SYNCHRONOUS COMPUTER-MEDIATED COMMUNICATION AND INTERACTION:

A RESEARCH SYNTHESIS AND META-ANALYSIS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Nicole Ziegler, M.A.

Washington, DC
July 1, 2013
SYNCHRONOUS COMPUTER-MEDIATED COMMUNICATION AND INTERACTION:
A RESEARCH SYNTHESIS AND META-ANALYSIS

Nicole Ziegler, M.A.

Thesis Advisor: Alison Mackey, Ph.D.

ABSTRACT

The interaction approach to second language acquisition (SLA) suggests that changes that occur during conversation facilitate second language development by providing learners with opportunities to receive modified comprehensible input and interactional feedback, to produce output, and to notice gaps between their interlanguage and the target language through negotiation for meaning (Gass & Mackey, 2007; Long, 1996). Numerous empirical studies and meta-analyses have demonstrated positive effects for interaction (e.g. Keck et al., 2006; Mackey & Goo, 2007). However, despite an increasing body of evidence examining synchronous computer-mediated communication (SCMC) from an interactionist perspective (e.g. Blake, 2000; Smith, 2004; Yilmaz, 2012), the efficacy of interaction in computer-mediated contexts has been less clear so far.

The current study reports on a synthesis and meta-analysis on the relative effectiveness of interaction in SCMC and face-to-face (FTF) contexts. The primary studies included in the analysis were journal articles or dissertations published between 1990 and 2012, retrieved through searches in CALL and applied linguistics journals and texts.

Results demonstrate that interaction in both SCMC and FTF had positive impacts on L2 development, with comparative findings revealing a small, although non-significant, effect for
interaction in SCMC on measures of L2 learning outcomes. When the development of learners’ productive and receptive skills were the focus, results showed a small effect for SCMC on learners’ productive skills and a small effect for FTF on learners’ receptive skills. Additionally, analyses investigated learners’ written and oral skills, this time with results indicating a small effect for interaction in SCMC on written skills and a small effect for FTF interaction on oral skills. In sum, these results reveal several promising trends regarding interaction in SCMC. Furthermore, there were no significant differences between the two modes, suggesting, importantly, that in this meta-analysis the mode of communication has no significant impact on the positive developmental benefits associated with interaction.

This research also found a number of issues regarding methodological quality, including inconsistent reporting of statistical tests and procedures. Finally, areas are identified within SCMC interaction research that are ripe for further investigation and suggestions are made for methodological improvements and directions for future research.
ACKNOWLEDGEMENTS

The research and writing of this dissertation would not have been possible without the support and guidance of many people. I am grateful to my dissertation committee for their understanding and encouragement, and for their scholarly direction. I am especially indebted to my mentor, Dr. Alison Mackey, without whom this dissertation would not have been possible. Alison believed in my abilities even when I had doubts, and provided me with direction and support every step of the way. I would like to thank Dr. Bryan Smith for his enthusiasm and insight, and Dr. Rusan Chen for his advice and patient instruction, as well as his willingness to meet with me numerous times to discuss data and analyses. All of my committee members contributed to my success, and I hope to emulate their excellence in research and mentorship.

I would also like to thank my other professors at Georgetown University who have provided me with exemplary linguistic training and intellectual inspiration: Dr. Heidi Byrnes, Dr. Cecilia Castillo-Ayometzi, Dr. Jeff Connor-Linton, Dr. Ron Leow, Dr. Dan Loehr, Dr. John Norris, Dr. Lourdes Ortega, Dr. Robert Podesva, Dr. Patrick Rebuschat, Dr. Cristina Sanz, Dr. Deborah Tannen, Dr. Andrea Tyler, Dr. Linda Van Guilder, and Dr. Elizabeth Zsiga. To Dr. Rebecca Sachs, for her excellent feedback and brainstorming sessions on a wide variety of topics. To Dr. Michael Long of the University of Maryland, thank you for your encouragement and support, and for your willingness to always discuss research ideas and future endeavors with honesty and humor. Thank you also to Dr. Paul Portner and Dr. Heidi Hamilton for their support throughout this journey. To Dr. Randi Reppen and Dr. Doug Biber of Northern Arizona University for first encouraging me to pursue a doctorate, and Dr. Kim McDonough of Concordia University for her patience and advice during countless office hours. Thank you.
To my fellow graduate students, thank you for helping to keep me grounded and on-task, particularly when I felt overwhelmed. I am especially grateful to Phillip Hamrick and Luke Amoroso for their friendship and their invaluable help on this and many other projects. I will miss lunches at Epicurean and hot-pot dinners, and the stellar conversation that was always a part of these outings. I would also like to thank my colleagues (Jessica Cox, Marisa Filgueras Gómez, Marissa Fond, Hanna Jan, Katie Kim, Julie Lake, Haein Lauren Park, Mackenzie Price, Mari Sakai, Corinne Seals, Ellen Serafini, Sheena Shah, Laura Siebecker, Kerstin Sondermann, Kaitlyn Tagerelli, Laura West, Sarah Young, and Germán Zárate-Sández, and those whose names I may have omitted here) for the joy I have experienced through being your friend. Thank you to Erin Esch and Manela Diez for their tremendous patience in answering my questions and emails, and for their invaluable assistance regarding the administrative aspects of my degree. You have all been an important part of my Georgetown experience, and I am grateful.

I would like to dedicate this dissertation to my family: my husband, Traveler Terpening, my parents, Jack and Connie Ziegler, and my in-laws, Kim and Gordon Terpening. Your constant support, your endless hours of childcare, and your steady stream of encouragement were instrumental to my success. I will be forever grateful to you all for the time you spent taking care of me, whether it was bringing me a hot meal while I sat at the computer or letting me sleep late after a long night of coding data. And to my daughter, Cedar Katerinka, for always reminding me what is most important in life. Thank you for all that you have done to help me during this journey-this dissertation is for you.

Many thanks,

Nicole Ziegler
# TABLE OF CONTENTS

Chapter I: Introduction .................................................................................................................. 1

Chapter II: Review of the Literature ............................................................................................... 9
  2.1. The Interaction Approach ....................................................................................................... 9
    2.1.1. Historical foundations of the interaction approach .......................................................... 9
    2.1.2. Key constructs in the interaction approach ....................................................................... 12
    2.1.3. Input .............................................................................................................................. 12
    2.1.4. Output ........................................................................................................................... 13
    2.1.5. Feedback and negotiation .............................................................................................. 15
    2.1.6. Interaction and computer-mediated communication ...................................................... 17
    2.1.7. Learner characteristics and their role in interaction ...................................................... 24
    2.1.8. Comparing interaction in FTF and CMC ...................................................................... 27
  2.2. Research Questions ............................................................................................................. 30

Chapter III: Study Design and Methodology .................................................................................. 33
  3.1. Study Identification and Retrieval ......................................................................................... 33
  3.2. Inclusion and Exclusion Criteria ............................................................................................ 36
    3.2.1. Inclusion criteria ............................................................................................................ 37
    3.2.2. Exclusion criteria .......................................................................................................... 39
  3.3 Analysis .................................................................................................................................. 41
    3.3.1. Coding procedures ....................................................................................................... 41
    3.3.2. Study filtering process .................................................................................................. 43
    3.3.3. Substantive features ...................................................................................................... 44
    3.3.4. Methodological features ............................................................................................... 48
  3.4. Statistical Analysis ............................................................................................................... 50
    3.4.1. Fixed effects and random effects model ......................................................................... 51
    3.4.2. Calculation of effect sizes ............................................................................................ 53
    3.4.3. A Priori methodological decisions ................................................................................. 58

Chapter IV: Results ....................................................................................................................... 59
  4.1. Descriptive Results ............................................................................................................... 59
    4.1.1. Research distribution ................................................................................................... 59
    4.1.2. Publication bias ............................................................................................................ 59
    4.1.3. Research setting ............................................................................................................ 61
    4.1.4. Learner characteristics ................................................................................................. 64
4.1.5. Research design ................................................................. 66
4.1.6. Statistical analyses and reporting ........................................ 67
4.2. Meta-Analytic Results ............................................................... 68
4.2.1. Overall effectiveness of interaction in SCMC compared to FTF .... 69
4.2.2. Effectiveness of interaction in SCMC compared to FTF for receptive and productive outcomes .................................................. 74
4.2.3. Effectiveness of interaction in SCMC compared to FTF for oral and written outcomes .............................................................. 76
4.2.4. Mode of SCMC interaction, type of feedback, modified output, and target feature .............................................................. 79
4.3. Methodological Factors ............................................................. 84
4.3.1. Research setting (second or foreign language) ......................... 84
4.3.2. Research context (laboratory or classroom) ............................... 85
4.3.3. Interlocutor .............................................................................. 87
4.4. Summary of Results ................................................................... 89

Chapter V: Discussion ...................................................................... 92
5.1. Research Question 1: Compared to Face-to-Face Interaction, How Effective is Synchronous Computer-Mediated Interaction at Promoting L2 Learning Outcomes? .............................................. 92
5.2. Research Question 2: Compared to Face-to-Face Interaction, How Effective is Synchronous Computer-Mediated Interaction at Promoting Receptive and Productive L2 Learning Outcomes? ......................... 97
5.3. Research Question 3: Compared to Face-to-Face Interaction, How Effective is Synchronous Computer-Mediated Interaction at Promoting Oral and Written L2 Learning Outcomes? .............................. 104
5.4. Research Question 4a: How Does the Nature of the Research Setting (Foreign Language vs. Second Language) Impact the Effectiveness of Interaction in SCMC? ................................................................. 110
5.5. Research Question 4b: How Does the Type of Research Setting (Classroom vs. Laboratory) Impact the Effectiveness of Interaction in SCMC? .............................................................................. 114
5.6. Research Question 4c: How Does the Type of Interlocutor Impact the Effectiveness of Interaction in SCMC? .............................................................................. 115
5.7. Research Question 4d: How Does the Type of Target Feature Impact the Effectiveness of Interaction in SCMC? .............................................................................. 123
5.8. Research Question 4e,f,g: How Does the Mode of Interaction, Type of Feedback Provided During the Interaction, and Presence or Absence of Modified Output Opportunities
Mediate the Relationship between Interaction in SCMC and L2 Learning Outcomes? ................................................................. 127
5.8.1. Mode of interaction .......................................................... 127
5.8.2. Type of interactional feedback ........................................... 128
5.8.3. Presence or absence of modified output opportunities ......... 132

Chapter VI: Limitations and Future Research ......................................................... 136
6.1. Limitations ........................................................................ 136
6.2. Expanding the Research Agenda .............................................. 141
6.3. Implications ..................................................................... 150
   6.3.1. Theoretical implications .............................................. 150
   6.3.2. Methodological implications ....................................... 152
   6.3.3. Pedagogical implications ............................................ 155
6.4. Conclusion ..................................................................... 158

References .............................................................................. 161
LIST OF FIGURES

Figure 1. Included studies with their years of publication ........................................ 59
Figure 2. Publication bias: Funnel plot by effect sizes .............................................. 61
Figure 3. Forest plot of standardized mean effect sizes for overall L2 learning outcomes .......................................................... 72
Figure 4. Comparison of SCMC and FTF communication modes on overall L2 learning outcomes .................................................. 73
Figure 5. Comparison of SCMC and FTF communication modes on L2 receptive outcomes .......................................................... 75
Figure 6. Comparison of SCMC and FTF communication modes on L2 productive outcomes ...................................................... 76
Figure 7. Comparison of SCMC and FTF communication modes on L2 oral outcomes .............................................................. 77
Figure 8. Comparison of SCMC and FTF communication modes on L2 written outcomes ................................................................. 79
LIST OF TABLES

Table 1. Coding characteristics................................................................. 41
Table 2. Included studies by research characteristics ......................... 63
Table 3. Included studies by learner characteristics .......................... 65
Table 4. Descriptive effect size statistics for overall L2 learning outcomes ...... 70
Table 5. Descriptive effect size statistics for receptive L2 outcomes .......... 74
Table 6. Descriptive effect size statistics for productive L2 outcomes .......... 74
Table 7. Descriptive effect size statistics for oral L2 outcomes ................... 77
Table 8. Descriptive effect size statistics for written L2 outcomes ............ 79
Table 9. Summary of results: Comparative effects............................... 90
Table 10. Summary of results: Comparative effects of moderator variables........ 91
Chapter I: Introduction

The interaction approach to second language acquisition (SLA) suggests that conversational adjustments occurring during communication benefit second language (L2) development by providing learners with opportunities to receive modified comprehensible input and interactional feedback, to notice gaps between their interlanguage and the target language features, and to produce output (Gass & Mackey, 2007; Mackey, Abbuhl, & Gass, 2012; Mackey & Gass, 2006). By providing learners with negotiation opportunities, interaction facilitates acquisition through ‘the connection of input, internal learner capacities, particularly selective attention, and output in productive ways’ (Long, 1996, p. 451-452). These negotiations for meaning provide opportunities for learners to identify gaps between their interlanguage and the target language and produce modified output (Gass & Mackey, 2012; Long, 1996; Swain, 1985, 1995, 2005), thereby offering learners both receptive and productive linguistic benefits.

Having evolved from an hypothesis to an extended framework (Gass & Mackey, 2006, 2007; Jordan, 2005; Mackey, 2012), interaction research has expanded from initial examinations into the impact of conversational adjustment on learners’ second language development (Gass & Varonis, 1994; Mackey, 1999) to investigations into a wide range of interactional factors and processes, including the relationship between specific interactional constructs, such as various types of implicit and explicit corrective feedback, modified output, and learner outcomes (e.g. Ellis, 2007; Loewen & Philp, 2006; Oliver & Mackey, 2003). In addition, recent interaction research has investigated cognitive factors, including working memory capacity and attentional control (e.g.
Mackey, Adams, Stafford, & Winke, 2010; Mackey & Sachs, 2012), and social factors, such as peer relationships and attitudes (e.g. Philp & Mackey, 2011; Philp & Tognini, 2009). Numerous studies and meta-analyses have demonstrated the benefits of interaction on second language development (see Keck, Iberri-Shea, Tracy-Ventura, & Wa-Mbaleka, 2006; Mackey & Goo, 2007 for recent reviews), with research having investigated the impact of these features in lab and classroom based contexts with a variety of participant populations, providing empirical evidence for the benefits of interaction for children and adults, as well as for learners from diverse first language (L1) backgrounds.

Although interaction in face-to-face contexts is largely accepted as beneficial to L2 development (Keck, et al., 2006; Mackey & Goo, 2007), the relationship between interaction and learner outcomes within computer-mediated environments is less clear. In order to assess whether interaction is equally effective at promoting L2 development in CALL and face-to-face contexts, research has sought to establish the ability of computer mediated communication (CMC) to direct learners’ attention to specific target language features and to provide learners with opportunities to negotiate for meaning, to receive comprehensible input and corrective feedback, and to produce modified output (Smith, 2004). The growing body of research examining CMC and interaction suggests positive benefits within technology supported environments (e.g. Blake, 2000; de la Fuente, 2003; Lai & Li, 2011; Sauro, 2011; Sauro & Smith, 2010; Smith, 2004, 2005, 2010), with some studies indicating advantages within computer-mediated contexts over face-to-face contexts. For instance, research has demonstrated that text-based synchronous computer mediated communication (SCMC), which refers to real time interaction such as a live chat, may result in increased and more equitable learner participation (Beauvois, 1992;
Kelm, 1992; Kern, 1995; Warschauer, 1996), as well as greater quality of language production (Chun, 1994; Kelm, 1992; Kern, 1995; Warschauer, 1996), than in face-to-face interaction. In addition, CMC may result in greater amounts of learner output (Chun, 1994; Kelm, 1992) and may also provide learners with a less stressful environment than face-to-face interaction (Chun, 1998), potentially leading to reduced anxiety and improved willingness to communicate. CMC may also promote increased participation by less active students when compared with more active, dominating students (Warschauer, 1996; Kern, 1995; Chun, 1994) or encourage students’ risk-taking and hypothesis testing regarding language (Meunier, 1998).

Overall, CMC research has provided compelling evidence that interactional features found to be beneficial to L2 development in face-to-face contexts, such as negotiation for meaning and modified output, can and do occur in CALL environments (Beauvois, 1992; Kötter, 2003; Iwasaki & Oliver, 2003; Lee, 2001). For instance, tasks commonly found in face-to-face interaction research (Pica, Kanagy, & Falodun, 1993), such as decision-making and information gap tasks, have been shown to be successful in eliciting learner interaction and negotiation in a CMC environment (Blake, 2000; Pellettieri, 1999). Studies have also investigated negotiation for meaning, which includes conversational moves such as comprehension checks, confirmation checks, and clarification requests made in response to failures in communication, during SCMC by comparing different communication breakdowns leading to instances of negotiation (Toyoda & Harrison, 2002) and the frequency of negotiation routines when learners interacted with native (Tudini, 2003) or non-native speaking interlocutors (Fernández-Garcia & Martinez-Arbeloiz, 2002).
Similar to face-to-face interaction research (Keck et al., 2006; Mackey & Goo, 2007), lexical items have also been found to be more facilitative of negotiation than grammatical items in CMC (Blake, 2000; Fernández-Garcia & Martínez-Arbeiza, 2002; Pellettieri, 2000; Tudini, 2003). For example, Pellettieri found that a greater number of communication problems triggering negotiation occurred over lexical items than grammatical items, with the same results reported by Tudini (2003). Within face-to-face contexts, Williams (1999) found that the majority of Language Related Episodes (LREs) related to lexical features, with learners frequently failing to notice morphosyntactic features. More recently, Mackey, Gass, and McDonough (2000) also found that learners were more likely to perceive interactional feedback on lexical than on morphosyntactic and phonological items. In addition, de la Fuente’s (2003) results indicated that learners in both face-to-face and CMC negotiated interaction groups demonstrated greater comprehension on lexical items than learners in a non-negotiated interaction group. Long (2006) has also argued that the type of target form may be an influencing factor on the effectiveness of interactional feedback in facilitating L2 development, and the results of Mackey and Goo’s (2007) meta-analysis, which suggest that interaction was more effective in supporting lexical development than grammatical development, provide further evidence that learners may participate in negotiation more frequently regarding lexical items than grammatical items. Overall, these findings indicate that oral, face-to-face interaction and CMC display many of the same features and learning outcomes, with learners receiving similar opportunities for negotiation and feedback in both contexts.

Although research has indicated promising results regarding the effectiveness of interaction in facilitating L2 development in computer-mediated contexts, studies vary
along a variety of dimensions, including oral and written modalities, settings, and learner and interlocutor characteristics. In addition, methodologies and investigative techniques have evolved during the last two decades, with more recent, contemporary studies seeking to incorporate features from mainstream second language acquisition (SLA) (e.g. Chapelle, 2009; Smith, 2004; 2005; 2009; 2010). Following calls from researchers within the field of computer assisted language learning (Chapelle, 1997, 2005), these recent studies have sought to move CALL research from the fringes of L2 acquisition research by integrating theoretical models and frameworks from the field of SLA (Doughty & Long, 2003; Ortega, 2009; Salaberry, 2000) and by utilizing more rigorous research methodologies, such as control and comparison groups, to examine the design and effectiveness of computer mediated materials and applications (Chapelle, 2001), leading to appreciable changes in the nature of computer-based L2 instruction and research. In order to assess the progression of methodologies and to promote the advancement of findings from seminal CALL studies (Chapelle, 1994, 1998; Chun, 1994; Pelletieri, 1999), as well as improve our understanding of the empirical findings to date and to illuminate potential areas for future investigations of interaction within CALL contexts, a synthetic and meta-analytic approach is clearly needed.

