reading, listening, etc.) in order for the effects to be observable; yet the results suggest a contradictory picture. Caution is, however, necessary in interpreting the present finding. First, the confidence intervals around the two average effect sizes overlap, and the observed difference is not large ( $= 0.09$ ); hence, although the Q test indicated a significant difference, the observed difference is not trustworthy. Second, three very large effect sizes are contributed by the studies in the long treatment category (Chen, 2008; Satar & Özdener, 2008; Xiao, 2007). The excessive weight that these outliers exert on the average effect size may result

in its size being larger than that of the short treatment duration. A third and even more important caveat is that the long treatment duration category in the present meta-analysis had more study samples associated with the one-group only pretest/posttest design than the short treatment category (five of the six one-group pretest/posttest designs in the long treatment group and only one in the short treatment group). In light of this imbalance, it is impossible to ascertain whether the difference between the average effect sizes of the two treatment duration groups of study samples was related to treatment length (that is, long vs. short), or merely to the dominant presence of the one-group pretest/posttest design condition in the long treatment group. Thus, as research design and treatment duration are confounded, an answer to this research question remains only tentative.

## 5 Discussion

By carefully following procedures typically performed when conducting meta-analyses, and providing detailed information on the decisions we made at each step of the meta-analysis, we have achieved our major aim of examining whether communication mediated by computers/technology can be at least as effective as face-to-face interaction. We were also able to provide a comprehensive profile of primary studies in this domain by calculating frequencies of substantive features identified in our codebook. The methodological features that researchers typically manipulate in (quasi)experimental studies were also located and compared across studies to reveal the potential or consistent effects of such virtual communication on oral development in the L2 classroom. We found that in primary studies of CMC for oral development, opinion exchange is the most prevalent type of task used to elicit communication between L2 learners. Tasks that were more likely to trigger negotiations and prompt output (e.g. jigsaw and information gap) were rarely used. This research synthesis also reveals a tendency to include a comparison group (e.g. face-to-face) or a real control (no CMC) and administer a pretest to establish the equivalence of treatment and control groups in the studies to accurately pin down the treatment effect. In terms of the assessment, researchers of the primary studies have a preference for formal assessments shortly after the treatment to determine its effects on oral development. Naturalistic data elicited while learners were engaged in the interactions were rarely used or analyzed to determine such effects. Regarding the type of assessment to determine oral proficiency, there is a tendency to use researcher-developed, performance-based tasks that involve two-way communication such as oral interviews; very few studies used high-stakes, standardized language tests. We also noted that holistic assessment is the overwhelming scoring strategy adopted by the researchers, and when individual oral components were assessed separately, fluency and accuracy were the two most common indices used, followed by pronunciation, syntactic, and lexical components in oral production.

Beyond descriptive analyses, we carried out a series of moderator analyses of important study features with an aim to systematically establish the conditions in which CMC may be more beneficial for oral development. The evidence based on the calculation and aggregation of effect sizes contributed by the 25 studies included in this meta-analysis has provided the following insights. First, concerning the overall effectiveness of CMC on oral performance, a small to moderate and statistically significant mean effect size of 0.40 was found for the contrasts between CMC treatment groups and non-CMC groups, suggesting a greater likelihood of improved oral performance if L2 learners were exposed to communication

scaffolding enhanced by technology. The difference in oral performance between the contrast groups could be up to 0.65 standard deviation units. This result corroborated findings from Lin, Hung and Liou (2013), Taylor (2009), Yun (2011) and Grgurović, Chapelle and Shelley (2013), whose studies revealed a small to moderate effect of CALL instruction compared to traditional language pedagogy without access to any kind of technology. Our results lend further support to the finding that, across the various conditions in which technology is used for various language learning purposes, it is at least equal or even superior to instruction without technology (Grgurović *et al.*, 2013). Second, there was a potential relationship between size of effect and (1) task type, (2) treatment length, (3) measurement type, and (4) oral component. However, the observed relationship has to be interpreted as suggestive rather than conclusive due to gaps in the data. Some categories under comparison involved only a few eligible studies, preventing us from making reliable conclusions based on uneven or insufficient evidence.