Although previous syntheses and meta-analyses have identified areas within CALL needing further study, such as the exploration of computer-mediated technologies in K-12 contexts (Liu, Moore, Graham, & Lee, 2002), or the use of learner strategies and pedagogical tools unique to SCMC modalities (Sauro, 2011), no studies to date have comprehensively investigated the relationship between interaction, CMC, and L2 development. As the field of CALL continues to grow, there is an increasing number of
studies examining computer-centered contexts from an interactionist perspective, and
with the increasing use of technology in the foreign and second language classroom, there
is a growing need to better understand the pedagogical implications of interaction within
a computer-mediated context. A comprehensive synthesis and meta-analysis will allow
both researchers and educators to gain important insights into the effects of interaction on
learners’ L2 development in computer assisted environments.

Due to the growing body of work investigating L2 development in CMC from an
interactionist perspective, the collected research supports a comprehensive synthesis of
primary study findings to date. The present dissertation uses research synthesis and
meta-analysis methodologies to review and compare the results and methods found in
interaction research within CMC during the last two decades. Research syntheses,
through the systematic review of previous studies, offer researchers the opportunity to
compare and combine results across numerous primary studies, to conclusively answer
specific research questions, and identify gaps in research agendas and methodologies
(Norris & Ortega, 2006). Through the standardization of results and by comparing effect
sizes across multiple, primary studies, synthetic research provides a comprehensive
quantitative assessment of research practices within a particular domain (Lipsey &
Wilson, 2001; Norris & Ortega, 2000). By investigating and integrating accumulated
research, the synthetic approach is able to identify patterns and inconsistencies in
methodologies and findings (Cooper & Hedges, 1994) and examine broader research
questions than those of the original, individual studies (Miller & Pollock, 1994). In other
words, although the generalizability of single studies is limited by sampling qualities,
such as sample size and setting, by including all of the empirical studies relevant to a set
of research questions, meta-analyses enhance the generalizability of results (Felix, 2008). For instance, in their meta-analysis on task-based interaction, Keck et al. (2006) were able to investigate the impact of task design features, including task essentialness and opportunities for pushed output, as independent variables on the dependent variable of task effectiveness. In addition, syntheses and meta-analyses examine the impact of specific treatments, such as the impact of corrective feedback on L2 development, without the restrictions on objectivity often introduced by selective or convenience sampling (Suri, 2000). Conducting meta-analytic reviews provides researchers with the opportunity to produce new findings that go beyond the results of any of the individual primary studies, generating conclusions across studies and allowing researchers to propose specific recommendations for future investigations.

Building on previous meta-analytic and synthetic research (Mackey & Goo, 2007; Sauro, 2011), the present dissertation examines the last two decades of interaction research within the field of CALL, investigating whether well-established effects of interaction on L2 development can be found in computer-mediated contexts, and how empirical investigations of interactional features in CALL have evolved from the pre-2000 era to more recent, contemporary studies. Focusing on the advancement of methodologies and findings from early CALL studies (Chapelle, 1994, 1998; Chun, 1994; Pelletieri, 1999), as well as the use of changing data collection techniques during the development of the field, such as the recent use of eye-tracking (Smith, 2010) and virtual learning environments (Canado, 2010), this research investigates the relationship between L2 development, SCMC, and interactional features, including corrective feedback, negotiation, and opportunities for modified output, thereby furthering our
understanding of which features of a computer-mediated environment might influence the success of L2 acquisition.

Responding to the increasing use of technology in educational contexts, this research makes valuable contributions to our understanding of the interface between SLA theory and computer-mediated environments, and provides important data regarding the effectiveness and implementation of SCMC in the second and foreign language classroom. Furthermore, the results of the current study have implications for the individual research fields of CALL and interaction by identifying areas of potential improvement in the methodology and by exploring possible research agendas in CALL-based SLA interaction research, furthering our understanding surrounding the utilization of new and emerging media for language use and L2 learning, and helping to shape the policy and practice of future computer-mediated empirical investigations that will move the field forward.
2.1. The Interaction Approach

2.1.1. Historical foundations of the interaction approach

The foundations of the interaction approach are rooted in the late 1970s and early 1980s research of Evelyn Hatch and Michael Long, respectively. This research, beginning with Hatch’s seminal paper examining NS-NNS discourse (1978), called attention to the idea that that learners’ knowledge of form and structure could be acquired through their conversational interactions with more advanced interlocutors, highlighting the idea that language learning, including the development of syntactic knowledge, ‘evolves out of learning how to carry on conversations, out of learning how to communicate’ (Hatch, 1978, p. 63). Building on Hatch’s work, Long (1980, 1981, 1983) investigated NS-NNS interactions, leading him to propose the concepts of negotiation and interactional modifications, which he suggested occurred when more advanced interlocutors adjusted their speech in order to repair miscommunications and provide learners with more comprehensible input. Long’s original interaction hypothesis proposed that conversational interaction, particularly with native speakers, was the necessary and sufficient condition for SLA (1981) as it provides learners with opportunities to receive input made comprehensible by interactional modifications taking place through the negotiation arising from communication issues.

Initial research focused on describing the variation between native and non-native speaker interactions (e.g. Gass & Varonis, 1985; Long, 1980; Pica, Young, & Doughty, 1987) and how interactional modifications may influence learners’ comprehension. For
instance, Long (1983) identified and described specific interactional modifications used by interlocutors to avoid and repair communication issues, including comprehension checks and clarification requests, as well as repetition, reduced rate of speech, and accommodating unintended changes in topic. Long later proposed an updated version of the interaction hypothesis, emphasizing that through negotiation for meaning, especially negotiation leading to interactional adjustments by a more competent interlocutor, learners may receive feedback on their utterances, directing their attention to mismatches between their interlanguage and target language features and driving them to modify their utterances in an effort to repair communicative misunderstandings. The interactional modifications arising from these negotiations facilitate L2 development by connecting input, internal learner capacities, and output in productive ways (Long, 1996, p. 451-452), highlighting the importance of conversational interaction in SLA.

Although early research demonstrated the potential for negotiation in facilitating SLA, studies focused primarily on the learning conditions, rather than on how these features of negotiation, specifically clarification requests, confirmation checks, and comprehension checks, could lead to language learning outcomes. As Pica (1994) pointed out in her overview article on negotiation, little attention had been devoted to conducting empirical investigations of the direct connection between interaction and L2 development, a necessary step in discovering whether negotiation was in fact successful in facilitating SLA. One of the first studies to address the empirical link between interaction and learning, Ellis, Tanaka, and Yamazaki (1994) investigated the influence on interaction on learners’ comprehension and lexical development within a classroom-based environment. Seeking to establish a direct relationship between negotiation and
learning, the researchers investigated whether learners receiving interactionally modified input demonstrated better comprehension and retention of L2 vocabulary than learners receiving premodified or unmodified input. Findings demonstrated that learners receiving interactionally modified input received more input than those learners receiving premodified or unmodified input, suggesting that negotiation of meaning may lead to increased opportunities for input. In addition, results indicated that the frequency of modifications was not significantly related to learners’ comprehension, suggesting that learning is facilitated by the negotiations and modified input arising from the modifications, rather than the modifications themselves. Gass and Varonis (1994) also demonstrated the potential impact of interaction on learners’ development through their investigation of the effects of interaction on production. NS-NNS dyads assigned to interactive or non-interactive conditions performed two description tasks. Findings indicated modified input from the NS interlocutors was significantly related to task performance, immediate gains in comprehension, and delayed effects on production.

Overall, these two seminal studies demonstrated the potential benefits of interaction for L2 learning, laying the foundation for future investigations on the relationship between conversational interaction and development. Since then, the interaction approach has inspired a vast body of research examining direct and indirect links between interactional features and L2 development, and has been featured in a plethora of handbooks and encyclopedias on second language acquisition and applied linguistics research (e.g. Doughty & Long, 2003; Ellis, 2008; Gass & Mackey, 2012; VanPatten & Williams, 2007).
2.1.2. Key constructs in the interaction approach

Grounded in the notion that communication is the driving force behind language learning, with the research of the past few decades investigating and illuminating the relationship between communication, learning, and mediating factors such as noticing and attention, the interaction approach posits that exposure to linguistic information (as input), opportunities for language production (as output), and positive and negative feedback on learners’ production (through interaction) are constructs critical to understanding second language learning processes (Gass & Mackey, 2006).

2.1.3. Input

Regardless of theoretical orientations, researchers generally agree that in order to successfully acquire a second language, learners must be exposed to the target language. Input, which can be defined as the language that a learner is exposed to through listening, reading, writing, speaking, or other mediums, is a critical aspect of the second language learning process. As Long (1983) points out, access to comprehensible input is a necessary characteristic of language acquisition, with lack of input leading to little or no language development. The concept of comprehensible input has played a key role in SLA, with a variety of theories indicating its importance. However, not all approaches have agreed on whether input alone is sufficient for L2 acquisition. For instance, Krashen (1984, 1994) proposed the Input Hypothesis, which claimed that learners acquire language by understanding input that contains forms slightly beyond their current level of linguistic competence (i + 1). Although input is necessary, in order for acquisition to occur, learners ‘affective filters,’ defined as their levels of anxiety and attitudes towards
the target language, must be sufficiently low to allow the input to be internalized. As summarized by Mackey (2012), although input is a key component of the interaction approach, unlike Krashen’s hypothesis, the interaction approach does not consider input on its own to be sufficient for successful L2 development. Rather, it is the ways in which learners interact with input and their interlocutors through the interactional processes of negotiating for meaning, giving and receiving corrective feedback, and producing modified output, that are likely to lead to L2 development (Mackey, 2012; Mackey et al., 2011).

2.1.4. Output

Although input is crucial to L2 development, researchers have demonstrated that learners also need opportunities to produce language. For instance, contrary to Krashen’s (1984) claim that speaking was the result of acquisition, rather than the cause, suggesting that production did not play an important role in L2 development, Swain (1985, 1995) demonstrated through her work in French immersion contexts that in order for learners to develop productive language skills, output opportunities were necessary. Examining elementary aged children enrolled in a French immersion program, Swain (1985) found that although learners’ receptive skills were native-like, their productive skills were deficient when compared to native French-speaking children of the same age. Arguing that learners’ low proficiency was the result of a lack of opportunity to produce language, Swain suggested that output played a critical role in SLA, particularly comprehensible or pushed output, eventually leading to the formulation of her influential Output Hypothesis (1985) as a complement to Krashen’s Input Hypothesis. This hypothesis suggested that
learners need opportunities to produce output in which they would be ‘pushed’ to make adjustments in order to be understood by their interlocutors. This production not only provides practice, potentially promoting fluency and automatization, but also allows learners to test out hypotheses regarding the target language, directs them to attend to target language forms, and to notice any gaps between their interlanguage and the target language (Swain, 1995).

Although the initial formulation of the Output Hypothesis equated comprehensible output with pushed output, later interpretations expanded the concept of comprehensible output to include all utterances that are understandable to the interlocutor (Van den Branden, 1997). Modified output, or pushed output, which is defined as the reformulation of a learners’ utterance in response to feedback or self-monitoring (Mackey, 2012), is believed to facilitate L2 development by driving learners to modify their production in more target-like manner (Swain, 1985, 1995; Shehadeh, 2002; McDonough, 2005; McDonough & Mackey, 2006). However, as Mackey (2012) points out, the process of modifying output can be as important to language development as the actual modification, with learners still benefiting whether they produce correct target forms or not. Overall, Swain’s Output Hypothesis was integrated into later formulations of the Interaction Hypothesis (Long, 1996), highlighting the importance of output, particularly pushed output, as an integral component of the interaction approach as it provides learners with opportunities to develop fluency and automaticity, test hypotheses about the L2, and receive corrective feedback, thereby drawing their attention to mismatches in their interlanguage and the target language.
2.1.5. Feedback and negotiation

Long’s (1996) update to the Interaction Hypothesis also addressed the ways in which interaction provides negative evidence, defined as input that provides direct or indirect evidence of ungrammatical forms, to the learner, thereby facilitating L2 development. This evidence indicates to learners that there was an issue with their language production, potentially drawing their attention to gaps between their IL and the TL, thereby leading to L2 development.

Negative evidence is an important construct in other theoretical approaches as well. For example, White (1991, 2003) argued that learners would be unlikely to acquire language rules through input alone, as the positive input and evidence they receive does not contain information regarding unacceptable structures in the L2. In order for learners to successfully acquire knowledge regarding grammatical acceptability, White suggests that it would be necessary for them to identify instances in which structures cannot occur through indirect negative evidence. In other words, learners would need to be aware of the lack of a structure in certain contexts, as this provides them with needed information regarding acceptable L2 structures. Although White is approaching acquisition from a more formal perspective, the importance of negative evidence in SLA remains clear. In general, by drawing learners’ attention to their linguistic errors and the gaps between their interlanguage and the target language, negative evidence creates an environment supportive of SLA.

The mechanism through which negative evidence is provided is through interlocutors’ feedback on learners’ L2 utterances (Leeman, 2003, 2007), often during the process of negotiation for meaning. Negotiation can be provided by other learners, native speakers
of the target language, or instructors. Through negotiation, input can be modified to be made more comprehensible, as well as more salient (Mackey, 2012). In addition, negotiation may prime learners to be more attentive to future input, raising their awareness of specific features of the target language and providing them with multiple opportunities to confirm or disconfirm hypotheses they have formed regarding the L2. Negotiation was originally operationalized to include confirmation checks, clarification request, and comprehension checks (Long, 1983). Confirmation checks are utterances designed to “elicit confirmation that the utterance has been correctly heard or understood by the speaker” (Long, 1983, p. 137), while clarification requests are “expressions designed to elicit clarification of an interlocutor’s preceding utterance” (Long, 1983, p. 137). Comprehension checks are designed to confirm that an interlocutor has been understood by attempting “to anticipate and prevent a breakdown in communication (Long, 1983, p. 136).

Although negotiation has been traditionally operationalized and analyzed according to these definitions, this important concept has expanded in recent research to include interactional modifications occurring in response to other forms of implicit and explicit feedback, including recasts and metalinguistic feedback, as well as modified output (Mackey, 2012). Within this expanded operationalization, recasts, defined as the reformulation of all or part of a learner’s immediately preceding utterance in which one or more non-target like item(s) are replaced by the corresponding target form, and where the focus is on meaning rather than form or object (Long, 2007), have been shown to represent positive as well as negative evidence (Mackey, 2012). Recasts have been shown to be facilitative of L2 development as they enhance the salience of the target feature and
direct learners to contrast their erroneous utterance with their interlocutors’ reformulation (Goo & Mackey, 2013; Doughty & Varela, 1998; Leeman, 2003; Long, 1996, 2007), thereby focusing learners’ attentional resources on the target form. However, research has also demonstrated that the impact of recasts on learning may be affected by a variety of factors, such as target feature (Egi, 2007), learners’ proficiency and developmental readiness (Mackey & Philp, 1998), setting (Oliver, 2000), task characteristics (Révész, Sachs, & Mackey, 2011), and learners’ individual differences (Mackey, 2012). Although some studies have found that recasts may be unlikely to lead to acquisition (e.g. Lyster, 2004; Lyster & Ranta, 1997) research has empirically demonstrated the effectiveness of recasts on learners’ L2 development (see Mackey & Goo, 2007 for a review). Overall, research has shown that receiving feedback and participating in negotiation may support learners’ L2 development by providing both positive and negative evidence, which as Gass and Mackey (2006) point out, can aid learners in noticing their erroneous utterances and focusing their attention on the target language, thereby preparing learners to be more observant regarding future instances of linguistic input and the testing of their linguistic hypotheses.

2.1.6. Interaction and computer-mediated communication

During the past few decades, CALL research has sought to situate itself more firmly within the theoretical foundations of a variety of SLA approaches, with the interactionist perspective serving as the predominant approach in current SLA-oriented CALL research (Chapelle, 2005). In a series of articles, Chapelle suggested that CALL research not only needed to be more firmly grounded within SLA theory, but that there was a need for
researchers to inform their studies using research methods from instructed SLA (Chapelle, 1997) and to design more pedagogically effective computer-based activities (Chapelle, 1998). Reacting to early CALL research, which suffered from the same limitations as early interaction research, specifically a lack of detail regarding the description of learner interaction during instruction (Long, 1980), Chapelle (1990) pointed out that much of the initial investigations failed to empirically describe learners’ processes and experiences during CALL activities, instead focusing on the description of the computer activities used in the language classroom. According to Chapelle (1998), the most critical questions which CALL research could address were 1) what kind of language does the learner engage in during a CALL activity, and 2) how good is the language experience in CALL for L2 learning. Although the first question focuses on the description of language that learners receive and produce, the second question is evaluative in nature. Chapelle (1997) suggested that in order to effectively evaluate the quality of learners’ language in L2 tasks, it is necessary to function according to certain assumptions about the types of language facilitative of L2 development. By grounding CALL-based investigations in established SLA perspectives, such as that of the interaction approach, researchers would have specific observable features of learners’ interlanguage on which to focus their investigations. In other words, Chapelle’s articles highlighted the idea that learner language is central to investigations of L2 development, an area of inquiry that was being pursued in the interaction literature at the time.

However, within the body of CALL research in the early 1990s, few studies had sought to examine the nature of learners’ interactions. For example, in her investigation of computer-assisted class discussion, Chun (1994) examined the quantity and quality of
language produced by German as a foreign language learners. Using transcripts of learners’ synchronous computer-mediated communication on a LAN, Chun examined the number and length of turns, grammatical complexity, and type and amount of different discourse structures used by learners in their text-based interactions. Results indicated that learners exhibited features of interaction, including the provision of corrective feedback and the use of negotiation for meaning in the form of clarification requests.

Chun’s study was one of the first investigations of learner language in a computer-mediated environment, with findings demonstrating that the same features found to occur in face-to-face interaction also took place in SCMC. This research, although descriptive rather than experimental in nature, provided important evidence that learners’ communicative competence could benefit from computer-mediated interaction, even though students were interacting in a written rather than oral mode. By demonstrating the existence of interactional features in text-based chat, Chun’s research also showed the potential for future CALL investigations from an interactionist perspective. As Chapelle (1997) points out, oral interactions comprise the discourse intended for language learning (p.23), and Chun’s findings indicate that this same language occurs in written conversations between learners in computer-mediated communicative contexts. In addition, Chun’s research demonstrates the occurrence of negotiation for meaning in response to misunderstandings. By encouraging learners to focus on the meaning of the interaction and the dynamics of the discourse, rather than a specific goal-oriented task, Chun provided learners with opportunities to focus on their language without sacrificing the communicative value. Focus on meaning is central to the interactionist approach (Long, 1996), and communicative tasks, although not widely used during early CALL
research, are critical to exploiting the language learning potential of a computer-mediated context.

Urging researchers to not only use CALL to examine interactional processes, but to also investigate the effectiveness of interaction for L2 development, Chapelle (1997) stressed the importance of conducting CALL research situated within SLA, rather than in psychology, computational linguistics, or educational technology. Specifically, Chapelle addressed the need to use hypotheses from interaction research as guidelines for investigations into learner language and development within CALL contexts. By grounding CALL based research in the interaction paradigm, researchers would have a framework within which to examine specific aspects of interaction thought to be facilitative of SLA, such as comprehensible input, modified output, and negotiation of meaning. Overall, Chapelle called for researchers to design and use empirical and pedagogical activities that would direct learners’ attention to form, as well as encourage modified output in response to negotiation moves. Conducting CALL research from the interactionist perspective would thus provide researchers with information useful for improving the design and evaluation involved in computer-mediated tasks. Crucially, studies that answered this call not only provided empirical investigations testing the potential of computer-mediated contexts to facilitate SLA, but also an additional context in which to test SLA hypotheses already supported in face-to-face classroom environments. Through the use of such complementary methodologies, CALL research had the potential to make valuable contributions to researchers’ understanding of the processes and experiences involved L2 language learning.
Building on the idea that CALL might be developed to reflect ideal conditions for L2 development and focusing on the potential pedagogical applications of interaction in CALL, Chapelle (1998) identified seven features of interaction that could be directly applied to instruction in a computer-mediated environment, including making key linguistic features salient, offering modified input and supporting modified interaction between the learner and the interlocutor, and providing opportunities for learners to be active participants in the task and to notice their errors, to modify their output, and receive comprehensible output. These principles of instructional design were based on hypotheses regarding ideal SLA conditions (Long, 1996; Pica, 1994), and were later expanded upon by Doughty and Long (2003) to include features of task-based language learning and teaching, providing CALL researchers with theoretically sound foundations on which to base hypotheses regarding L2 development in computer-mediated contexts. Using these instructional designs to investigate the basic tenets and features of the interaction approach in CALL contexts, researchers sought to establish the ability of computer mediated communication (CMC) to direct learners’ attention to specific target language features and to provide learners with opportunities to negotiate for meaning, to receive comprehensible input and corrective feedback, and to produce modified output (Smith, 2004). Following calls from researchers to approach CALL SLA from a more theoretically grounded perspective (Chapelle, 1994, 1997, 1998, 2001), researchers began to examine interactional features within CMC. Similar to FTF interaction research, the initial investigations in CALL sought to describe interactional features, focusing on observational studies of discourse in CMC.
For instance, Pellettieri (2000) examined intermediate level Spanish learners’ negotiation of form and meaning during different types of communication tasks, including jigsaw and guided conversation tasks, during text chat. Findings indicated that the patterns of interaction found in CMC, referred to in her study as network-based communication (NBC), were similar to those found in NNS oral conversation, with instances of negotiation triggered by inappropriate responses or lack of comprehension. Pellettieri also found that learners used clarification requests and confirmation checks to negotiate for meaning, and provided one another with corrective feedback during meaning-focused exchanges. These results led Pellettieri to conclude that CMC provided an environment facilitative of negotiation of meaning and interaction, and importantly, that because learners in a CMC context may have more time to process and monitor their language, CMC may play a “significant role in the development of grammatical competence among classroom language learners.” (Pellettieri, 2000, p.59). Pellettieri also suggested that synchronous text-chat may be more beneficial to learners’ L2 development than FTF contexts as it provides learners with a visual representation of their own output, thereby directing learners’ attention to their language and allowing students to “practice and gain control over more cognitively demanding aspects of grammar that otherwise might not be so frequently practiced in classroom oral interaction” (p.82). Providing further support for Chapelle’s (1998) instructional designs, Pellettieri also called for CMC tasks to be designed to promote learners’ need to exchange information, providing multiple opportunities for interaction, and subsequently, L2 development.

Blake (2000) also used communicative text-chat tasks within a computer-mediated environment in order to encourage negotiation between fifty intermediate learners of
Spanish. Participants completed two-way information gap tasks, in which each partner holds half of the information needed for task completion, thereby requiring learners to share and converge on a single outcomes, one-way information-gap tasks, in which only one partner is provided with the relevant information and the other must elicit the data needed for task completion, and decision-making tasks, in which all interlocutors have access to the information needed for task resolution, although they are not required to converge on one single outcome. One-way information gap tasks were identified as the most effective at promoting negotiation. In addition, in results similar to those by Pellettieri (2000), Blake (2000) found that these negotiations were most often triggered by lexical items, with 75% and 95% of all negotiations occurring in response to lexical misunderstandings. Results also indicated that learners increased their metalinguistic awareness of their own L2 vocabulary development, suggesting that the benefits of interaction in CMC have communicative as well as potentially developmental benefits.

Similarly, Smith (2003) also found that task type plays an influencing role in the ways that learners negotiate for meaning in CMC. Examining fourteen learner dyads, Smith examined the amount and types of negotiation occurring when learners encountered novel lexical items during jigsaw and decision making tasks. Providing additional evidence supporting Pellettieri (2000), results demonstrate that learners do negotiate for meaning when miscomprehension occurs in a computer-mediated environment. In addition, findings provide some evidence that negotiation is affected by type of task, with more negotiation occurring in the decision making task than the jigsaw task. Although the negotiation patterns identified in this study were similar to those found in FTF environments, Smith argues for an expanded model of negotiation specific to
CMC. In contrast to oral contexts, Smith suggests that a CMC model of negotiation must account for a delay between the initial trigger and the following negotiation, as the absence of adjacency turns in text-based chat often leads to extended delays in learners’ responses to negotiation triggers. Importantly, this study not only answered Chapelle’s calls by providing detailed, descriptive evidence of what occurs during learner interactions in a computer-mediated context, but it sought to move the field forward by proposing an expanded model of negotiation specific to a computer-mediated context. This model took into consideration the features unique to CMC, providing a more accurate and representative method for coding and assessing learner negotiation in text-based chat.

2.1.7. Learner characteristics and their role in interaction

Turning to the impact the interlocutor may have on negotiation, Toyoda and Harrison (2002) examined discourse moves between Japanese NS and NNS, investigating whether interactional adjustments by a NS or a more competent interlocutor lead to negotiation in CMC in a manner similar to face-to-face interaction. Rather than using focused tasks, the dyads engaged in open-ended chat conversation. Using a discourse analytic approach, the authors identified nine categories of negotiation of meaning triggered by difficulties in understanding, including recognition of a new word, misuse or misunderstanding of a word, typing error, grammatical error, inappropriate segmentation, abbreviated sentencing, sudden topic change, slow response, and intercultural communication gap. In addition, many of these negotiations led learners to produce modified output by both NS and NNS interlocutors. Results indicate that many of the same interactional features
found in FTF interaction also occur in CMC, and similar to those results of Pellettieri (2000), Toyoda and Harrison suggest that by providing learners with the opportunity to visually review their interaction through chat logs, their attention will be directed to their interlanguage, thereby potentially leading to improvement (2002, p. 96).

Tudini (2003) also investigated learner and native speaker dyads, examining whether live chat between NNS and NS of Italian provided negotiation opportunities during open-ended tasks. Results indicated that negotiation for meaning and modified output opportunities do occur in synchronous CMC, with interlocutors also providing instances of corrective feedback during the interactions. Findings also demonstrated that lexical and morphosyntactic difficulties triggered the most instances of negotiation, supporting previous research (Blake, 2000; Pellettieri, 2000; Smith, 2003) and providing evidence that CMC chat may support L2 development in ways similar to oral interaction. Kötter’s (2003) research also drew on interactionist approaches, developing categories for examining learner data produced in text-based multiuser object-oriented (MOO) collaborations. In contrast to other studies (e.g. Blake, 2000; Pellettieri, 2000), the tasks in Kötter were not designed specifically to foster negotiation. Nonetheless, results showed that learners engaged in frequent comprehension checks and clarification requests, and provided each with other with positive feedback to indicate understanding and agreement. Kötter’s results reveal that learners in the MOO environment produce similar interactional moves as those found in FTF interaction, providing further evidence that CMC can contribute to SLA.

Overall, these studies answered Chapelle’s (1994, 1997, 1998) call to provide empirical descriptions of learner interactions in computer-mediated environments,
demonstrating that not only does negotiation occur in CMC in patterns similar to FTF interactions, but that the nature of CMC, particularly synchronous text-chat, may provide additional benefits of interaction. For instance, because learners are presented with a written record of their interaction, they may be directed to attend more closely to the form and content of the input, while still maintaining the real-time feel of conversation (Pellettieri, 2000; Smith, 2003; Toyoda & Harrison, 2002). This additional processing time may provide learners with more opportunities to receive modified input, and produce modified output and corrective feedback, directing their attention to gaps between their interlanguage and the target language.

However, the focus of these discourse-oriented studies was to identify and describe instances of negotiation in CMC, with some researchers using indirect comparisons to oral interaction research to make claims regarding the comparable effectiveness of CMC in supporting interaction. For instance, Kötter (2003) compares the amount of learner generated clarifications, elaborations, and reformulations within the MOO to the those performed by learners in Long and Porter’s (1985) and Pica and Doughty’s (1985) face-to-face oral interactions. Although the raw amounts and ratios of negotiation seem to indicate the MOO may be more supportive, Kötter’s conclusions must be interpreted with caution due to the differences in tasks used within the studies. Within the MOO, learners were not required to participate in controlled, structured tasks, but were encouraged to converse freely. In other words, participants completed non-goal directed two-way opinion and information exchange tasks. Pica and Doughty’s participants, on the other hand, completed one-way decision-making tasks, which may be more likely to lead to less negotiation than two-way tasks (Pica, Kanagy, & Falodun, 1993). As Pica et al.
(1993) suggest, and as research within CMC demonstrates (e.g. Blake, 2000; Smith, 2003), type of task has been found to significantly impact the amount of negotiation, suggesting that in their haste to demonstrate the comparability of CMC to FTF interaction, researchers may have inferred greater conclusions than the available data was able to support.

2.1.8. Comparing interaction in FTF and CMC

Investigations have also sought to make direct comparisons between interaction in FTF and CMC, demonstrating the existence of interactional similarities between the two modes. For example, research has demonstrated that both contexts provide learners with opportunities to negotiate for meaning (Fernández-García & Martinez-Arbeiz, 2002) and provide an environment facilitative for corrective feedback, including recasts (Lai & Zhao, 2006). Although research investigating interactional features and processes within CMC expanded to include direct comparisons between computer and oral contexts, most of these studies were descriptive in nature and did not seek to provide direct connections between CMC, interaction, and SLA. Answering the call to address evaluative methods and learning outcomes (Chapelle, 1997, 1998), CALL research sought to move beyond examining whether features of interaction can and do occur in CMC to investigating whether interaction in CMC directly benefits L2 development.

For instance, de la Fuente (2003) investigated whether interaction occurring in CMC was as effective as FTF interaction in promoting receptive and productive lexical knowledge. Participant dyads completed an information gap task designed to foster negotiation and modified output. Using a pretest-posttest-delayed posttest design, results
indicated that learners in both the FTF and CMC groups demonstrated receptive and productive gains in L2 vocabulary development. Although not statistically significant, the oral interaction group (FTF) outperformed the CMC group on both written and oral measurements, and the FTF group demonstrated significantly higher productive skills on both immediate and delayed posttests while the CMC group did not. These results suggest that, contrary to previous claims that CMC directs learners’ attention to form by providing greater salience than oral interaction, FTF may be more beneficial for immediate oral productive acquisition. However, de la Fuente (2003) imposed strict limits on the time permitted for negotiation, possibly inhibiting learners from taking part in the negotiations necessary to successfully notice and acquire the target forms. As previous research has suggested that CMC negotiations are likely to experience potentially extended delays between triggers and indicators (Smith, 2003), these temporal restrictions may have impacted the results. For instance, there is a short time delay in text-chat between the initiation of the utterance and its receipt by the interlocutor. More importantly, SCMC does not adhere to the same pattern of turn adjacency found in FTF interaction. Rather, negotiation triggers may be followed by an indicator, but are initially unaddressed, only to be answered later after a repeat indicator in subsequent turns. These delays between triggers, indicators, and responses is referred to as a split negotiation routine (Smith, 2003), and highlight the possibility that negotiation may be hampered by time restrictions due to the complex nature of split negotiation routines in SCMC.

Comparative studies between CMC and face-to-face environments have also provided encouraging evidence for the use of technology in the classroom, with findings indicating higher oral proficiency scores for learners participating in online and face-to-face
interactions when compared to learners participating in only face-to-face contexts (Payne & Whitney, 2002). Recent research by Murphy (2011) has demonstrated the importance of corrective feedback provided during social interaction in asynchronous and synchronous chats, with results suggesting that feedback contributed to learners’ improved confidence and self-evaluation, while the social aspects served as a source of motivation for students, thereby showing the positive impact of CMC on affective factors.

In addition, studies have shown that text-based CMC provides extended opportunities for learners to notice target language features (Warschauer & Kern, 2000), with some findings suggesting that the use of a CALL context may result in heightened and sustained student interest through greater interaction with a variety of materials and tasks (Strambi, 2001). According to Smith (2004) and Chapelle (2001), text-based CMC might provide learners with additional saliency as well as the opportunity to re-read and review previous text, thereby allowing learners to more closely focus on target forms while not impacting the communicative success of the interaction. In addition, the added time available to learners may provide more processing time, potentially leading to improved comprehension, more accurate production, and a higher quality of interaction than might have occurred in a face-to-face context (Pellettieri, 1999). Recent research in CMC has also examined the effect of specific types of technology on the quantity and type of negotiation learners produce, with studies comparing the amount of negotiated repair moves in voice and text chat (Jepson, 2005) and the relationship between negotiated interaction and the use of scrolling or cursor movement (Smith, 2009). Smith (2010) has also investigated the relationship between learners’ eye movements and uptake, while
Wang (2006) examined the role of visual cues, such as facial expressions or gestures, in negotiation routines during video conferences. However, studies have also found results demonstrating improved performance in FTF interaction over computer-mediated environments. For example, Baralt (2010) examined the effects of increased task complexity in FTF and CMC interactions. Results indicated that increases in cognitive complexity led to increased levels of processing and awareness, which were significant predictors of L2 development within the context of the study, suggesting that complex tasks may not be as facilitative of L2 learning in CMC as in FTF contexts.

Overall, the continued integration of theory and methods from the interactionist perspective during the last few decades has strengthened the theoretical foundations of CALL research, providing empirical results on which the field can continue to build. However, although individual research reports have provided promising results regarding the benefits of interaction in CMC, results are difficult to generalize across studies due to variation in setting, design, and participant population. Considering the lack of conclusive, generalizable findings, there is a clear need for a statistically sound review of the effects of interaction on learners’ L2 development within computer-mediated environments.

2.2. Research Questions

This dissertation provides a synthesis and meta-analysis of the recent trends and findings on the role of interaction in second language development in CALL, as well as a critical evaluation of the methodological quality of CALL-based interaction research during the past few decades. Specifically, the current research focuses on CALL research
from 1990 to 2012 and investigates the relationship between interaction, computer-mediated contexts, and L2 development by addressing the following research questions:

1. Compared to face-to-face interaction, how effective is synchronous computer-mediated interaction at promoting L2 learning outcomes?

2. Compared to face-to-face interaction, how effective is synchronous computer-mediated interaction at promoting receptive and productive L2 learning outcomes?

3. Compared to face-to-face interaction, how effective is synchronous computer-mediated interaction at promoting oral and written L2 learning outcomes?

4. To what extent do the following factors impact the degree to which interaction facilitates L2 development in computer-mediated environments?
   a. Context (foreign language and second language)
   b. Setting (laboratory and classroom)
   c. Interlocutor characteristics (learner, native speaker, instructor, other)
   d. Type of target feature (syntactic, morphological, phonological, lexical, pragmatic, productive)
   e. SCMC mode of interaction (text-based chat, video chat, audio conferencing, multimedia, or bimodal chat)
   f. Type of interactional feedback provided during the interaction (recasts, comprehension checks, clarification requests, confirmation checks, prompts, metalinguistic feedback, elicitation, repetition)
   g. Opportunity to produce modified output (encouraged, discouraged, not controlled)
5. To what extent have research practices and statistical procedures been reported thoroughly?

Because one of the goals of this dissertation is to provide a critical assessment of the methodological quality in CALL-based interaction research during the last two decades as well as provide a quantitative assessment of the main effects and relationships between interaction, L2 development, and CMC, the first four questions focus on study-generated evidence, whereas the fifth question combines study-generated and synthesis-generated evidence (Cooper, 1998). Through the quantitative coding of studies’ methodological features, including study design, the reporting of statistics, and validity, the current research examines the progression of study quality in CALL-based interaction research over the past few decades. In addition, the last research question yields important information regarding how interaction research has been conducted in SCMC, as well as the ways in which individual variables, such as research context and setting, might interact with one another. Furthermore, the current study systematically examines the progression of research methodologies, including positive changes in research practices and reporting. This approach not only offers a comprehensive look at interaction in CALL environments, but suggests possible research agendas for future CALL and SLA investigations.
Chapter III: Study Design and Methodology

3.1. Study Identification and Retrieval

According to Lipsey and Wilson (2001), the most effective strategy in locating relevant studies for possible inclusion is to use multiple sources, including review articles, references in published studies, bibliographic databases, bibliographic reference volumes, relevant journals, conference programs and proceedings, dissertation and theses databases, and experts in the target area. Because it is unlikely that any one source will allow for the identification of all potentially relevant studies, the use of multiple and varied sources is not only necessary, but imperative to the success of synthetic research.

Following previous research (Plonsky & Gass, 2011; Smith & Lafford, 2009; VanPatten & Williams, 2007), journals were identified as comprising the primary means through which CALL and SLA research is disseminated to the field at large. Journals are also often readily available through both print and electronic sources, ensuring access to the relevant research reports. Primary studies published as books or within edited volumes were excluded based on the premise that the available number of sources would prevent the retrieval of a representative sample. For example, certain volumes or collections were difficult to procure through library access, potentially biasing the sample due to availability.

In order to retrieve the relevant journal reports for the current synthesis, a number of sources were consulted. The initial stage of data retrieval involved a comprehensive and exhaustive search using the online research databases Linguistics and Language Behavior Abstracts (LLBA), Education Resources Information Clearinghouse (ERIC), and
Education Full Text to identify potentially eligible empirical studies and associated literature, such as previous reviews or syntheses. In addition, the online databases ProQuest and Dissertation Abstracts International were searched to retrieve relevant unpublished research reports.

Previous syntheses and meta-analyses (Keck et al., 2006; Mackey & Goo, 2007; Norris & Ortega, 2000) have elected to exclude unpublished research for a variety of reasons, including convenience, to avoid a biased sample, and to represent a higher quality of research. By excluding this “fugitive literature,” these previous analyses have a larger potential for publication bias, as a greater number of studies reporting statistically significant results are usually accepted for publication than studies with non-significant findings. In addition, published studies are more likely to have larger effect sizes than unpublished studies (Begg, 1994; Lipsey & Wilson, 1993; Smith, 1980), potentially resulting in an upward bias in the final analyses’ accumulated effect sizes (Lipsey & Wilson, 2001). To avoid these possible methodological issues, many meta-analysts have encouraged the inclusion of unpublished work in order to have a sample more representative of a particular research domain and to avoid introducing any bias into the size of effects that might be found (Hunter & Schmidt, 2004; Konstantopoulos & Hedges, 2004; Lipsey & Wilson, 2001; Norris & Ortega, 2006; Rosenthal, 1994). In order to obtain a larger sample of relevant research, dissertations were selected for inclusion in the current study, as they have been subjected to review by experts in the field, and though unpublished, are carefully designed and might be considered comparable quality to published work. In addition, dissertations are likely to include detailed information regarding methodology and statistical analyses, providing the necessary information for
the calculation of effect sizes. Rather than excluding these studies solely due to publication status, dissertations were retrieved, evaluated, and included in the sample based on the eligibility criteria discussed below. Using the same key words used in the search of published information, the electronic database ProQuest Dissertations and Theses was searched for relevant studies.