Bearing in mind that the conclusions may be suggestive rather than conclusive, we found that studies that relied on naturalistic data (transcript) produced a smaller and negative effect size than those that employed elicited data. Among the elicited data, reading aloud was the task that could elicit the best oral performance from students. Two exploratory reasons may be provided here. Naturalistic data were collected mostly when L2 learners were engaged in interactions, during which they endeavored to undertake several rounds of hypothesis-testing, and therefore the data might reveal many interlanguage features, which resemble native-like utterances. On the other hand, elicited data were mostly collected after a certain amount of exposure of the L2 learners to the CMC treatment instruction, and some performances were called for to reflect the treatment effect. In other words, the time point at which these two types of data were collected might explain the difference. The former was collected while the treatment pedagogy was in progress, while the latter was collected after the treatment had been completed. The latter type of data might be more reliable and more robust since in these studies time was allotted for the exposure of the subjects to an experimental treatment. Furthermore, the naturalistic data were mostly collected from one-shot studies on a selective interaction, which might range from several minutes to one hour. The ability of such selective data to represent the effectiveness of CMC instruction may be questionable. Further, in his study arguing that the interpretation of CMC data might be very different depending on the data collection method, Smith (2008) revealed potential flaws in using printed chat logs as the major and sole source of CMC data. He claimed that chat logs are useful in interpreting CMC interaction; however, a hard copy transcript of the chat might not successfully capture “interesting elements of online interaction such as scrolling” (p. 89). This is because the L2 learners might have edited their messages before they hit “send”. This implies that naturalistic data collection using chat logs might not be sensitive enough to uncover what actually happens in the interactions in sufficient detail.

Among the eighteen studies that used performance-based assessment tasks to elicit data, reading aloud generated the largest effect size. More challenging tasks such as oral interviews, information exchange and responses to topics yielded smaller effect sizes ranging between 0.46 and 0.52. This result may be plausible given that reading aloud, as a planned oral reading of printed materials, does not require L2 learners to resort to their linguistic repertoire for oral production. As a typical strategy used in reading classrooms to foster comprehension and critical thinking, reading aloud does not engage learners in language processing that requires the same cognitive effort as required in authentic two-way oral

communication. That is, this task only taps very superficial oral communication skills. On the other hand, in two-way communication tasks such as oral interviews, L2 learners are pushed to come up with appropriate reactions based on specific topics and target linguistic features. The simultaneous demands on the limited capacity of working memory, the learners' linguistic repertoire and their ability to use appropriate communication strategies when confronted with communication breakdowns allow tasks of this kind to more accurately diagnose oral proficiency. The discussion, however, is tentative considering the evidence that the number of eligible studies included in each category revealed a big gap. Furthermore, to accurately ascertain whether assessment tasks have an effect on oral performance, future research would be needed in which all variables are held constant and oral performances are elicited using different assessment tasks for each participant.

Regarding the tasks employed in the primary studies to engage students in various forms of interaction, the meta-analysis found superior effect of decision-making on oral performance to other tasks such as opinion exchange, jigsaw and information gap. Pica *et al.* (1993) proposed different ways in which features of activities and goals can be realized in different communication-oriented tasks, and established a typology that consists of five types of tasks, namely jigsaw, information gap, problem solving, decision making, and opinion exchange. This typology is based on task relationships, requirements, goals, and outcomes and their possibilities of providing participants with comprehensible input, providing feedback on production and assisting them in modifying their interlanguage. Based on this typology, the optimal conditions for a task to trigger L2 learning/acquisition include collaboration between the L2 participants in which they work hard to understand each other, provide or receive the necessary feedback to sustain mutual communication, and negotiate any miscomprehension to complete a task that has only one acceptable outcome. Following these principles, jigsaw and information gap tasks are more restrictive in the kind of relationship participants can develop and in their interaction requirements; therefore they are believed to be more likely to induce L2 learning compared to other types of tasks. In this meta-analysis, however, jigsaw studies produced a negative effect on oral development, and information gap studies were only slightly more effective than opinion exchange. Surprisingly, decision-making that does not require mutual interaction and has more than one possible outcome produced the largest effect size. The larger gains obtained in the decision-making study should not be taken as definitive, however; it is merely indicative, as only one of the primary studies involved decision-making. The relatively larger number of studies that focused on opinion exchange, although not effective, revealed a concentration of interest in using less restrictive tasks in virtual discussions. To test if task type would differentiate the effect of CMC on oral development, more studies using jigsaw, information gap, and decision-making would have to be conducted.