The next phase of retrieval was conducted using the online search function for targeted L2, applied linguistics, and CALL publications. Five CALL-specific journals (CALICO Journal, CALL-EJ Online, Computer Assisted Language Learning, ReCALL, Language Learning and Technology) and fourteen applied linguistics journals (Applied Linguistics, Journal of Applied Linguistics, Canadian Modern Language Review, Foreign Language Annals, International Journal of Educational Research, Language Learning, Language Testing, International Review of Applied Linguistics, Language Teaching Research, The Modern Language Journal, Second Language Research, Studies in Second Language Acquisition, System, TESOL Quarterly) were targeted for relevant studies and literature. These journals are some of the most commonly referenced journals in the fields of SLA and applied linguistics, with sixteen having been identified by Smith and Lafford (2009) as relevant and beneficial for tenure and promotion by CALL and applied linguistics professionals.

A key word search was used to locate potentially relevant research in the above sources. Within the selected databases, an anywhere search, which allows for the identification of target studies including the search term in the title, keyword, and abstract subfields, was first used to search for combinations of the following key words and terms: computer-mediated communication, computer assisted language learning,
synchronous chat, asynchronous chat, multimedia, interaction, input, output, feedback, corrective feedback, recast, uptake, negotiation, tasks, task-based language learning, task-based instruction, and language learning technology.

Using studies identified in the first two phases, the third phase of study retrieval utilized forward searches and citation chaining through the online search engine Google Scholar. Reference sections of CALL and interaction related book chapters were also reviewed to locate all relevant research, and experts in the field were consulted regarding potentially eligible studies and reports. At this time, only English language research reports were included in the sample. However, it is hoped that a future analysis might also include relevant research from publications in other languages in order to provide a more complete synthesis of the current research conducted in the global fields of SLA and CALL. Overall, this multiple source approach is designed to yield a large and representative body of research examining the relationship between interaction and computer-mediated environments, thus providing an adequate sample size for synthesis and analysis.

Regarding temporal parameters, studies published between 1990-2012 (inclusive) were considered for inclusion. These two decades were selected as they roughly correspond to the initial investigations of interaction within a CALL context, and provide sufficient data to contrast changing trends and methodological practices across the 1990s and initial decade of the 2000s.

3.2. Inclusion and Exclusion Criteria
543 studies relating to interaction and synchronous computer-mediated communication were identified through the retrieval process and were reviewed in order to determine relevance to the proposed research questions. For the quantitative meta-analytic aspect of the present dissertation, studies were reviewed using the following inclusion and exclusion. However, as Lipsey and Wilson (2001) point out, eligibility criteria often undergo revision once coding is underway. For example, studies whose eligibility may have been ambiguous under the initial inclusion criteria might be located, warranting further revision. Thus, finalizing the eligibility criteria was an iterative process, and the inclusion/exclusion criteria evolved during the coding phase of the project. However, as criteria changed, all new criteria were applied retroactively to any studies coded prior to the change. The final inclusion and exclusion criteria are included below.

3.2.1. Inclusion criteria

1. The study was published in a peer reviewed journal between 1990 and 2012. This temporal range was selected in order to capture the initial investigations as well as the most recent research on interaction in SCMC. This range was also thought to provide sufficient data to contrast changing methodological practices across the 1990s and initial decade of the 2000s.

2. The study was a dissertation completed between 1990 and 2012. This range was selected for the same reasons described above.

3. The study examined a target language that was either a second or foreign language for the participants.
4. The study compared interaction in SCMC to interaction in FTF, with both groups participating in interactional activities. In other words, the study measured differences in L2 learning outcomes between a group participating in interaction in SCMC (operationalized as the treatment group) and a group participating in interaction in FTF (operationalized as the comparison group).

5. The study incorporated some form of synchronous computer-mediated communication (e.g. text-chat, video conferencing, audio conferencing, or bimodal chat).

6. The study was of an experimental or quasi-experimental design. Because this study was concerned with the comparative efficacy of interaction in two groups, FTF and SCMC, only studies of an empirical nature were included. Studies must have followed an independent between-groups design, in which participants were assigned to only one mode of communication. Designs could include comparison or ‘control’ groups where the ‘control’ participated in the same activities as the treatment group but in a FTF communication mode.

7. The study incorporated one or more open or closed interactive tasks (one-way or two-way information gap, information exchange, problem-solving, and decision-making tasks) or an interactive setting, such as conversation groups, that facilitated interaction through open group discussion. Tasks or group discussion were permitted to serve as the target treatment (in studies comparing more than one task type) or as an environment for the investigation of other features (e.g. corrective feedback).
8. The study examined the effects of SCMC on L2 learning outcomes, including measures of production, performance, and development. Outcome measures examining oral, written, receptive, and productive skills were included, as long as procedures and assessments were designed to evaluate learners’ linguistic, communicative, or socio-pragmatic knowledge and understanding.

9. The study measured L2 learning outcomes using a pretest-posttest or posttest only design. In other words, the study assessed and compared the impact of the interactional treatment and communication mode using custom made or standardized measures (such as ACTFL’s Oral Proficiency Interview). Although the inclusion of studies using custom-made tests may be criticized due to the difficulty inherent in comparing such measures across a range of participants, settings, and contexts, these types of tests are relatively common and are often used in studies appearing in peer-reviewed journals. Because these

10. The study included the necessary information for the calculation of effect sizes, including mean, standard deviation or variance, and sample size, for each group.

3.2.2. Exclusion criteria

1. The study incorporated exclusively asynchronous CMC (ACMC) and FTF interaction.

2. The study incorporated SCMC in combination with ACMC, but findings were not attributed to the synchronous aspects of the research.
3. The study compared interaction in SCMC and in FTF, but the FTF did not participate in any interactional activities. In other words, because the current research was concerned with the comparative efficacy of interaction, studies in which the FTF group only completed the testing measures were excluded.

4. The study compared interaction in SCMC and in FTF, but followed a within-groups repeated measures design, in which all learners participated in both conditions.

5. The study did not follow a quasi-experimental or experimental design. Instead, the study was observational, descriptive, or correlational in design, which although important to our understanding of interaction in SCMC, do not provide direct, data-based comparisons of L2 learning outcomes. In addition, these types of designs seek to answer different research questions, which would make any conclusions drawn from their summary effect sizes unreliable (Borenstein et al., 2009).

6. The study did not examine L2 production, performance, or development. For instance, the study may have examined learners’ perceptions of CMC or the paper may have focused on description or theoretical explication. Although valuable to the field as a whole, studies using only self-report measures or questionnaire data on learners’ opinions or perceptions of the efficacy of interaction in SCMC were excluded on the basis that they did not empirically examine the effects of interaction on L2 production, performance, or development.
7. The study did not examine L2 learners as the target population, but rather focused on the use of CMC as a pedagogical training tool. In these studies, learners were secondary to the instructor-centered research questions and outcomes. Although this is an important area of SCMC research, educators’ perceptions and experiences are the focus of the current study.

8. The study was a literature review, synthesis, or meta-analysis.

9. The study did not include the information necessary to calculate effect sizes.

10. Studies based on the same sample of participants but appearing in more than one journal or book were included only once as a unique study within the current sample in order to avoid the artificial inflation of sample sizes.

3.3. Analysis

3.3.1. Coding procedures

Fourteen unique sample studies satisfying the inclusion criteria were retrieved and identified for the current study. The author coded all of the studies according to the substantive and methodological features listed in Table 1, with a second and third rater coding 75% of the sample to ensure reliability. Interrater reliability was .91, and was calculated using simple percentage agreement.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition/operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Author(s) of each study report</td>
</tr>
<tr>
<td>Type of publication</td>
<td>Dissertation or peer-reviewed journal</td>
</tr>
<tr>
<td>Title of publication venue</td>
<td>Journal in which they study was published</td>
</tr>
<tr>
<td>Year of publication</td>
<td>Year in which the study was published</td>
</tr>
<tr>
<td>Learner characteristics</td>
<td></td>
</tr>
<tr>
<td>Participants’ L1</td>
<td>First language(s) of the participants</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Participants’ age</td>
<td>Mean age or age range of the participants</td>
</tr>
<tr>
<td>Proficiency level</td>
<td>Participants’ proficiency level (beginning, intermediate, advanced)</td>
</tr>
<tr>
<td>Proficiency measure</td>
<td>How proficiency was measured to classify participants (in-house placement, institutional designation, standardized test, self-report)</td>
</tr>
<tr>
<td>Participant L2 (target language)</td>
<td>The language in which learning outcomes were conducted</td>
</tr>
<tr>
<td>Academic status</td>
<td>Educational level of participants (secondary, high school, university)</td>
</tr>
</tbody>
</table>

### Research characteristics

<table>
<thead>
<tr>
<th>Target feature</th>
<th>Type of linguistic, communicative or socio-pragmatic feature that was the developmental or performance target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactional feedback</td>
<td>Type of general or specific feedback provided during the interaction (recast, clarification request, comprehension check, confirmation check, metalinguistic feedback, elicitation, repetition).</td>
</tr>
<tr>
<td>Modified output</td>
<td>Whether the study controlled learners’ opportunities to produce modified output, and whether opportunities were provided.</td>
</tr>
<tr>
<td>Task type</td>
<td>Type of interactional task or activity used for the interactional treatment (one-way information gap, two-way information gap, information exchange, problem-solving, decision-making).</td>
</tr>
<tr>
<td>Type of technology</td>
<td>The type of computer-mediated technology used during the interactional treatment (text, audio, video, multi-modal).</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Characteristics of interlocutor (native speaker, non-native speaker, researcher, teacher, or learner).</td>
</tr>
<tr>
<td>Research setting</td>
<td>Setting in which the study was conducted (second language or foreign language).</td>
</tr>
<tr>
<td>Research context</td>
<td>Context in which the study was conducted (laboratory or classroom).</td>
</tr>
<tr>
<td>Research design</td>
<td>Type of research design used by the study (observational, experimental, pretest-posttest, between groups, within groups).</td>
</tr>
<tr>
<td>Group design</td>
<td>Whether the study used a comparison or control group design.</td>
</tr>
</tbody>
</table>

### Statistical reporting

<table>
<thead>
<tr>
<th>Sample size</th>
<th>The sample size of the participants (total, per independent group).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>The mean of each outcome measure for each independent group.</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>Whether or not the study reported the standard deviation of the mean for each outcome measure for each independent group.</td>
</tr>
<tr>
<td>Statistical test</td>
<td>Type of inferential or non-parametric statistical.</td>
</tr>
</tbody>
</table>
### Effect size
Whether or not the study reported an effect size for inferential statistics.

### Confidence interval
Whether or not the study reported confidence intervals for inferential statistics.

### P values
Whether or not the study reported an exact value or a relative value.

### A priori alpha
Whether or not an a priori level of alpha was reported.

### Power analysis
Whether or not the results of a post-hoc or a priori power analysis were reported.

### Reliability-rater
Whether or not any measures of interrater reliability were reported.

### Reliability-instrument
Whether or not any measures of reliability were reported for materials (including Cohen’s kappa or alpha).

### Study filtering process

The retrieval process initially identified 543 studies for potential inclusion in the analysis. First, studies that did not examine language learners or did not involve SCMC or interaction were eliminated from the pool. Next, studies using only a qualitative or case-study approach were discarded from the sample. Studies were then categorized according to study design, with those using a one-shot or one-group design eliminated from the sample size. Finally, of the remaining studies using a two-groups pretest/posttest design, only those using an independent group design were included in the final sample. These fourteen studies met the eligibility criteria, and were coded and classified according to the selected substantive and methodological features described above. As is typical of synthetic research (Lipsey & Wilson, 2001), substantive features were examined as independent variables to determine the effects of study characteristics on the nature and scale of the primary studies’ findings. In addition, as one of the goals of this synthesis is to identify potential areas of improvement in research methodology, methodological features, such as number of treatments, duration of treatment per session,
statistical tests, and reported effect sizes, were also examined in order to determine the impact of methodology on the studies’ outcomes. To ensure that the investigated features were relevant to the selected research domains, experts in interaction and CALL reviewed the proposed features and definitions, and their recommendations were subsequently integrated into the coding book (Stock, 1994) used for the systematic review of the retrieved reports.

3.3.3. Substantive features.

Because one goal of the present dissertation is to investigate whether well-established benefits of interaction on L2 development can be found in computer-mediated contexts, the current study investigated the effects of five substantive features related to interaction and CALL: type of target feature (syntactic, morphological, phonological, lexical, pragmatic), type of interactional feedback (recast, clarification request, comprehension check, confirmation check, metalinguistic feedback, elicitation, repetition), modified output opportunities (participants were or were not provided with opportunities to modify their language production or it was not controlled), type of technology (text-chat, video-chat, multimedia, audio conferencing, bimodal chat), and type of interlocutor (learner, native speaker, instructor, other).

Building on previous meta-analytic research in interaction (Keck et al., 2006; Mackey & Goo, 2007), type of target feature was selected as a potentially influential variable on the findings of the primary studies. Research in interaction has demonstrated that interactional feedback may be more or less effective depending on the type of target feature, with findings suggesting feedback on lexical or phonological errors may be more
noticeable than on errors in morphosyntax (Carpenter, Jeon, MacGregor, & Mackey, 2006; Ellis, Basturkmen, & Loewen, 2001; Long, 1996; Long, Inagaki, & Ortega, 1998; Mackey, Gass, & McDonough, 2000; Sheen, 2006; Yang & Lyster, 2010). For example, Egi (2007) found that learners participating in interactive tasks processed feedback differently depending on whether the target feature was lexical or morphosyntactic. Results indicated that learners’ L2 performance was influenced by whether recasts were perceived as responses to content or as positive or negative evidence, suggesting that the interpretation and effects of corrective feedback may be mediated by type of target feature. For instance, Long (2006) suggested that implicit forms of feedback may be less effective if the target form is non-salient or communicatively irrelevant. Although recent research in SCMC has not found saliency to be a moderating factor (Yilmaz, 2012), it remains unclear how much of an influence target feature may have on the overall efficacy of interaction. In their meta-analysis, Mackey & Goo (2007) found that interactional benefits were greater for lexical items when compared to grammatical items. Similarly, research in SCMC has demonstrated that the type of target feature may play a role in the efficacy of the interaction, with findings demonstrating that lexical items seem to lead to more instances of negotiation than grammatical items (Blake, 2000; Fernández-Garcia & Martinez-Arbelaitz, 2000; Pellittieri, 1999; Smith, 2004; Toyoda & Harrison, 2002). This analysis examined the mediating influence of lexical, grammatical, socio-pragmatic, and phonological target features.

Type of feedback was also investigated as a potentially influential variable on the relationship between interaction and L2 learning outcomes. Although previous meta-analyses have demonstrated the overall effectiveness of various types of corrective
feedback in face-to-face interaction within classroom and laboratory settings (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), debate remains regarding advantages of one type of feedback over another. For example, results examining the positive benefits of implicit and explicit forms of feedback vary widely across studies, with some findings indicating that implicit forms seem to be more facilitative (e.g. Leeman, 2003), while others support the use of explicit feedback as more beneficial (e.g. Ellis, 2007; Ellis et al., 2006; Loewen & Erlam, 2006). Studies have also found no significant differences between the two types of feedback (e.g. Kim & Mathes, 2001). The wide variation of results suggests that the picture regarding the differential effectiveness of type of feedback in FTF contexts is still unclear.

In addition, the impact of specific types of feedback within computer-mediated contexts has not yet been systematically investigated. Little research within SCMC has examined implicit and explicit forms of feedback (although see Bower et al., 2003; Loewen & Erlam, 2006; Sauro, 2009; Yilmaz, 2012; Yilmaz & Yuksel, 2011). The current research examined both implicit and explicit types of feedback, focusing on the use of recasts, metalinguistic feedback, and negotiation moves, such as comprehension checks, clarification requests, and confirmation checks. Although the operationalizations of these types of feedback may differ across studies, they were treated as single constructs within the current research to facilitate analysis across a large body of work.

In her influential Output Hypothesis (1985, 1995), Swain suggests that the production of output is a necessary condition of SLA as it provides learners with opportunities to notice gaps between their IL and the target language and to test their hypotheses about linguistic forms. In addition, the production of output may also promote noticing, a
component argued to be critical to L2 acquisition (Schmidt, 2001), and offer learners opportunities to reflect on their and their interlocutors’ language. Despite its’ importance to L2 development, few studies have examined the direct link between modified output opportunities and acquisition (Adams, Nuevo, & Egi, 2011; Ellis, 1993; McDonough, 2004, 2005), with the majority of studies providing indirect evidence of the facilitative effects of modified output on L2 development (e.g. Mackey et al., 2003; Pica, 1988; Pica, Holliday, Lewis, & Morgenthaler, 1989; Shehadeh, 1999, 2002; Swain & Lapkin, 1995; Van den Branden, 1997). Previous research has also distinguished between tasks that encourage or require learners to produce modified output (Mackey & Oliver, 2002, Ellis, Loewen, & Erlam, 2006) and those that permit learners to complete the task without doing so (Leeman, 2003). Following Mackey & Goo (2007), the present study focused on whether or not learners were provided with the opportunity to produce modified output rather than whether or not they were required to produce pushed or modified output. One of the reasons for this decision was the lack of studies in SCMC research examining the direct impact of modified output on L2 development (although see Smith, 2008, 2009, 2010 for investigations of self- and other-initiated modified output).

Type of technology was also investigated. During the last two decades, options in technology have not only become more varied, but also more complex. Because computers have become an increasingly common part of the educational process, it is necessary to investigate how different technologies might affect L2 learning processes and outcomes (Felix, 2005). A variety of studies have compared the impact of different computer-mediated methods on L2 development (e.g. Lo, Wang, & Yeh, 2004; Ma, Nicolas, Gierts, & Vandamme, 2002; Smith, Alvarez-Torres, & Zhao, 2003; Yanguas,
2010, 2012), with results revealing interesting differences across mode on L2 learning outcomes. For example, Blake (2008) found a higher number of repair moves occurred in voice chat than text chat, while Yanguas (2010) found advantages for an audio-conferencing group over video and FTF. These findings provide promising information regarding the differential efficacy of technological mode, however, more research is needed to determine if certain approaches are more facilitative of L2 development. This meta-analysis examined different domains within SCMC, such as text-chat, video conferencing, audio conferencing, and bimodal chat, as possible factors impacting learners’ L2 development.

3.3.4. Methodological features

Building on the findings of previous research (Felix, 2005, 2008; Liu et al., 2002), this study coded methodological features in order to examine and describe the methodological quality of the primary research. The methodological features investigated were: the context of the research, treatment setting, learner characteristics, and the reporting of statistical procedures. The selection of these features was guided by previous research within interaction (Keck et al., 2006; Mackey & Goo, 2007) and CALL (Felix, 2005, 2008; Suri, 2000; Zhao, 2003).

Research in interaction has suggested that there may be differences between lab and classroom contexts, with some researchers proposing that the quantity and quality of interactional feedback in classrooms may be highly variable and may be task or context-dependent (Foster, 1998). Previous research has also suggested that feedback provided in laboratory environments is more focused and consistent (Sheen, 2004), and therefore may
be more supportive of L2 development (Nicholas, Lightbown, & Spada, 2001). However, recent meta-analyses (Russell & Spada, 2006; Mackey & Goo, 2007) have indicated that interaction is facilitative of L2 development in both classroom and lab contexts. Although much of the research on CMC has been conducted within a laboratory setting, a growing number of studies within CALL have been conducted in a classroom context (Felix, 2008), highlighting the need for investigation into the overall effects of interaction on L2 development in different CMC research settings.

In addition, this study examined whether the educational context of the primary study influenced L2 development. Limitations to learner-learner interactions have been pointed out specifically with respect to EFL contexts (Fujii & Mackey, 2009; Mayo & Pica, 2000; McDonough, 2004; Philp & Tognini, 2009), with previous research suggesting that learners in a variety of contexts may produce low rates of interactional feedback, due to cultural preferences (Burrows, 2008; Fujii & Mackey, 2009), or other affective factors, such as perceptions of proficiency differences or learner attitude (Philp, Walter, & Basturkmen, 2010; Watanabe & Swain, 2007; Yoshida, 2008). Few studies in CALL research have directly compared second and foreign language contexts, highlighting the need for further investigations into the effects of interaction in diverse contexts.

Because one of the goals of this dissertation is to assess the progression in methodological sophistication over the last two decades of research, research design was selected as a set of features to be investigated. Following previous meta-analyses (Keck et al., 2006; Felix, 2005, 2008; Mackey & Goo, 2007), this synthesis coded the primary research for study design, reporting practices, and use of statistical analyses. Empirical studies within the field of CALL have been criticized for poor research design (Chapelle,
1997; Felix, 2008; Liu et al., 2002), underscoring the need for a comprehensive review of research practices. In addition, studies were coded according to the following learner characteristics: 1) participants’ L1, 2) target language, 3) proficiency level, and 4) educational context (elementary, secondary, or post-secondary). Statistical procedures were also investigated, as this has been an area noted for poor reporting in prior research (Felix, 2008). Studies were coded for their selection of statistical test (e.g. ANOVA, non-parametric tests), as well as whether they reported details regarding effect sizes, confidence intervals, exact $p$ values, and preset alpha levels.

### 3.4. Statistical Analysis

All analyses were conducted using the Biostat meta-analysis software program *Comprehensive Meta-Analysis Version 2.0* (CMA), developed by experts in the field of meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005). This all-inclusive meta-analysis program provides researchers with a wide range of tools in a flexible and user-friendly format. It is robust enough to facilitate analysis using different data formats, and provides results for both fixed and random effects models, allowing researchers to compare and contrast findings across the different models. In addition, CMA offers the user a variety of options for customizing and exporting forest plots, which have been identified as one of the more reliable visual interpretations of meta-analytic data (Bax et al., 2009) and were previously difficult for researchers to produce (Lewis & Clarke, 2001). In addition, the program produces a reliable assessment of heterogeneity and publication bias, and has been referred to as the ‘most sophisticated stand-alone software
package for meta-analysis’ by researchers in the social sciences, such as Littell, Corcoran, & Pillai (2008, p. 146).

3.4.1. Fixed effects and random effects model

Meta-analyses are based on either the fixed-effects or random-effects model depending on the assumptions of the data. Fixed-effects models assume that all studies included in the sample share a common effect size. In other words, the factors that might influence the effect size are constant across all the studies, making the true effect size the same. Random-effects models, on the other hand, assume that effects are normally distributed and vary due to heterogeneous factors. Selecting the appropriate model is critical (Hedges, 1982; Lipsey & Wilson, 2001), as each model is likely to generate varying results. The decision regarding which model to use should be based on a number of factors, including the researcher’s understanding of whether the collected studies share a common effect size and the ultimate goals of the analysis. Because the participants and treatments used in the sample studies varied in ways that may have influenced the outcomes, for example treatment length and intensity, it is logical to assume that there is not a common effect size. In addition, the goal of the current analysis is to make generalizations regarding interaction in SCMC across a range of populations and scenarios, rather than drawing conclusions for identical, narrowly defined participants, providing additional support for the selection of the random-effects model over the fixed-effects model. The studies included in the current study have been sampled from a distribution of effect sizes and from the published literature, further supporting the selection of a random effects model (Borenstein et al., 2009). To further ensure that the
correct model was selected for the current analysis, the test of homogeneity, also known as a $Q$ test, was used to statistically determine whether the between-studies variance observed is zero.

In order to assess whether the average effect sizes estimate the same population effect size, $Q$ tests were conducted to examine the distribution of the various effect sizes for each contrast. The null hypothesis of homogeneity was tested for each comparison, and with the exception of the receptive skills contrast, all null hypotheses were rejected, indicating that there are differences in true effect sizes across studies, $p < .05$. These results indicate that the variation in the effect sizes is larger than expected from within-subject sampling error alone. Instead, there are differences across effect sizes that are likely caused by individual study characteristics, such as sample size, that are randomly distributed, leading to the selection of a random-effects model for all comparisons included in the current study. However, it should be noted that although the $Q$ test resulted in a non-significant result for the comparison of receptive skills, $p > .05$, a random-effects model was deemed appropriate due to the small number of effect sizes ($k = 7$) based on small participant samples used in this particular comparison (Lipsey & Wilson, 2001). As Borenstein et al. (2009) point out, a non-significant $p$ value does not always indicate evidence of consistent effect sizes, as it may instead be due to low power. In addition, the 14 unique studies included in the sample were conducted by independent researchers and were thus unlikely to have participants or treatments that functioned equivalently across studies. Random-effects models are able to account for this random variability at both the study and participant level, making this model more appropriate for the current research than a fixed-effects model. Furthermore, study weights are argued to
be more balanced under the random-effects model (Borenstein et al., 2009), as the 
random-effects model estimates the mean of a distribution of effects, allowing the meta-
alyst to use a wider range of study features to explain observed variability. Based on 
the results of the $Q$ tests, as well as the small sample size and independent nature of the 
included studies, a random-effects model was selected for all of the following 
comparisons.

3.4.2. Calculation of effect sizes

A variety of effect sizes were utilized in the current analysis, and were selected 
depending on the statistics reported in the selected studies. Previous meta-analyses on L2 
interaction and corrective feedback research have used Cohen’s $d$ (Keck et al., 2006; Li, 
2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006) to estimate 
effect sizes, as the descriptive statistics necessary for computing this statistic are 
commonly reported. For instance, Cohen’s $d$ is based on means, standard deviations, and 
sample size, which are usually found in the literature, particularly studies published in the 
last decade. However, not all studies used in the current analyses reported the descriptive 
statistics needed for the calculation of Cohen’s $d$. For instance, one study (Sykes, 2005), 
reported only the means and $p$ value of the post-hoc comparisons. In order to include the 
effect size in the analysis, STATA was used to estimate the standard deviation based on 
the reported $p$ value. For studies that failed to report the mean or standard deviation, 
separate calculations were conducted for studies reporting only sample size ($n$) and $t$ or $F$ 
values following Lipsey & Wilson (2001):
Contrasts for effect sizes were based on a number of two-variable group comparisons. Although the ideal comparisons would stem from studies comparing the effects of a single experimental condition to a single control condition on a single dependent variable (Norris & Ortega, 2000), interaction research does not always follow such a straightforward and simple study design. Instead, many studies use a comparison group rather than a true control group, which would not receive any interactive treatment. However, because one of the goals of this study is to compare FTF and SCMC groups, effect sizes for the current meta-analysis were calculated by comparing the treatment group, in this case the SCMC group, with the FTF comparison group. To facilitate this comparison, SCMC data were consistently coded as “experimental” data and FTF data were coded as “baseline” data. Following this classification system, a positive comparative effect would suggest an advantage for SCMC while a negative comparative effect would indicate an advantage for FTF.

Effect sizes were then calculated for each separate group on immediate posttests. Although previous research has included effect sizes based on delayed posttests (Keck et al., 2006; Mackey & Goo, 2007), only three studies in the current sample reported delayed posttest scores (Baralt, 2010; de la Fuente, 2003; Yanguas, 2012). Due to the insufficient amount of data, delayed posttests were not included in the current analysis. Because effect sizes were calculated for each of the possible contrasts in each study according to the inclusion criteria, multiple effect sizes were obtained from a single study. In other words, for studies that examined the acquisition of different features,
separate effect sizes were calculated for each feature. Although some meta-analysts have argued against this approach (Light & Pillemer, 1984), instead suggesting that effect sizes should be combined within studies to produce one average effect size for each individual study, multiple effect sizes extracted from a single study can provide a clearer assessment of the impact of features of interest on the efficacy of the target treatment. For instance, Keck et al. (2006) elected not to combine multiple within-study effect sizes in order to gain a better understanding of the influence of certain features, such as task type or type of target feature, on the effects of interaction. This same method was applied in the current study, and allowed for the investigation of the effects of specific features on the efficacy of interaction in SCMC at promoting L2 learning outcomes, providing a targeted and overall global perspective.

However, including an effect size for each outcome measure used within individual studies can lead to an improper estimate of the precision of the summary effect (Borenstein et al., 2009). Because the separate outcomes, such as measures of oral and written production, have been completed by the same set of participants, the data is not independent, although it is treated as such by the statistical program. This non-independence of data points can lead to an underestimation of the standard error, which may result in unreliable statistical results (Lipsey & Wilson, 2001). In order to correct the variance in order to account for the relationships between data points, and therefore increase the robustness of the calculations, effect sizes were combined across similar outcomes measures. For example, in order to compute an accurate effect size for overall L2 learning outcomes, combined effect sizes were calculated from all of the outcome measures within one study, resulting in a ‘one study, one effect size’ for the main
analysis. For the analyses of productive and receptive skills and oral and written skills, multiple effect sizes were averaged for each category of outcome measures within individual studies. In other words, for the analysis of productive skills, a combined effect would be averaged from the oral and written production measures, providing an average of the appropriate dependent variable. This approach allowed for a robust examination of multiple outcomes, providing summary effects for the impact of mode on overall L2 learning outcomes, as well as the difference in effect sizes across various skills.

To review, the first contrast examined the impact of the independent variable of mode, operationalized as FTF and SCMC interaction, on the dependent variable of overall L2 learning outcomes, operationalized as posttest scores. Learning outcomes were defined here as any measure of L2 oral or written performance, including accuracy and complexity, or development, including both productive and receptive measures of lexical, syntactic, morphological, and phonological knowledge. This approach was used to obtain a global perspective on the differences, if any, between interaction in FTF and SCMC modes on learners’ overall development and learning.

In order to attain a more finely-grained analysis of the impact of modality on specific learning outcomes, additional two-variable contrasts were conducted to examine the effects of FTF and SCMC interaction on learners’ productive (speaking and writing) and receptive (listening and speaking) skills, as well as the effects of FTF and SCMC interaction on learners’ oral and written skills. Previous research has demonstrated differences in learners’ performance on productive and receptive tasks (e.g. Benati, 2005, 2009; Collentine, 2005; De Jong, 2005; DeKeyser & Sokalski, 1996; Erlam, 2003; Farley, 2001; Izumi, 2002; Lee & Bonati, 2007; Morgan-Short & Bowden, 2006; Nagata,
and oral and written tasks (e.g. Grabowski, 2005; Granfeldt, 2008; Ferrari & Nuzzo, 2009; Houck & Gass, 1996; Kuiken & Vedder, 2008, 2009, 2011, 2012; Martinez-Flor, 2006; Yu, 2010), with a recent meta-analysis providing evidence that comprehension and production based instruction results in differential learner performance on posttests (Shintani, Li, & Ellis, 2013). Overall, this research suggests the possibility that interaction may have a differential impact on the learning and development of specific features across modes.

In sum, the following effect sizes were calculated:

a. Overall average effect size for interaction treatment
b. Average effect size for productive and receptive target features
c. Average effect size for oral and written target features
d. Average effect size for second and foreign language settings
e. Average effect size for classroom and laboratory-based contexts
f. Average effect size for type of interlocutor (researcher vs. student, NNS vs. NS)

Because the effect sizes reported in this study are comparative, SCMC data were coded as ‘experimental’ while FTF data were coded as the ‘baseline’ data. Effect sizes in the negative direction indicate an advantage for FTF interaction while effect sizes in the positive direction indicate an advantage for SCMC interaction. Confidence intervals of 95% were also calculated in order to provide information regarding the reliability of the mean effect sizes and the differences across groups. Confidence intervals indicate the
range of values within which the target mean is likely to be (Lipsey & Wilson, 2001), providing useful information regarding the precision of the estimated effect sizes.

3.4.3. *A priori* methodological decisions

Because the overall sample size of this meta-analysis was small with 14 unique studies and 37 total effect sizes, there was a possibility that outliers would skew the data, potentially producing misleading results. The CMA software program provides for the examination of change in the mean effect size when one study is removed from the sample. Using this approach, outliers were defined as any study resulting in a change in the mean effect size by more than .05 standard deviation units. Because one study identified as an outlier in one contrast may not necessarily act as an outlier in another contrast, analyses were performed for each comparison in order to maintain as much data as possible for each individual contract. Analyses indicated that none of the identified studies were extreme outliers, therefore, all remained in the final sample. Alpha levels were set at .05 for all analyses.
4.1. Descriptive Results

4.1.1. Research distribution

Fourteen studies comparing the effectiveness of interaction in FTF and SCMC contexts met the inclusion criteria for the current meta-analysis. Of these, eight (48%) were published in the following refereed journals: CALICO, Computer-Assisted Language Learning, Modern Language Journal, and System. The remaining six (42%) were unpublished doctoral dissertations. Figure 1 demonstrates the publication frequency of research comparing the effectiveness of interaction in FTF and SCMC contexts.

![Publication Year](image)

*Figure 1. Included studies with their years of publication*

4.1.2. Publication bias

In order to explore for the possibility of publication bias, a funnel plot analysis was used to analyze the current sample. A biased sample would be more likely to yield an
overestimated effect size due to the possibility that published studies are more likely to have significant effects. If publication bias were present, the bottom of the funnel plot would show a higher concentration of studies on one side of the mean than the other. This type of distribution would reflect the tendency for smaller studies with larger than average effect sizes, making them more likely to achieve statistical significance, to be published (Borenstein et al., 2009). Figure 2 demonstrates that the majority of effect sizes were equally distributed around the mean, indicating the absence of publication bias. Studies with larger sample sizes appear towards the upper portion of the funnel and are relatively evenly distributed about the mean, with the graph indicating that medium and larger scale studies with medium effect sizes were well represented. In order to search for potential missing values that would alter the mean effect size, a trim-and-fill analysis was conducted. Trim-and-fill analysis, developed by Duval and Tweedie (2000), provides information regarding where missing studies would be likely to fall within the analysis, and then recalculates the combined effect with these potentially missing studies. Using a random-effects model, the Trim and Fill procedure did not indicate any missing values that would change the obtained average effect size.
In addition, Cooper’s Fail-Safe N (1979, 1998) was conducted to assess the possible impact of relevant studies missing from the current sample. The 14 combined effect sizes from the analysis resulted in a $z$-value of 1.60 and a non-significant $p$ value, $p = .11$, suggesting that the observed effects in the following analyses are authentic and are not due to the presence of bias.

4.1.3. Research setting

The fourteen studies included in the current study represented a variety of instructional settings, research contexts, and learner characteristics. The majority of studies (71%) occurred in a classroom-based environment ($n = 10$), with only four of the studies (29%) in the sample taking place in a laboratory context. Seventy-nine percent ($n = 11$) of the studies were conducted in a foreign language (FL) context, with the remaining three studies (21%) in a second language (SL) context. The target language

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{funnel_plot.png}
\caption{Publication bias: Funnel plot by effect sizes.}
\end{figure}
contexts included Spanish as a Foreign Language (43%), English as a Second Language (21%), English as a Foreign Language (14%), German as a Foreign Language (14%), and Japanese as a Foreign Language (7%). Table 2 summarizes the research characteristics of the included studies.
Table 2. Included studies by research characteristics

<table>
<thead>
<tr>
<th>Setting</th>
<th>( k )</th>
<th>Context</th>
<th>( k )</th>
<th>Interlocutor</th>
<th>( k )</th>
<th>Outcome Measure*</th>
<th>( k )</th>
<th>SCMC mode*</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab</td>
<td>4</td>
<td>FL</td>
<td>11</td>
<td>Native speaker</td>
<td>2</td>
<td>Open-ended production</td>
<td>11</td>
<td>Oral</td>
<td>3</td>
</tr>
<tr>
<td>Classroom</td>
<td>10</td>
<td>SL</td>
<td>3</td>
<td>Non-native speaker</td>
<td>12</td>
<td>Close-ended response</td>
<td>4</td>
<td>Text</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Selected response</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*In some studies, more than one outcome measure or mode was used, so the sum of these categories is not the same as the total number of studies.
4.1.4. Learner characteristics

Eleven different participant L1s were reported across the various studies, with some studies reporting diverse learner populations including speakers of English (64%), Spanish (29%), Chinese (21%), Korean (14%), Portuguese (7%), French (7%), Korean (7%), Arabic (7%), and Thai (7%). One study (Petersen, 2010) also reported L1s of Amharic and Urdu, although the exact number of learners with these L1s was not stated. Most of the studies were conducted in university contexts (79%), while the remaining 21% were in high-school contexts. Participants’ ages ranged from 14 to 25 years old, although 57% of the studies failed to report an age range or mean age for their participants, highlighting the need for more rigorous reporting of participant characteristics. Eleven studies (79%) reported participants’ proficiency level as intermediate and three reported beginning proficiency levels (21%), however, only one study (Petersen, 2010) considered the developmental readiness of participants. Most studies used institutional status to designate learners’ general proficiency level (64%), a finding similar to that of Mackey and Goo (2007) and Keck et al. (2006), while others used in-house placement tests (7%) or standardized placement tests (7%). Twenty-nine percent did not specify how learners’ proficiency was categorized and measured. Table 3 illustrates learner characteristics across the studies included in the sample.
## Table 3. Included studies by learner characteristics

<table>
<thead>
<tr>
<th>Age</th>
<th>L1</th>
<th>L2</th>
<th>Proficiency</th>
<th>Proficiency measure</th>
<th>Academic status</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-18</td>
<td>Chinese</td>
<td>1</td>
<td>English</td>
<td>Beginning</td>
<td>Institutional status</td>
</tr>
<tr>
<td>19-25</td>
<td>English</td>
<td>8</td>
<td>German</td>
<td>Intermediate</td>
<td>In-House</td>
</tr>
<tr>
<td>NR*</td>
<td>Spanish</td>
<td>1</td>
<td>Japanese</td>
<td></td>
<td>Standardized</td>
</tr>
<tr>
<td></td>
<td>Mixed</td>
<td>4</td>
<td>Spanish</td>
<td></td>
<td>Not reported</td>
</tr>
</tbody>
</table>

*Not reported
4.1.5. Research design

Because this analysis sought to compare the effects of interaction in FTF and SCMC contexts, every study included in the sample had a comparison group. Only three studies used a true control group, in which learners did not participate in any interactional treatment or task, as part of their design (21%). Overall sample sizes included in the analysis had a wide range from 20 to 70 learners total (M = 44.5, SD = 17.27), including only the comparison and treatment groups. Group sizes ranged from nine participants to 36, with a mean group size of 17.33, SD = 8.10. Studies varied in how length and duration of treatment was reported, with frequency of treatment ranging from one session to 16 sessions taking place over one semester.

Outcome measures were coded according to Gass & Mackey’s (2007) text on data elicitation measures in order to provide comparability to previous meta-analytic research in interaction (Keck et al., 2006; Mackey & Goo, 2007). Three types of outcome measures were used in the included studies, with open-ended prompted productions, such as oral interviews and written production tasks, being the most commonly used (79%). Closed-ended prompted response tasks were the next most frequent outcome measure (29%), followed by prompted response measures, such as multiple-choice tests (21%). Naturalistic production was not found to be a measure of learning outcomes in any of the studies in the current sample. Overall, production measures were used in the majority of studies as assessment tasks, with the majority of studies favoring oral production tasks over written production (71%).