To determine what aspects of oral performance CMC is more likely to enhance compared to face-to-face communication, we classified the dependent variables measured in the primary studies into six categories: holistic, accuracy, fluency, pronunciation, lexical, and syntactic. We found that studies which adopted a holistic measure yielded larger effect sizes (0.50) than studies which measured separate aspects of oral performance. Among the five individually measured oral components, studies that measured accuracy and fluency yielded small and negative effect sizes. The mean effect size calculated from studies that measured lexical aspects of oral proficiency was the largest compared to studies that measured pronunciation and syntactic features. However, the observed differences were minimal and

did not reach a significant level, indicating that CMC may be equally facilitative for oral development at the lexical, syntactic and pronunciation levels. As mentioned earlier, researchers in this domain discussed the benefits of CMC on L2 learning from an Interactionist perspective. The temporality and modality features of CMC provide cognitive benefits such as increased planning time and memory traces. These benefits are believed to be able to assist language learning over the course of negotiated interaction. Smith's (2004) study testing the Interactionist hypothesis found that L2 learners engaged in negotiated interaction over the course of task completion when they experienced communication breakdowns as a result of unfamiliar lexical items. L2 learners have also been observed to use and practice vocabulary beyond that related to the topic (Fitze, 2006).

The suggestion that CMC may bring about negative effects on fluency and accuracy is not surprising. In his study reviewing research that has investigated the effects of planning on L2 oral production, Ellis (2009) found that planning has a beneficial effect on fluency, confirming its role in oral production. In this meta-analysis, only three studies relied on an asynchronous mode of communication alone, while the remaining 22 studies involved synchronous communication or both. Planning is not possible when rapid interaction is taking place in real time, so synchronous communication might provide little chance to facilitate speaking fluency. Similarly, the rapid turn-taking in most chat room discussions may put accuracy at stake (Blake, 2009).

With regard to treatment length, the difference between short treatment studies lasting for up to fifteen weeks and longer ones lasting more than fifteen was close to statistical significance. The larger effect sizes were found for shorter treatment studies – a finding that runs against our expectations that a longer and more intensive exposure to oral practice would bring about better results. However, the amount of immersion time might not be the sole factor in determining the effectiveness of CMC in terms of oral development; the quality of interaction and the tasks designed to practice the interaction might mediate such an effect.

## 6 Conclusion and suggestions

This meta-analysis has presented an overview of empirical studies in the research domain of computer-mediated communication used for facilitating L2 oral development from 2000 to 2012. Rigorous and exhaustive searches found 25 studies that met the inclusion criteria. We have acknowledged throughout the text that the results obtained for the research questions prescribed for this meta-analysis are suggestive rather than definitive, primarily due to the weaknesses associated with the research design, the explicitness of the task description, and the quality of the reporting of results allowing for appropriate synthesis. Furthermore, all CMC tools in this meta-analysis were conflated without further considering their unique affordances. Zhao, Alvarez-Torres, Smith and Tan (2004) showed empirically that modality, spatiality, temporality and identity (the four features associated with most CMC sub-technologies) can be realized in different forms, which can greatly affect students' online behaviors and how interaction is carried out and shaped. Although it is beyond the purpose of this meta-analysis to synthesize the effects of differential CMC technological affordances on oral proficiency development, we would acknowledge that not being able to distinguish characteristics of CMC tools used in the primary studies is a limitation of the study. The paper has discussed several caveats regarding the findings; nevertheless, we might provide a number of suggestions for future research in the CMC domain as follows.

*Future research should provide adequate details of the assessment procedures and tasks when measuring L2 oral proficiency.*

Across the 25 studies included in the current meta-analysis, a vast range of oral abilities were assessed and assessment tasks carried out. Similarly, the way that oral assessments were incorporated into the research design varied tremendously, along with the depth of investment with which these assessments were conducted (Thomas, 2006). The assessment tasks could be as simple as reading aloud an assigned passage, or as complicated as full employment of a standardized oral proficiency test such as ACTFL. Most researchers did not justify their choice of specific assessment procedures or tasks, nor were operational definitions of oral proficiency provided. The above information regarding how particular assessment procedures were selected is important in helping us understand whether there is a valid connection between what the treatment instruction is designed to improve or facilitate, and whether those targeted areas were actually assessed. The operational definitions of oral proficiency and its sub-skills are also needed for research result comparison and replication purposes.