Previous research has suggested that researchers may elect to test learners in the same modality as the treatment (Gass & Mackey, 2006). However, some studies in SCMC have
sought to examine the effectiveness of written modalities, such as text-chat, in facilitating learners’ oral development (e.g. Abrams, 2003). Although three studies in the current sample used audio or video chat (Bueno-Alastuey, 2011; Sykes, 2005; Yanguas, 2012), the majority utilized text-chat as the treatment mode. This preference for pairing text-chat treatment tasks with oral production assessment measures might be explained by researchers’ desire to provide evidence of the crossover benefits of interaction in SCMC. For instance, although much of the research uses text-chat rather than video or audio chat, interaction in SCMC still provides positive benefits for learners’ oral production skills, suggesting that learners’ improvement is not limited to the mode of the treatment task, but instead transfers to novel skills in the assessment task. In other words, the written interaction of text-chat can aid learners in improving their oral interaction, demonstrating the potential of SCMC to benefit learners written and oral production. In addition, researchers’ frequent use of oral production tasks provide a more readily comparable assessment measure with which to examine the similarities and differences of interactional benefits in FTF and SCMC contexts.

4.1.6. Statistical analyses and reporting

Although a large sample of studies were identified through the initial search and retrieval process (N= 543), many were eliminated during the coding phase due to incomplete reporting of statistical data necessary to calculate effect sizes (e.g. Beauvois, 1998; Salaberry, 2000; Warschauer, 1996). Of the fourteen studies that met the inclusion criteria, all but one used inferential statistics for the analyses. The most commonly used statistical test was ANOVA (64%), followed by t-tests (29%) and ANCOVA (21%), to
examine the differences in learning outcomes between groups. Previous analyses have also found ANOVA to be the most frequently used inferential test (Keck et al., 2006; Mackey & Goo, 2007, Plonsky & Gass, 2011). However, despite the prevalence of inferential statistics, only one of the studies published in peer-reviewed journals (Yanguas, 2012) reported testing the assumptions necessary for the selected statistical procedure, such as tests of normality. Four dissertations in the sample stated that assumptions had been checked before running the statistical analyses. (Baralt, 2010; Hirotani, 2010; Lin, 2009; Petersen, 2010), but 71% of the sample failed to mention an *a priori* alpha level. Similar to the results of Keck et al. (2006), less than half the studies (43%) reported an exact *p* value. Only five reported effect sizes (36%), four of which were dissertations. Previous research has highlighted the need to report effect size, a feature critical to understanding the amount of variance observed in the dependent variable due to the influence of the independent variable. Based on the data in the current sample, there is still ample room for improvement in the reporting of this important feature. In addition, more than half of the sample (57%) failed to report any reliability measures for interrater reliability or internal validity and reliability of testing materials, while none of the studies reported confidence intervals. Although these issues in reporting might impact readers’ ability to thoroughly understand the results, nearly all of the studies (except Conaim & Wong, 2004) within the current sample relied on statistical significance to interpret the findings.

### 4.2. Meta-Analytic Results
The current meta-analysis combined effect sizes from 14 unique study reports to compare the effectiveness of interaction in FTF and SCMC contexts on L2 learning outcomes. Effect sizes were combined to compare the benefits of interaction in FTF and SCMC as measured on immediate posttest scores. In addition, effect sizes were calculated to assess the impact of interactional mode on the development of learners’ productive and receptive abilities and learners’ oral and written skills.

4.2.1. Overall effectiveness of interaction in SCMC compared to FTF

Table 4 provides the effect size statistics for the first comparison, where $k$ represents the number of unique studies and $N$ represents the total number of individual participants in the selected studies.
Table 4. Descriptive effect size statistics for overall L2 learning outcomes

<table>
<thead>
<tr>
<th>k</th>
<th>N</th>
<th>Total effect sizes</th>
<th>Combined effect sizes</th>
<th>Effect size (d)</th>
<th>SE_{SM}</th>
<th>Z-Value</th>
<th>95% CI</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>623</td>
<td>37</td>
<td>14</td>
<td>.13</td>
<td>.125</td>
<td>1.071</td>
<td>-.11 to .38</td>
<td>29.481</td>
</tr>
</tbody>
</table>
According to the random-effects model, the overall weighted mean effect size of the 14 combined effect sizes ($d = .13$) indicates that there is no significant difference ($p = .284$) in learning outcomes as measured by posttests when learners participate in either FTF or SCMC interactional modes. According to Cohen’s effect size magnitude (1988, 1992), $d \leq .2$ is considered to be a small effect size. The effect size for the current comparison is small, suggesting that interaction in SCMC contexts provides a slight advantage over FTF contexts in facilitating L2 learning outcomes, as it accounts for approximately 13% percent of the variance due to mode of interaction. These results indicate that groups participating in interaction in SCMC contexts do not have significantly better learning outcomes than those learners in FTF contexts. Figure 3 provides a forest plot of the current analysis, illustrating the average effect size for the benefits of interaction in SCMC contexts compared to FTF contexts on L2 learning outcomes. Individual study weights are indicated by the size of the box with which they are marked, with larger effect sizes indicated by larger boxes.
**Figure 3.** Forest plot of standardized mean effect sizes for overall L2 learning outcomes.
Figure 4 shows the individual mean effect sizes for each group individually, demonstrating that interaction treatments in both SCMC and FTF result in large effect sizes for overall L2 learning. Although interaction in SCMC ($d = 1.13$) resulted in higher scores on immediate posttests than FTF interaction ($d = .84$), indicating that interaction in both FTF and SCMC contexts lead to significant improvements ($p < .001$) in L2 learner outcomes, the difference between the modalities is not statistically significant.

*Figure 4. Comparison of SCMC and FTF communication modes on overall L2 learning outcomes*
4.2.2. Effectiveness of interaction in SCMC compared to FTF for receptive and productive outcomes

The descriptive statistics for the second set of contrasts are provided in Table 5 and Table 6. Treatments focused on receptive learner outcomes including reading and listening as measured by immediate posttests, resulted in a small effect for FTF interaction, $d = -.12$, $p = .460$. However, the extremely small sample size of individual effect sizes ($n = 7$) included in this comparison highlights the need to interpret the results of this particular contrast with caution. Due to the small sample size, this comparison may be lacking the precision of the other contrasts. In order to make a more robust assessment of the impact of interaction in SCMC contexts on learners’ receptive skill development, more studies are needed.

Table 5. Descriptive effect size statistics for receptive L2 outcomes

<table>
<thead>
<tr>
<th>$k$</th>
<th>Total effect sizes</th>
<th>Effect size ($d$)</th>
<th>SE$_{SM}$</th>
<th>Z-Value</th>
<th>95% CI</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7</td>
<td>-.12</td>
<td>.165</td>
<td>-.738</td>
<td>-.45 to .20</td>
<td>1.712</td>
</tr>
</tbody>
</table>

For productive measures, including oral and written production, results indicated a small effect size for SCMC, $d = .15$, which was not significant, $p = .230$. This finding suggests that SCMC has a small effect on learners’ productive skills when compared to learners participating in FTF interaction. However, it should be noted that effect size for the SCMC group in Bueno-Alastuey’s (2010) study was quite large ($d = 3.20$), which may have resulted in a slightly inflated average effect size for this specific comparison.

Table 6. Descriptive effect size statistics for productive L2 outcomes

<table>
<thead>
<tr>
<th>$k$</th>
<th>Total effect sizes</th>
<th>Effect size ($d$)</th>
<th>SE$_{SM}$</th>
<th>Z-Value</th>
<th>95% CI</th>
<th>Q</th>
</tr>
</thead>
</table>
Results indicated large effect sizes on both receptive and productive measures in SCMC ($d = 1.31; d = .97$) and FTF ($d = 1.44; d = .78$), respectively. Figure 5 provides the average effect sizes of the benefits of interaction in SCMC and FTF contexts on receptive L2 learning outcomes, while Figure 6 illustrates the mean effects for productive L2 learning outcomes.

*Figure 5. Comparison of SCMC and FTF communication modes on L2 receptive outcomes*
4.2.3. Effectiveness of interaction in SCMC compared to FTF for oral and written outcomes

Meta-analytic descriptive statistics for the third contrast can be found in Tables 7 and 8. Analyses indicated large effect sizes for SCMC in oral ($d = 1.04$) and written ($d = .91$) learning outcomes and large and medium effect sizes for FTF in oral ($d = 1.24$) and written ($d = .38$) measures, suggesting that learners in all groups experienced improvement due to the interactional treatments. Figure 7 illustrates the effect sizes of both SCMC and FTF communication modes on L2 oral outcomes.

Figure 6. Comparison of SCMC and FTF communication modes on L2 productive outcomes
In the contrast examining the effectiveness of interaction in FTF and SCMC contexts, findings indicate a small effect size on the development of oral skills for FTF over SCMC contexts, \( d = -0.04 \), although this difference was not significant, \( p = 0.89 \). Table 7 provides the descriptive statistics for the comparative efficacy of interaction in FTF and SCMC on learners’ oral outcomes.

**Table 7. Descriptive effect size statistics for oral L2 outcomes**

<table>
<thead>
<tr>
<th>( k )</th>
<th>Total effect sizes</th>
<th>Effect size (( d ))</th>
<th>( \text{SE}_{\text{SM}} )</th>
<th>Z-Value</th>
<th>95% CI</th>
<th>( Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>11</td>
<td>-0.04</td>
<td>0.163</td>
<td>-0.297</td>
<td>-0.297 to 0.343</td>
<td>33.352</td>
</tr>
</tbody>
</table>

*Figure 7. Comparison of SCMC and FTF communication modes on L2 oral outcomes*
In the comparison of FTF to SCMC contexts, results indicated a small effect size in favor of SCMC for written learning outcomes, $d = .16$, $p = .123$, indicating that approximately 16% of the variance in learner performance on written outcomes can be explained by the mode of interaction, with findings suggesting that interaction in SCMC results in small effects for learners’ written language skills. However, these results should be interpreted with caution, as the effect size calculated for the FTF group in Yanguas (2012) was much larger than the other sampled studies ($d = 3.20$), which may have inflated the overall average effect size.

In addition, the total number of effect sizes used to calculate the written contrast is small ($n = 8$). According to Borenstein et al. (2009), with a small number of effects, the estimate of the between-studies variance is more likely to be in error. Because the average effect size is based partially on this variance, there is a greater likelihood of an unreliable effect when small sample sizes are used for the calculation. Although this statistical approach may be preferred to ad-hoc approaches, such as vote-counting, the impact of mode may have a smaller or larger effect than demonstrated here. Rather than drawing definitive conclusions based on these findings, additional studies examining the impact of mode on learners’ written language skills should be conducted in order to form a more robust sample size, thus strengthening the conclusive validity of the findings of this comparison. In addition, other meta-analytic work might adapt the eligibility criteria in order to increase the sample size of included studies, providing the data necessary for more reliable conclusions.
Although one of the intended goals of this analysis was to investigate the impact of SCMC mode, including audio, video, and text chat, on the average effect size, the studies included in the current sample did not yield sufficient variation for statistical analyses.
The majority of studies (79%) used a text-chat feature, such as Yahoo messenger, Google chat, or Interchange, as the mode of interaction, with only three (21%) using audio or video chat. Although video chat software programs are widely available, based on the current sample, text-chat is by far the most commonly used modality for SCMC. In addition, despite growing interest in multimodality in SCMC (Collentine, 2010), in which participants can interact using more than one form of communication, no multi-modal studies were included in this research. Overall, after an extensive search of relevant studies, only a handful examining video or audio forms of SCMC were identified (Jepson, 2005; Lee, 2007, Yanguas, 2010, in addition to the studies included in the current sample). Based on the data presented in this study, this small yield of oral SCMC research highlights the need for more diverse investigations using a wider range of available technology to examine the effects of interaction on learning outcomes. Oral SCMC offers a fertile environment in which researchers can investigate established L2 theories and hypotheses in a potentially under-investigated context, with previous studies demonstrating potential benefits not only for L2 linguistic (Yanguas, 2012) and pragmatic (Sykes, 2005) development, but also for learners’ social development (Yamada & Akahori, 2007) and anxiety levels (Müller-Hartmann & Schocker-von Ditfurth, 2010). Although the lack of studies examining oral SCMC in the present sample makes any comparisons across modes difficult, these results identify the need for more research to facilitate robust statistical analyses of average effect sizes across different types of SCMC, providing a positive contribution to the field by bringing attention to this gap in the literature.
Similarly, although all of the studies included in the current analysis utilized interactional treatments where feedback and negotiation is likely to occur, the type of feedback taking place during treatment was not always reported. Two studies reported using recasts during the treatment tasks (Baralt, 2010; Petersen, 2010) and three described the use of meta-linguistic feedback (Lin, 2009; Payne & Whitney, 2002; Sullivan & Pratt, 1996) as part of the interaction. Two studies indicated that negotiation in general took place (de la Fuente, 2003; Yanguas, 2012), and one stated that feedback was provided although there was no specific indication of the type (Kost, 2004).

Although it may be possible to speculate about the type of interactional feedback that may have taken place during these interactions based on the type treatment task, such as the potential for information gap tasks to result in learners negotiation for meaning (Pica et al., 1993), without clear documentation from the primary research, statistical analyses examining individual feedback types are likely to produce inaccurate results due to the lack of classification in the primary research. The remaining six studies (43%) did not provide any specific information regarding the provision or use of feedback during the interactional treatments.

Although the lack of information regarding feedback types prevented a statistical comparison in the current analysis, the question concerning the differential effectiveness of specific feedback types in FTF and SCMC contexts remains. The substantial body of research examining the effects of different types of corrective feedback (see Li, 2010, and Russell & Spada, 2006, for recent reviews), ranging from implicit forms such as recasts to more explicit types like meta-linguistic correction, demonstrates the keen interest in the field at addressing which type of feedback may best facilitate L2 development. The
current research reveals that this agenda has not been pursued as thoroughly in computer-mediated contexts as in face-to-face contexts, and although the extensive search phase of the current study identified experimental studies (Ayoun, 2001; Bationo, 1991; Loewen & Erlam, 2006; Sauro, 2009; Sachs & Suh, 2007; Yilmaz, 2012) investigating the impact of implicit and explicit corrective feedback in SCMC L2 development, the overall number of studies specifically focusing on feedback type remains limited. Research has demonstrated the potential impact of communication mode on the provision of feedback (Yilmaz, 2012; Yilmaz & Yuksel, 2011), with learners receiving recasts in SCMC outperforming those that received feedback in FTF contexts. However, based on the data included in this study, findings indicate that this area remains underexplored, as a relatively small number of studies examining the relationship between different types of corrective feedback, L2 development, and communication mode, were found. These results highlight the need for further investigation into this promising area, particularly because SCMC offers a new context in which researchers might explore how individual types of feedback differentially affect L2 development. Additional research in this area would also make valuable theoretical and empirical contributions to the ongoing debate regarding the efficacy of implicit and explicit feedback (Long, 2007; Lyster, 2004), as well as provide important information regarding the potential benefits of using and providing corrective feedback in computer-mediated contexts. Overall, although there was not sufficient information in the current study to warrant a statistical contrast between individual types of feedback, this finding is important nonetheless, as it identifies an area of SCMC research in need of further exploration.
Previous studies have investigated the role of modified output in L2 development (Egi, 2010; Mackey et al., 2003; Mackey & Goo, 2007; Swain, 1985, 1995, 2005), with findings indicating that modified output plays an important role in L2 learning. However, only a handful of studies have directly investigated the impact of modified output opportunities on facilitating learners’ acquisition of target forms (Adams, Nuevo, & Egi, 2011; Ellis, 1993; McDonough, 2004, 2005). Based on the data included in the current study, this area of research remains under-investigated in both FTF and SCMC contexts, as none of the studies selected for the current analysis directly examined the efficacy of modified output opportunities on L2 learning outcomes. Furthermore, none of the current studies explicitly mentioned modified output as part of the research design, suggesting that modified output opportunities were uncontrolled for in each of the studies in the sample. This finding underscores the need for more research examining modified output opportunities in order to further our understanding regarding the role of modified output in SCMC based interaction.

Previous FTF interaction research has demonstrated differential effects depending on the type of target feature, finding that interaction appears to have a significantly more positive impact on lexical rather than grammatical features (Mackey & Goo, 2007). Although studies have shown that interaction facilitates the acquisition of both grammatical and lexical forms, lexis seems to trigger negotiations more frequently than morphosyntax in both FTF (e.g. Gass & Alvarez-Torres, 2005; Jeon, 2007; Loshky, 1994; Mackey & Goo, 2007) and SCMC environments (e.g. Blake, 2000; Blake & Zyzik, 2003; Fernández-Garcia & Martínez-Arbelaitz, 2002; Pellettieri, 1999; Shekary & Tahririan, 2006; Smith, 2005; Toyoda & Harrison, 2002). The current analysis coded for
type of target feature, and although findings reveal a positive effect for interaction in SCMC on overall learning outcomes, there were not enough individual studies focusing on similar target features to warrant a more finely-grained statistical analysis. Instead, percentage values are reported here to illustrate the diversity of the sample. Twenty-nine percent (n = 4) assessed the acquisition of grammatical targets (29%), 14% investigated the acquisition of lexical features (n = 2), and 7% (n = 1) explored the development of pragmatic knowledge. The remaining studies examined fluency (14%), complexity (7%), or general oral or written proficiency (29%). Although more research is needed in order to examine the effect of mode of interaction on specific learning outcomes, the current results provide promising information regarding the overall impact of interaction in SCMC on L2 learning outcomes. The wide variety of target items featured in the current sample, coupled with the positive overall average effect size on L2 learning outcomes, demonstrates the potential of SCMC interactional treatments to improve performance and development on a diverse set of learner measures.

4.3. Methodological Factors

4.3.1. Research setting (second or foreign language)

One of the goals of the current analysis is to examine whether the research context had any impact on the observed mean effects. Medium average effects were found for both SCMC (d = .57) and FTF (d = .33) for second language contexts, while foreign language contexts resulted in large average effect sizes of .96 for SCMC and .76 for FTF. Individual analyses revealed that each individual mode resulted in significant medium and large effects for learning outcomes, while the comparison between the two
communication modes revealed a small effect for SCMC foreign language contexts, \( d = .08 \), and a small effect in favor of SCMC in second language contexts, \( d = .33 \), although neither effect was statistically significant, indicating that neither context has a substantial impact on the overall effectiveness of either mode.

### 4.3.2 Research context (laboratory or classroom)

Although much of the body of research investigating interaction was conducted in laboratory contexts (e.g. de la Fuente, 2002; Ellis, 2007; Ellis & He, 1999; Ellis et al., 1994; Jeon, 2007; Linnell, 1995; Mackey, 1999, 2006; Mackey & Oliver, 2002; Mackey & Philp, 1998; Mackey & Silver, 2005; McDonough, 2005; McDonough & Kim, 2009; Silver, 2000; Philp, 2003), previous meta-analyses have demonstrated that interaction is also beneficial for L2 development in classroom contexts (Mackey & Goo, 2007).

Results of the current analysis indicate large effect sizes for research conducted in laboratory contexts in both SCMC (\( d = 1.01 \)) and FTF (\( d = .95 \)), and medium to large effect sizes for research conducted within the classroom (SCMC, \( d = .82 \); FTF, \( d = .52 \)). Similar to previous findings in interaction (Mackey & Goo, 2007), lab studies overall resulted in a significantly larger effect size (\( d = .98 \)) than classroom-based research (\( d = .68 \)). Both contexts demonstrate substantial effects, indicating that L2 learning does occur in lab-based and classroom environments.

However, the current research is most concerned with the comparative effectiveness of SCMC and FTF modes. Therefore, lab and classroom based studies were compared in terms of the impact of communication mode on the efficacy of interaction in facilitating L2 learning outcomes. Contrasts of interaction in FTF and SCMC environments revealed
a very small effect size for SCMC in laboratory settings, $d = .06$, although it was not significant. This finding suggests a slight trend in favor of computer-mediated laboratory contexts over FTF lab-based contexts, a result potentially influenced by social pressure and anxiety. For example, in a lab-based environment where learners interact one-on-one with an interlocutor, a situation that could potentially cause anxiety and therefore impact L2 learning, there may be an advantage for computer-mediated contexts over FTF.

Previous studies have suggested that one of the benefits of computer-mediated interaction is the potential to reduce learners’ anxiety levels (Chun, 1994; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996; Warschauer, 1996; Abrams, 2003). As Beauvois (1992) points out, there may be less immediacy in SCMC than FTF, providing learners with additional time to process input and produce output. This added time may be beneficial to learners with greater levels of anxiety or for learners with low proficiency levels. In addition, Kern (1995) found that introverted learners may be more likely to participate in SCMC contexts, with students indicating on self-report measures that they ‘freer’ to take part in the interaction. Furthermore, the trend in favor of SCMC may be partially explained by the written mode of text-chat used by the majority of studies in the current sample. Some researchers have argued that oral production is one of the leading causes of learner anxiety (Gregersen & Horwitz, 2002; Hauck & Hurd, 2005; Krashen, 2003), a claim that would suggest computer-mediated environments may be advantageous over FTF interactions.

In classroom-based environments, results indicated that there was a small effect found in favor of SCMC, $d = .18$. This trend suggests that there is a slight advantage for SCMC in foreign and second language classrooms. However, this finding was not significant,
suggesting that there is no significant difference in L2 outcomes between learners participating in FTF or SCMC classroom-based interaction. This result provides encouraging evidence for the use of SCMC in a wide range of classrooms, as the studies included in the sample represent diverse educational settings. For instance, the classroom contexts examined here include a range of target languages in SL and FL settings in both intensive and traditional classrooms. Research has shown that there are important differences between individual types of classrooms (Sheen, 2004), highlighting the potential of interaction in SCMC to promote L2 learning in a variety of classroom contexts. Although additional research directly comparing the efficacy of interaction in SCMC lab and classroom environments would further our understanding of the influence of mode on L2 learning outcomes, these findings provide encouraging evidence that interactional treatments in computer-mediated contexts have a wide range of pedagogical applications.

4.3.3. Interlocutor

This meta-analysis also investigated whether the effects of mode of interaction would be influenced by the type of interlocutor participating in the interaction. For this comparison, studies were categorized according to native speaker status (non-native vs. native speaker) rather than by proficiency. Previous research has shown that NNS-NNS dyads may use more interactional features (Gass & Varonis, 1994; Long & Porter, 1985; Pica et al., 1996; Varonis & Gass, 1985) and provide more modified output opportunities (Sato & Lyster, 2007) when compared to NS-NNS dyads. However, studies have also demonstrated that NS interlocutors may provide more feedback (Mackey et al., 2003;
Oliver, 2002), suggesting that there are empirical differences in learners’ production and development depending on interlocutor type. In the current study, a small effect size was found in favor of SCMC for interactions with native speaker interlocutors, $d = .09$, while a small effect was found for non-native speaker interlocutors in SCMC interactions, $d = .11$. These findings suggest that type of interlocutor accounts for approximately 10% of the variance overall. Differences were not significant, however, and due to the small number of effect sizes for native-speaker interlocutors ($n = 5$), results should be interpreted with caution as the between groups variance may be in error due to the small sample size.

In addition, because previous research has demonstrated variation in feedback provision, uptake, and L2 development across learner-learner and researcher/teacher-learner dyads (e.g. Toth, 2008; Zhao & Bitchener, 2007), this study also sought to examine the influence of these interlocutor differences on the effectiveness of interactional treatments. A small effect size was observed for SCMC interactions in which the researcher or teacher served as the interlocutor, $d = .12$, while a trivial effect was found in favor of FTF interactions in which students interacted with other students, $d = .01$. None of the comparisons were statistically significant, $p > .05$. These results should be interpreted cautiously, however, as there may be errors in the between-study variance due to the small number of effect sizes generated for this contrast. Only 21% ($n = 3$) of the current sample consisted of studies using researcher or teacher interlocutors, with a teacher acting as interlocutor in only one study. Because teacher-learner dyads have been shown to potentially benefit learners through greater feedback provision and increased uptake (Zhao & Bitchener, 2007), this factor warrants further investigation.
within a computer-mediated environment. Based on the studies included in the current sample, findings indicate that few SCMC studies have explored this issue, and due to the potentially important pedagogical implications of such information, researchers should consider examining dyad composition more thoroughly.

Although previous studies have found differences in learners’ performance and production depending on the proficiency of their interlocutors (e.g. Davis, 2009; Kawaguchi & Ma, 2012; Kim & McDonough, 2008; Storch & Aldosari, 2013; Watanabe, 2008), there was not sufficient variation within the current sample to warrant statistical investigation of this variable. For example, intermediate learners were the target population for the majority of the sample (n = 11), with the remaining few studies using beginning learners (n = 3). This relatively homogenous sample of participant proficiency levels highlights the need for more investigations into the effects of interaction in computer-mediated contexts on beginning and advanced learners’ development. Because learner proficiency may influence the efficacy of interaction in computer-environments, additional studies examining these questions are needed to build a body of research large enough to warrant a meta-analytic approach. Overall, these results have identified a number of gaps in the literature based on the data of the current sample, and more data are needed in order to gain a better understanding of the role that type of interlocutor plays in the treatment effects of interaction in SCMC contexts.

4.4. Summary of Results

This study sought to compare the effectiveness of interaction in SCMC with interaction in FTF contexts on the development of L2 learning outcomes. Findings
indicate a small average effect size for SCMC interaction on L2 learning and performance, as measured by immediate posttests. The comparative effects of interaction in FTF and SCMC contexts on learners’ oral and written skills development was also examined, with results showing a small effect for FTF interaction on oral skills and a small effect for SCMC on written skills. Additionally, analyses examined the impact of interaction on learners’ productive and receptive skills, with results indicating a small effect size for FTF on the development of learners’ receptive skills and a small effect size for SCMC on learners’ productive skills. Table 9 provides a summary of the comparative effects.

<table>
<thead>
<tr>
<th></th>
<th>k</th>
<th>d</th>
<th>p</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall L2 outcomes</td>
<td>14</td>
<td>.13</td>
<td>.284</td>
<td>.125</td>
<td>-.11</td>
<td>.38</td>
</tr>
<tr>
<td>Receptive measures</td>
<td>7</td>
<td>-.12</td>
<td>.460</td>
<td>.165</td>
<td>-.45</td>
<td>.20</td>
</tr>
<tr>
<td>Productive measures</td>
<td>14</td>
<td>.15</td>
<td>.230</td>
<td>.124</td>
<td>-.09</td>
<td>.39</td>
</tr>
<tr>
<td>Oral measures</td>
<td>11</td>
<td>-.04</td>
<td>.890</td>
<td>.163</td>
<td>-.30</td>
<td>.34</td>
</tr>
<tr>
<td>Written measures</td>
<td>8</td>
<td>.16</td>
<td>.123</td>
<td>.106</td>
<td>-.04</td>
<td>.37</td>
</tr>
</tbody>
</table>

In order to determine if there were any statistically significant differences due to moderator variables, interlocutor, setting, and context were also examined as possible intervening factors. Small effects were found in favor of SCMC contexts for NNS-NNS and NS-NNS interactions, as well as for interactions between researchers or teachers and students. Student-student interactions, on the other hand, resulted in a small effect in favor of FTF interaction. Analyses found small effects for SCMC interaction in SL and FL contexts, as well as laboratory and classroom settings. Due to the small number of studies available in the current sample, the following factors were not examined for effect
sizes: type of target feature, mode of computer-mediated interaction, type of interactional feedback, and modified output opportunities. Importantly, none of the comparisons resulted in statistically significant results, suggesting that there are no empirical differences in the effectiveness of interaction in promoting L2 development and learning in SCMC when compared to FTF contexts. Table 10 provides a summary of the results examining moderator variables.

**Table 10. Summary of results: Comparative effects of moderator variables**

<table>
<thead>
<tr>
<th></th>
<th>k</th>
<th>d</th>
<th>p</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL*</td>
<td>6</td>
<td>.33</td>
<td>.115</td>
<td>.208</td>
<td>-.08</td>
<td>.74</td>
</tr>
<tr>
<td>FL</td>
<td>11</td>
<td>.08</td>
<td>.147</td>
<td>.592</td>
<td>-.21</td>
<td>.34</td>
</tr>
<tr>
<td>Research context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>4</td>
<td>.06</td>
<td>.614</td>
<td>.116</td>
<td>-.17</td>
<td>.29</td>
</tr>
<tr>
<td>Classroom</td>
<td>10</td>
<td>.18</td>
<td>.182</td>
<td>.138</td>
<td>-.09</td>
<td>.45</td>
</tr>
<tr>
<td>Interlocutor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS*</td>
<td>5</td>
<td>.09</td>
<td>.542</td>
<td>.151</td>
<td>-.20</td>
<td>.39</td>
</tr>
<tr>
<td>NNS</td>
<td>12</td>
<td>.11</td>
<td>.440</td>
<td>.137</td>
<td>-.16</td>
<td>.37</td>
</tr>
</tbody>
</table>

Note: Comparisons marked with an * were calculated using multiple outcome measures due to the small sample size.
Chapter V: Discussion

Previous research has demonstrated the efficacy of face-to-face interaction in promoting L2 development (Keck et al., 2006; Mackey & Goo, 2007), as well as the impact of individual factors, such as type of target feature or focus of corrective feedback, on its efficacy. However, despite the growing body of research examining interaction in synchronous computer-mediated contexts, the effectiveness of interaction in computer-mediated environments was less clear. The current meta-analysis sought to assess not only the overall effectiveness of interaction in SCMC at promoting L2 learning outcomes when compared with FTF interaction, but also examine what factors might impact its efficacy. An additional goal of this analysis was to provide a critical review of the methodological factors in SCMC interaction research during the last few decades, identifying areas for potential improvement and future research. The following section discusses each of the research findings in more detail and, based on the current results, proposes directions for future primary research examining the relationship between interaction, SCMC, and L2 learning.

5.1. Research Question 1: Compared to Face-to-Face Interaction, How Effective is Synchronous Computer-Mediated Interaction at Promoting L2 Learning Outcomes?

The first research question sought to compare the overall effectiveness of interaction in facilitating L2 learning outcomes in FTF and SCMC modes. The large average effect sizes calculated from the immediate posttests of the fourteen studies included in the
sample reveal that interaction in both contexts plays a beneficial role in promoting L2 learning outcomes, with SCMC contexts resulting in a large effect size of 1.13 and FTF contexts resulting in a large effect size of .84. Although the average effect size for computer-mediated contexts was slightly larger than in FTF contexts, with results indicating a small effect size in favor of SCMC, no statistically significant differences were found between the two modes. These findings, which suggest that there is no difference between the effects of interaction in computer-mediated contexts and face-to-face contexts on measures of L2 performance and development, may be due to a number of factors.

For instance, previous research has provided evidence that negotiation patterns and interactional features found in face-to-face interaction occur in SCMC (Blake, 2000; Kötter, 2003; Pellettieri, 2000; Smith, 2003; Toyoda & Harrison, 2002). SCMC has also been found to provide opportunities for learners to negotiate for meaning, as well as give and receive corrective feedback (Fernández-Garcia & Martínez-Arbeláiz, 2003; Lai & Zhao, 2008). These studies demonstrate that interaction in computer-mediated modes provides similar opportunities as FTF interaction for negotiation for meaning and its benefits for L2 learning, providing learners with an environment facilitative of L2 development. Considering that empirical evidence seems to suggest that interaction in electronic environments provides learners with similar developmental opportunities as in face-to-face contexts (e.g. Abrams, 2003; Beauvois, 1992; de la Fuente, 2003; Kern, 1992; Kelm, 1995; Lee, 2001; Iwasaki & Oliver, 2003; Kötter, 2003; Tudini, 2003; Toyoda & Harrison, 2002; Pellettieri, 2000; Salaberry, 2000; Sauro, 2009, 2012; Shekary

In addition, despite demonstrated differences in interactional patterns between SCMC and FTF modes, such as variation in turn-taking strategies and the existence of different discourse features (Toyoda & Harrison, 2002), studies examining direct connections between interaction in SCMC contexts and learning have found results comparable to those in FTF research. For instance, Smith (2004) found that NNS-NNS dyads participating in task-based interaction in a SCMC environment experienced significant gains on vocabulary test, while Sauro’s (2009) results show that corrective feedback delivered during SCMC interactions led to improvement in learners’ grammatical competence. These results are similar to those found in traditional face-to-face contexts, such as de la Fuente’s (2002) and Ellis et al.’s (1994) study of the impact of negotiated interaction on lexical acquisition and Mackey and Oliver’s (2002) investigation of corrective feedback on L2 morphosyntactic development. Overall, these studies provide evidence that SCMC interaction leads to results similar to those found in FTF modes on lexical and grammatical items, as each mode provides learners with opportunities to negotiate for meaning as well as provide and receive feedback, providing further explanation for the results of the current analysis.

Overall, direct comparisons of L2 learning in FTF to SCMC contexts have led to differential results across communication mode, with some studies finding advantages for FTF (e.g. Baralt, 2010; de la Fuente, 2003; Petersen, 2010) and others for SCMC (e.g. Bueno-Alastuey, 2011; Payne & Whitney, 2002), making it difficult to make any conclusive findings. The present findings show that despite the potential advantages
unique to each form of communication, such as the additional visual input of gestures and facial expressions in FTF and additional online planning time in text-based SCMC, there is no significant difference in the current sample in overall learning outcomes due to communication mode, suggesting that there are no significant differences between computer-mediated contexts and face-to-face interaction on overall L2 development. These findings have important pedagogical implications, as they support the use of SCMC interactive tasks in distance learning contexts as well as in traditional classrooms as a way to improve learners’ L2 development.

However, an important caveat to these findings is the fact that the average effect sizes were calculated for immediate posttests only. Only three studies in the current sample reported delayed posttest scores (Baralt, 2010; de la Fuente, 2003; Yanguas, 2012), making any comparisons statistically unreliable. The lack of a delayed posttest contrast limits the generalizability of the findings, particularly as research has suggested theoretical grounds (Gass, 1997; Lightbown, 1998) and empirically demonstrated that interactional features may have delayed effects (e.g. Mackey, 1999; Muranoi, 2000; Sheen, 2007; Van den Branden, 1997). For instance, in one of the first studies to note this delayed effect, Gass & Varonis (1994) found that learners’ comprehension improved on delayed measures, suggesting that the impact of the interaction was not immediate. Mackey (1999) found significant increases in question production on delayed posttests, suggesting that the effects of interactional feedback such as recasts may be more evident over time. Mackey and Philp (1998) also found that recasts may have a delayed effect, with their results showing improvement in learners’ question forms on delayed posttest measures. Muranoi’s (2002) study of Japanese EFL learners found improved accuracy on
delayed measures for learners who had participated in interaction, while Takashima & Ellis (1999) and Carroll (2001) found that learners receiving corrective feedback demonstrated greater gains than learners who did not receive corrective feedback on both immediate and delayed measures, suggesting that the developmental effects were sustained. The results of Ellis, Loewen, & Erlam’s (2006) study on the differential effects of corrective feedback on learners’ acquisition of past tense ‘-ed’ indicate that the effects of recasts appear to endure over time. Leeman’s (2003) investigation of Spanish foreign language learners showed that learners receiving recasts maintained improved accuracy from immediate to delayed posttests, while Sheen’s (2007) results indicate sustained improvement from immediate to delayed posttest measures, providing additional support for the enduring effects of interactional feedback. In addition, in his recent meta-analysis, Li (2010) found that recasts had a greater effect on delayed tests than on immediate and short-term delayed posttests, providing conclusive findings that the developmental effects of recasts may be delayed, but are sustained over time. Clarification requests also resulted in medium effect sizes on long-term delayed posttests, providing further evidence for the enduring influence of interactional feedback on L2 performance and production.

These findings are similar to those of Mackey and Goo (2007), who found a larger effect for overall interactional feedback on delayed posttests, indicating an increased effect for feedback over time. When examining specific types of feedback, recasts resulted in large average effect sizes for immediate, short-term, and long-term delayed posttests, providing further evidence for the efficacy of recasts in sustaining L2 development. These findings clearly demonstrate the long-term and lasting impact of interaction on L2 learning outcomes in predominately FTF settings (although both Li,
2010, and Mackey & Goo, 2007, include a small number of studies in computer-mediated contexts). The absence of delayed posttests in the current sample highlights the need for more research investigating the long-term effects and potential durability of interactional benefits in SCMC. Although results for the immediate effects on interaction in SCMC are promising, more longitudinal research is needed to enhance our knowledge of the role of interaction on learners’ long-term learning.

5.2. Research Question 2: Compared to Face-to-Face Interaction, How Effective is Synchronous Computer-Mediated Interaction at Promoting Receptive and Productive L2 Learning Outcomes?

While results revealed large effects for receptive skills in both SCMC ($d = 1.31$) and FTF ($d = 1.44$), medium effect sizes were found for both SCMC ($d = .82$) and FTF ($d = .71$) modes for productive learning outcomes. The higher effect sizes for receptive skills are not unexpected, as it is relatively well accepted in SLA that learners are more likely to recognize linguistic forms than to produce them. According to the computational model of L2 processing (Ellis, 1997), learners must first attend to the input, reviewing it in short-term memory. The target feature is then integrated into long-term working memory, and accommodated (VanPatten, 1996). In other words, the learner’s interlanguage (IL) system adapts to accommodate the new feature. After the new feature has been incorporated in the learner’s IL, it is possible to produce the form through output. Empirical research seems to support the claim that receptive knowledge may be more readily acquired, as previous studies have demonstrated that learners often perform better on immediate measures of receptive learning than on productive learning (e.g. de la
Fuente, 2003; Webb, 2008; Yanguas, 2012; Zheng, 2012). In addition, studies show that comprehension based (CBI) and production based instruction (PBI) result in differential L2 development, with advantages for CBI in short-term receptive knowledge and PBI in sustained productive knowledge (see Shintani et al., 2013, for a recent meta-analytic review).

Although interaction provides learners with both comprehension and production opportunities through the provision of comprehensible input and modified output, thereby offering learners opportunities to build their receptive and productive knowledge, the overall findings of larger effect sizes for receptive knowledge would seem to support previous claims that receptive knowledge is acquired before productive. These findings may also be explained by Swain’s influential Output Hypothesis (1995), which posits that output is a necessary condition of SLA. Inspired by immersion programs where learners received rich comprehensible input but yet lacked high proficiency levels in grammatical and sociolinguistic competence, Swain proposed that learners need opportunities to talk and produce output. Production, particularly instances where learners were ‘pushed’ to produce output, was posited to aid learners in moving from semantic top-down processing to syntactic bottom-up processing. Although comprehension can occur with little syntactic analysis on the part of the learner, production requires the learner attend to the form and meaning of the utterance, ‘pushing’ themselves to produce socially appropriate and linguistically understandable output. In other words, comprehension requires less effort and processing than production, or as Swain states, “learners…can fake it, so to speak, in comprehension, but they cannot do so in the same way in production” (1995, p.127). When receiving input, learners may focus more on meaning,
leading to semantic processing. However, in order to produce language, learners must attend in some manner to the linguistic form they are attempting to utter to maximize their comprehensibility, thereby experiencing syntactic processing on some level. Other researchers have also argued that production requires deeper levels of processing than comprehension (Izumi, 2002). Whereas learners can attain a reasonable level of comprehension without necessarily focusing on any formal features, successful production requires learners to conceptualize, formulate, articulate, and monitor (Izumi, 2003). Output processing engages learners’ internal procedures of grammatical encoding and monitoring, promoting a deeper engagement of learner internal factors, such as selective attention, with environmental features, such as input and interaction.