*Future research should adhere to a well-established task typology when using tasks in CMC environments.*

In this meta-analysis, we posited five types of tasks on the basis of a preliminary reading of the task descriptions of the 25 studies. We then classified these five types of tasks based on the well-known and widely applied task typology in task-based language learning developed by Pica *et al.* (1993). Despite attempting to be as objective as possible, any judgment inevitably relies on the details available in the primary studies. Unfortunately, many studies offered little information on task design, especially details such as goals of the task and the specific target oral (sub)skills that the task was intended to improve. We urge future researchers to elaborate on the principles of the tasks used to achieve the proposed goals, supplemented with adequate details on the nature, procedures and intended goals of the tasks. Most important of all, researchers might refer to established task typologies for guidelines for developing appropriate tasks for language learning purposes. By doing so, results across primary studies will share the same basis for comparison.

*Future research should examine the delayed effect of CMC on oral performance, study less-researched target languages and expand the task repertoire.*

In the field of SLA, it is generally agreed that language proficiency requires a sustained period of time to develop. It is common practice that in empirical studies researchers establish a point in time at which learning outcomes are best estimated. However, when these contrived conditions are withdrawn we are not sure if the same results can be obtained in a natural setting. In the current meta-analysis, only a couple of studies measured the long-term effect of CMC on oral proficiency. It is therefore recommended that future research examine the delayed effectiveness of CMC for us to understand its impact on oral development on a long-term basis. Furthermore, the majority of studies in this meta-analysis look at the effects of CMC on the learning of English as a target language. More research is needed in other languages to provide a clearer picture of the potential of CMC for oral proficiency across several target languages. Furthermore, in addition to our call for more details to be given when describing tasks employed in the interaction process, we also urge

future research to use tasks that are more likely to bring about negotiated communication such as jigsaw and information gap tasks. Currently, the overwhelming majority of tasks were opinion-exchange or open/free discussion. It might not be difficult to suspect preferences for such tasks because they are less restrictive in their goals and outcomes, and tend to be more flexible in terms of the materials and procedures needed to carry them out. However, if the theoretical motivation for use of CMC in language learning lies in its potential for facilitating negotiation between L2 learners so that their output is more comprehensible, then the tasks should be designed to reflect that, taking account of the various features, affordances, and constraints of CMC techniques.

### Note

The following supplementary data of this meta-analysis is provided via the *ReCALL* journal website:

1. Effect sizes contributed by each primary study based on oral proficiency traits
2. Research design and type of data for each primary study
3. Measures, assessment task and reported reliability in primary studies
4. Coding of study characteristics
5. Meta-analytic data for the 25 included studies
6. Average effect sizes for naturalistic and elicited data conditions
7. Effect sizes for oral proficiency traits
8. Effect sizes for studies with different treatment durations

### Acknowledgements

The author would like to express her immense gratitude to the anonymous reviewers for their valuable feedback on this paper. Additionally, this research was supported by the National Science Council of Taiwan, grant number NSC99-2410-H-007-082-MY2.

### References

- \*indicates primary studies included in this meta-analysis.
- \*Abrams, Z. I. (2003) The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal*, **87**(2): 157–167.
- \*AbuSeileek, A. F. (2007) Cooperative vs. individual learning of oral skills in a CALL environment. *Computer Assisted Language Learning*, **20**(5): 493–514.
- \*Ahn, H. (2006) *The impact of individual learner characteristics and synchronous computer-mediated communication on language production in learners of English*. University of Arizona, unpublished PhD.
- \*Alastuey, M. C. B. (2010) Synchronous-voice computer-mediated communication: Effects on pronunciation. *CALICO Journal*, **28**(1): 1–20.
- \*Alastuey, M. C. B. (2011) Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer Assisted Language Learning*, **24**(5): 419–432.
- Beauvois, M. H. (1998) Conversations in slow motion: Computer-mediated communication in the foreign language classroom. *Canadian Modern Language Review*, **54**(2): 198–217.
- \*Blake, C. (2009) Potential of text-based internet chats for improving oral fluency in a second language. *The Modern Language Journal*, **93**(2): 227–240.



- \*Blake, R. (2000) Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning and Technology*, **4**(1): 120–136.
- \*Blake, R., Wilson, N., Pearson, N., Cetto, M. and Pardo-Ballester, C. (2008) Measuring oral proficiency in distance, face-to-face and blended classrooms. *Language Learning & Technology*, **12**(3): 114–127.
- \*Chang, Y.-Y. (2007) The potential of synchronous text-based computer-mediated communication for second language acquisition. *Issues in Information Systems*, **8**(2): 355–361.
- \*Chang, Y.-Y. (2008) *A study of synchronous text-based computer-mediated communication and language learning outcomes*. Unpublished PhD thesis. Taipei: Tamkang University.
- \*Chen, F. (2008) *The effects of voiced-based synchronous and asynchronous CMC on EFL learners' oral proficiency*. National Taiwan Normal University, unpublished MA.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. New York: Routledge.
- Cooper, H. (2009) *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: Sage Publications.
- Cummins, J. (1986) Cultures in contact: Using classroom microcomputers for cultural exchange and reinforcement. *TESL Canada Journal*, **3**(2): 13–31.
- Dinsmore, T. H. (2006) Principles, parameters, and SLA: A retrospective meta-analytic investigation into adult L2 learners' access to Universal Grammar. In: Norris, J. M. and Ortega, L. (eds.), *Synthesizing research on language learning and teaching*. Philadelphia: John Benjamins, 53–90.
- Ellis, R. (1999) *Learning a second language through interaction*. New York: John Benjamins.
- Ellis, R. (2009) The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, **30**(4): 474–509.
- Fitze, M. (2006) Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning & Technology*, **10**(1): 67–86.
- 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.
- Grgurović, M., Chapelle, C. A. and Shelley, M. (2013) A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, **25**: 165–198.
- Hatch, E. (1978) Discourse analysis and second language acquisition. In E. Hatch (ed.), *Second Language Acquisition*. Rowley, Mass.: Newbury House.
- \*Huang, H.-T. D. and Hung, S.-T. A. (2010) Effects of electronic portfolios on EFL oral performance. *Asian EFL Journal*, **12**(2): 192–212.
- Hunter, J. and Schmidt, F. (2004) *Methods of meta-analysis*. London: Sage Publications.
- Iwashita, N. (2010) Features of oral proficiency in task performance by EFL and JFL learners. In: Prior, M. T., Watanabe, Y. and Lee, S.-K. (eds.), *Selected Proceedings of the 2008 Second Language Research Forum*. Somerville, MA: Cascadilla Proceedings Project, 32–47. <http://www.lingref.com/cpp/slrf/2008>
- Iwashita, N., Brown, A., McNamara, T. and O'Hagan, S. (2008) Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, **29**(1): 24–49.
- Konstantopoulos, S. and Hedges, L. V. (2004) Meta-analysis. In: Kaplan, D. W. (ed.), *Sage handbook of quantitative methodology for the social sciences*. New York: Sage Publications, 281–297.
- \*Kost, C. R. (2004) *An investigation of the effects of synchronous computer-mediated communication (CMC) on interlanguage development in beginning learners of German: Accuracy, proficiency and communication strategies*. University of Arizona, unpublished PhD.
- Krashen, S. D. (1985) *The input hypothesis: Issues and implications*. London: Longman.
- Levelt, W. J. M. (1989) *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- \*Li, L. J. (2008) *The impact of different levels of instructional strategy on synchronous CMC in improving college students' oral English proficiency and learning motivation*. National Cheng Kung University, unpublished MA.

- Li, S. (2010) The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, **60**(2): 309–365.
- Lin, S. M. (2009) *How computer-mediated communication affects ELL students' writing processes and writing performance*. Oklahoma University, unpublished doctoral dissertation.
- Lin, W.-C., Huang, H.-T. and Liou, H.-C. (2013) The effects of text-based SCMC on SLA: A meta-analysis. *Language Learning & Technology*, **12**(3): 123–142. <http://www.llt.msu.edu/issues/june2013/v17n2.pdf#page=128>
- Lipsey, M. W. and Wilson, D. B. (2001) *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Loewen, S. and Erlam, R. (2006) Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, **19**(1): 1–14.
- \*Lord, G. (2008) Podcasting communities and second language pronunciation. *Foreign Language Annals*, **41**(2): 374–389.
- Meskill, C. (1999) Computers as tools for sociocollaborative language learning. In: Cameron, K. (ed.), *CALL: Media, design and applications*. Lisse: Swets & Zeitlinger.
- Norris, J. M. and Ortega, L. (2000) Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, **50**: 417–528.
- Norris, J. M. and Ortega, L. (2006) *Synthesizing research on language learning and teaching*. Amsterdam/Philadelphia: John Benjamins.
- Ortega, L. (1999) Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, **21**: 109–148.
- Ortega, L. (2003) Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, **4**(24): 492–518.
- Payne, J. S. and Ross, B. M. (2005) Synchronous CMC working memory and L2 oral proficiency development. *Language Learning & Technology*, **9**(3): 35–54.
- \*Payne, J. S. and Whitney, P. J. (2002) Developing L2 oral proficiency through synchronous CMC: Output, working memory and Interlanguage development. *CALICO Journal*, **20**(1): 7–32.
- Peterson, M. (2009) Learner interaction in synchronous CMC: A sociocultural perspective. *Computer Assisted Language Learning*, **22**(4): 303–321.
- Pica, T., Kanagy, R. and Falodun, J. (1993) Choosing and using communication tasks for second language research and instruction. In: Crookes, G. and Gass, G. (eds.), *Tasks and second language learning*. Clevedon: Multilingual Matters, 9–34.
- \*Pyun, O. C. (2003) *Effects of networked language learning: A comparison between synchronous online discussions and face-to-face discussions*. Ohio State University, unpublished MA.
- Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin*, **86**: 638–641.
- \*Sanders, R. E. (2005) Redesigning introductory Spanish: Increased enrollment, online management, cost reduction and effects on student learning. *Foreign Language Annals*, **38**(4): 523–532.
- \*Satar, H. M. and Özdener, N. (2008) The effects of synchronous CMC on speaking proficiency and anxiety. *The Modern Language Journal*, **92**: 595–613.
- Sauro, S. and Smith, B. (2010) Investigating L2 performance in text chat. *Applied Linguistics*, **31**(4): 554–577.
- \*Sequeira, C. A. (2009) *Synchronous computer-mediated communication and second language proficiency*. University of Oregon, unpublished PhD.
- Simpson, J. (2005) *The discourse of computer-mediated communication: A study of online community*. University of Reading, unpublished PhD.
- Smith, B. (2003a) Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, **87**: 38–58.
- Smith, B. (2003b) The use of communication strategies in computer-mediated communication. *System*, **31**: 29–53.

- Smith, B. (2004) Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition*, **26**: 365–398.
- Smith, B. (2008) Methodological hurdles in capturing CMC data: The case of the missing self-repair. *Language Learning & Technology*, **12**(1): 85–103.
- Smith, B., Alvarez-Torres, M. and Zhao, Y. (2003) Features of CMC technologies and their impact on language learners' online interaction. *Computers in Human Behavior*, **19**: 703–729.
- \*Sun, Y.-C. (2012) Examining the effectiveness of extensive speaking practice via voice blogs in a foreign language learning context. *CALICO Journal*, **29**(3): 494–506.
- Taylor, A. M. (2009) CALL-based versus paper-based glosses: Is there a difference in reading comprehension? *CALICO Journal*, **27**(1): 147–160.
- Thomas, M. (2006) Research synthesis and historiography: The case of assessment of second language proficiency. In: Norris, J. and Ortega, L. (eds.), *Synthesizing research on language learning and teaching*. New York: John Benjamins, 280–298.
- \*Volle, L. M. (2005) Analyzing oral skills in voice e-mail and online interviews. *Language Learning & Technology*, **9**(3): 146–163.
- Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- \*Wang, C. Y. (2010) *A study comparing the effects of synchronous CMC and FTF interaction on L2 oral proficiency development for students with various working memory capacities*. National Tsing Hua University, unpublished dissertation.
- \*Xiao, M. (2007) *An empirical study of using internet-based desktop videoconferencing in an EFL setting*. Ohio University, unpublished PhD.
- \*Yang, M. L. (2006) *Exploring Taiwanese EFL students' responses to synchronous CMC: Effects on language use, learning and transfer, and perceptions*. University of Texas at Austin, unpublished PhD.
- Yun, J. (2011) The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, **24**(1): 39–58.
- Zhao, Y., Alvarez-Torres, M. J., Smith, B. and Tan, H. S. (2004) The non-neutrality of technology: A taxonomy of CMC technologies. *Journal of Educational Computing Research*, **30**: 23–55.
- \*Zheng, S. N. (2010) *The transferability from synchronous computer-mediated communication to oral discussion*. National Tsing Hua University, unpublished MA.