In light of these differences, the larger effect sizes for receptive measures on immediate posttests in both FTF and SCMC might be attributed to the shallower semantic processing and attention to meaning associated with comprehension. The superficial processing of input during comprehension is less likely to lead to sustained learning of the target structure, while the deeper levels of processing in production, on the other hand, push learners to perceive the relationship between target structure and meaning through grammatical encoding. Production, then, encourages learners to attend to the forms as well as integrate them into the IL, providing a fertile environment for learners to notice and compare any mismatches between their language and the target language, thereby facilitating L2 development. Overall, measures requiring learners to produce rather than simply recognize the target forms are more likely to lead to intake, integration, and later output.
However, because the current analysis only includes immediate posttest measures, it is unclear how durable these effects for receptive and productive learning might be. For example, learners may have simply attended to the target forms, supporting their aural or visual recognition on immediate posttests, but failed to process the forms at a deeper level, thereby integrating them into their IL systems. This lack of processing might result in lower scores on later measures for receptive measures, while the more integrative processing associated with production (Izumi, 2002) might manifest in increased scores on delayed posttest measures. In their recent meta-analysis, Shintani et al. (2013) found that production resulted in more durable effects than comprehension, with results indicating a significant advantage for production-based instruction on delayed measures, but no significant differences between CBI and PBI on immediate posttests. Although CBI appeared to be more effective for developing receptive skills on immediate measures, a finding consistent with the results of the current research, no significant differences were found for delayed measures. In addition, interaction research in FTF contexts has resulted in similar findings, with delayed effects of interactional feedback found in a number of empirical studies (Mackey, 1999; Mackey & Philp, 1998; Mackey & Silver, 2005) and meta-analyses (Mackey & Goo, 2007). These differences between receptive and productive measures may also have been a result of whether the forms were new or partially acquired. Newly acquired forms may not surface in immediate measures, instead needing time to be sufficiently processed and integrated into learners’ IL systems. These newly processed target features may then result in an increase on later measures, a finding also in line with Pienemann’s Teachability Hypothesis (1985) and his more recent Processability Theory (1998). However, these trends were not
empirically assessed, as there were not sufficient data in the current sample to investigate long-term effects as measured by delayed posttests. In general, research in FTF contexts has demonstrated delayed effects depending on processing levels and corrective feedback, stressing the need for more longitudinal research in SCMC to test whether these trends remain consistent across communication mode.

Although large and medium effects were found for both FTF and SCMC modes for receptive and productive outcomes, small effects were found in favor of SCMC for productive measures and for FTF for receptive measures. No significant differences were found between the two communication modes for productive or receptive skills, suggesting that there are no substantial differences between the two modes on measures of receptive and productive learning. However, the trends in favor of SCMC on productive measures might be explained by several factors. For instance, production may require greater levels of processing than comprehension (Izumi, 2002, 2003; Swain, 1995). Because computer-mediated contexts have been argued to provide additional time for processing and planning (Beauvois, 1992; Pellettieri, 1999, 2000; Smith, 2004; Smith & Gorsuch, 2008; Warschauer, 1997), the small effect size in favor of SCMC may be due to this potential advantage in computer-mediated interaction. This is particularly true of text-chat exchanges, as learners are presented with a written record of the interaction, which may provide added opportunities for learners to attend more closely to the form and content of the input, while still maintaining the real-time feel of conversation (Pellettieri, 2000; Smith, 2003; Toyoda & Harrison, 2002).

These additional opportunities for noticing provide learners with more developmental support, particularly as the computational model of L2 processing (Ellis, 1997) requires
that learners must first attend, or notice, the input. Schmidt’s Noticing Hypothesis (Schmidt, 2001; Schmidt & Frota, 1986) also emphasizes the importance of noticing, specifically noticing of the formal features in the target language, as well as noticing gaps between a learner’s IL and the TL. According to this hypothesis, noticing is a necessary condition of SLA, underscoring its importance for L2 development. Because SCMC may increase learners’ opportunities to notice target items in the input, as well as notice gaps between their IL and the target language, this mode may be more facilitative of certain target forms than FTF interaction. For example, researchers have argued that text-chat provides learners opportunities to reflect on the discourse (Beauvois, 1992), to notice new target items as well as the gaps in their IL (Kelm, 1992), and enhance incidental noticing (Warschauer, 1997). Payne & Whitney (2002) found that learners reported noticing their mistakes more frequently in SCMC chat environments than in FTF interaction. Lai and Zhao (2006), although they did not compare FTF and SCMC interactions, found that text-chat resulted in learners’ improved noticing of errors and interactional feedback, providing further evidence of the possible benefits in SCMC to enhance noticing.

In addition, SCMC may be more effective at drawing learners’ attention to formal features within the input. For instance, Payne (1999) found that 50% of students said they focused more on form in text-chat than in FTF, suggesting that text-based SCMC provides learners with opportunities for increased monitoring and noticing of their own and their interlocutor’s production. Empirical research has also demonstrated that SCMC may encourage learners to focus on form (Blake, 2000; Salaberry, 2000; Shekary & Tahririan, 2006), as well as potentially enhance negotiation for form and meaning (Lee, 2002). The potential for SCMC to provide additional opportunities for learners to attend
to the input and notice the gap between their IL and the target language may lead to increased instances of intake, which according to the computational model discussed above, is a prerequisite for integration into the learners’ IL. In other words, interaction in SCMC may offer a small advantage over FTF interaction in promoting L2 learning, particularly on productive measures, by providing learners additional time to notice and process the target forms.

Although not statistically significant, a small advantage was found for FTF contexts over SCMC for the receptive language skills of listening and reading. This result might be due to the features of FTF interaction that provide learners with multiple opportunities to attend to the input. For instance, FTF interaction provides visual, verbal, and gestural cues, offering learners a variety of input sources to aid in the processing of spoken language. Learners in FTF interaction are exposed to a range of input, thereby potentially supporting the later recognition of the target forms. However, it is important to remember that only seven effect sizes were included in the contrast for receptive skills, suggesting that future research should consider dedicating more attention to the effects of SCMC on the development of learners’ receptive skills. Overall, the differences between productive and receptive learning across modes seem to suggest that these processes may be different, and learners may benefit from approaches designed to target individual abilities. Although the analyses did not indicate any significant advantage for one communication mode over the other, more studies are needed in order to tease apart the influence of interaction in SCMC on immediate and long-term L2 development.
5.3. Research Question 3: Compared to Face-to-Face Interaction, How Effective is Synchronous Computer-Mediated Interaction at Promoting Oral and Written L2 Learning Outcomes?

Research in CALL has demonstrated promising results regarding the effectiveness of SCMC in promoting both oral and written L2 development (Chun, 1994; Kost, 2004; Payne & Whitney, 2002; Sullivan & Pratt, 1996), and the results of the current analysis indicate large effect sizes for SCMC on oral ($d = 1.04$) and written ($d = .91$) measures. Face-to-face contexts also resulted in large and medium effect sizes for oral ($d = 1.24$) and written ($d = .38$) skills, indicating that positive benefits are found in both modes of interaction. Comparing the two contexts, SCMC resulted in a small effect size for written outcomes while a small effect was found for oral outcomes in FTF contexts, although the difference between the two modes was not statistically significant.

Although small, the average effect size in favor of interaction in FTF environments over SCMC contexts for the development of L2 speaking and listening skills is not entirely unexpected, particularly as many of the outcome measures used in the sampled studies were oral prompted production. Because the majority of studies in the current sample favored oral production tasks for assessment, learners participating in FTF interaction may have had a slight advantage over those in SCMC contexts, particularly written SCMC environments, due to the similarity between the treatment and testing tasks. Likewise, the slight advantage for interaction in SCMC for the development of learners’ written skills suggests that learners may benefit when the modality of treatment and assessment are the same. Another possible explanation for the advantage of FTF contexts for the development of learners’ oral skills is gesture. Available only in FTF or
video SCMC, of which there were only two studies included in the current analysis, gestures provide learners with important communicative information. For instance, early studies found that gestures may be used to supplement gaps in learners’ vocabulary knowledge (Bialystok, 1990; Faerch & Kasper, 1983). More recent work reveals that gestures are not restricted to a supplementary role, but instead complement speech and interaction. For example, Gullberg’s (1998) results indicated that learners used gesture to inform their interlocutors that they were searching for a word or needed help with an expression, providing information important to the success of the interaction. Research has also shown that gesture directly impacts L2 development, with studies demonstrating improved listening comprehension when learners were able to view video showing the speaker’s facial gestures in addition to hearing an audio track (Kellerman, 1992; Sueyoshi & Hardison, 2005). These promising findings suggest that gesture plays an important role in interaction, and may help to explain the trend in favor of FTF contexts for the development of oral skills. The role of gesture in L2 development also highlights the potential learning benefits of video SCMC, which has been shown to produce similar results to FTF interaction in both receptive and productive measures (Yanguas, 2010, 2012). Based on the data presented in the current sample, these findings underscore the need for more research into this relatively underexplored domain.

The current trends indicating an advantage for SCMC for written skills and FTF for oral skills seem to be consistent with claims that practice and learning are skills specific (DeKeyser, 1997, 2007), as learners seemed to experience better performance on posttest measures when the mode of treatment and assessment were the same. For example, DeKeyser (1997) examined the learning of four grammar rules, with variations in rules,
skills, and extent of practice across groups. For instance, some groups participated in comprehension-based practice while others completed production-based practice. After the treatment period, when learners were tested on a skill they had not practiced for their learned rule, their performance was markedly worse than their performance on the practiced skill. Participants who had practiced all the rules in both comprehension and production based skills performed equally well on all measures as those learners that had practiced only one rule in one domain. These data provide empirical evidence for skill-specific learning, suggesting that learners who acquire forms or rules in one skill will perform better in that skill than others. In other words, learners acquiring forms within an oral or written context will perform best on measures of that form that match the mode in which they acquired it. De Jong (2005) also found evidence for skill specificity; with results showing that extensive aural comprehension training led to increased processing speeds in comprehension but did not impact accuracy in production. De la Fuente’s (2003) results also illustrate these claims, as learners participating in written SCMC outperformed learners participating in FTF interaction on written measures of production and recognition, while learners in FTF modes outperformed the SCMC groups on oral measures. The observed trends in the current data are consistent with these predictions, although not all research has yielded similar results. For example, some researchers have found that the developmental benefits of interaction in SCMC are not confined to skill specific learning, with findings showing that written SCMC may gradually (Chun, 1994) or directly transfer from written to oral modalities (Beauvois, 1992; Kern, 1995; Payne & Whitney, 2002). Although both modes are interactive in nature, these conflicting results
seem to suggest that there are certain features unique to each mode that make direct comparisons difficult.

Although some researchers have pointed out the differences between FTF and SCMC interaction, such as in structure or turn-taking patterns (Smith, 2004; Toyoda & Harrison, 2002), other researchers have argued that video SCMC provides similar social signals as FTF interaction, as learners are still able to receive and produce visual and situational cues (Lee, 2007). Similarities in performance have been found for learners in video SCMC and FTF contexts, although these similarities do not seem to extend to audio SCMC interaction. For instance, Yanguas (2010) found differences in negotiation patterns between learners participating in audio interaction with learners taking part in FTF or video interaction, a difference attributed to the lack of visual input in the audio SCMC group. Learners in the audio group seemed to utilize other linguistic resources when deprived of the visual cues in video SCMC and FTF contexts. Results also indicated that overall interactional patterns in oral SCMC were more similar to those found in FTF contexts than those of written computer-mediated communication, suggesting that there are important implications of restricting learners’ input to aural or literal forms.

More recent research further supports these results, finding no significant differences between oral SCMC and FTF on written measures. However, advantages for audio SCMC contexts over video SCMC or FTF contexts were found on listening comprehension measures (Yanguas, 2012), suggesting that there may be positive benefits to audio conferencing for oral measures over other forms of SCMC. Mayer’s cognitive theory of multimedia learning (Mayer, 2001, 2005; Mayer & Moreno, 2003) may provide
an explanation for these results. Mayer’s theory proposes that visual and auditory information is processed through two individual channels, with each channel having a limited intake capacity. Each channel actively filters, selects, organizes, and integrates information in the input, suggesting that when one channel is inactive, resources are not dispersed over the two channels but are focused to one. Processing only audio information, rather than processing visual and audio input simultaneously through both channels, may lead to improved performance on aural receptive and productive measures.

These findings might also be explained by the lack of visual contact or gestural input typical of FTF interaction. For instance, the absence of multi-modal input might allow learners to concentrate more on the mode of interaction, a feature that may enhance the salience of the input (Yanguas, 2010). In addition, written SCMC may require more verbal explicitness due to the lack of visual contact, thereby drawing learners’ attention to form in the input and their subsequent output. This enhanced explicitness may also apply to learners’ use and provision of corrective feedback. Recent meta-analyses have indicated that explicit forms of corrective feedback, such as metalinguistic feedback, result in greater short-term gains than implicit forms of feedback (e.g. Li, 2010).

However, as the current analysis addresses immediate posttests only, the short-term effects in written outcomes found for SCMC may be due to the enhanced explicitness of written forms of SCMC.

Based on the data reported in the current study, there is very little research examining the differential effects of video and audio modes of SCMC on learners’ development, highlighting the need for more research in order to tease apart the nuances of multi-modal learning and skills development in SCMC interactions. Within the current sample, the
majority of studies used text-chat as the mode of interaction, which may have contributed to the small effect in favor of SCMC for written outcomes. However, more research is needed to determine the effectiveness of aural and visual forms of SCMC on learners’ written and oral production and recognition. Thus far, the body of research examining oral modes of SCMC seems to suggest that there are no significant differences between visual SCMC and FTF contexts (Yanguas, 2010, 2012), a finding clearly in need of further examination. As this important area of research continues to grow, future meta-analytic work may find positive benefits for computer-mediated contexts on the development of learners’ speaking and listening skills.

Overall, although small effect sizes were noted for both FTF and SCMC contexts, neither advantage was statistically significant, suggesting that there is no clear benefit for one mode or the other on the development of learners’ oral or written skills. However, the studies in the current sample were mainly conducted in the written SCMC mode, potentially producing a bias for written measures in the results. Based on the data in the current sample, results demonstrate that the efficacy of different SCMC modes for the development of oral and written skills remains murky, highlighting the gaps in the literature needing further investigation and underscoring the need to accumulate more studies examining the impact of different modes of SCMC on a variety of L2 skills, providing the data necessary in order to make more conclusive interpretations regarding specific skills development, especially the efficacy of cross-modality interaction treatments and assessments.
5.4. Research Question 4a: How Does the Nature of the Research Setting (Foreign Language vs. Second Language) Impact the Effectiveness of Interaction in SCMC?

Previous meta-analyses have demonstrated that foreign and second language setting act as influencing factors on the efficacy of interaction, with research in foreign language producing larger effect sizes for interaction than research in second language setting (Mackey & Goo, 2007). The current analysis provides further support for these findings, with within-groups effect sizes for SCMC and FTF in FL settings significantly larger than those in SL setting. Comparisons between FTF and SCMC revealed a small effect for SCMC \( d = .08 \) for FL settings and a small effect for SCMC \( d = .33 \) in SL settings. Although small, this positive trend in favor of SCMC may reflect some of the advantages researchers have attributed to computer-mediated contexts, such as its ability to reduce anxiety. Within SL settings, anxiety may have a greater effect on learner-learner interaction, as SL programs are more likely to consist of a student body with diverse L1s and cultural backgrounds. For example, research has demonstrated that learners from different cultural backgrounds may be hesitant to provide feedback because they feel their interlocutors may find it socially inappropriate (Mackey et al., 2003). For these learners, an SCMC environment, particularly text-chat, may be more facilitative of interaction due to the potential reduction in face-threatening acts. In foreign language environments, where learners may be more likely to be from a shared L1 and cultural backgrounds, cultural factors may play less of a role, explaining the lack of significant differences in the efficacy of SCMC and FTF interaction on L2 learning outcomes.

Furthermore, due to the additional processing and planning time in computer-mediated interaction, which provides learners with opportunities to reflect on what was
said before responding (Beauvois, 1992), SCMC interaction may place lower social
demands on learners, thereby reducing their levels of anxiety (Baralt and Gurzynski-Weiss, 2011). More introverted learners may feel more comfortable participating in
electronic conversations with their peers when some of the social pressures of face-to-
face communication are removed, allowing them to produce and participate more equally
in interactions (Kern, 1995; Warschauer, 1996). Together, these factors help to illuminate
the possible explanations for the small advantages noted for SCMC interaction. However,
these interpretations are speculative, particularly regarding SL settings, as most of the
studies meta-analyzed here were conducted in foreign language settings (11 out of 14, or
79%).

In order to gain a better understanding of the role setting may play in learner
development, researchers should consider pursuing empirical examinations of the role of
anxiety across mode (SCMC and FTF) and educational environment (SL and FL).
Although anxiety has been identified as one of the more influential factors on L2 learning
and performance (Horwitz, 2001), the majority of studies in SCMC rely on qualitative,
anecdotal evidence, providing only a partial picture of the role of this important
construct. Anxiety measures frequently rely solely on learner self-report questionnaires,
with very little quantitative, empirical research conducted on the role of anxiety in
computer-mediated interaction (although see Baralt & Gurzynski-Weiss, 2011). In order
to more objectively assess learners’ anxiety, as well as build a repository of numerical
results that researchers can use for future meta-analytic work, quantitative instruments
should be implemented in addition to perception measures. The use of quantitative
measures, such as surveys with Likert scale type questions, will generate descriptive
measures and facilitate correlational analyses between learner performance, development, and anxiety levels, providing important information regarding the effects of anxiety across mode and setting. Currently, very few studies empirically examine anxiety (e.g. Baralt & Gurzynski-Weiss, 2011) in SCMC within a FL setting, emphasizing the need for more research on the role of anxiety in both SL and FL settings. Through the collection of empirical, quantitative data on the role of anxiety in SCMC, researchers will be able to further our understanding of whether computer-mediated contexts reduce or heighten learners’ anxiety during interactions.

Overall, based on the data and findings of the current analysis, more research examining interaction in SL settings is needed in order to more accurately assess the impact of setting on the differential efficacy of FTF and SCMC interaction. One way that researchers might contribute to the existing body of work investigating computer-mediated interaction within SL settings is to collaborate with other SL programs within the same target country. Many universities have departments dedicated to second and foreign language instruction, however, researchers may find their studies to be limited by the number of available participants as individual programs may have a limited number of students from a specific proficiency level or with a particular L1. However, by using SCMC, researchers would be able to expand their recruitment opportunities through the addition of one or more virtual classes, potentially increasing the size and diversity of their participant population. Instead of forming dyads consisting of learners within the same classroom, researchers collaborating across similar institutions would be able to create dyads consisting of students from both locations, thereby increasing the sample size of the study. For some learners, computer-mediated interaction within a classroom
may feel artificial, as they may be able to visually see their interlocutor across the room even though they might be interacting digitally. For these learners, working with off-site interlocutors may help increase the ecological validity of the experiment. In addition, the potential increase in sample size is likely to lead to greater a priori statistical power (a topic addressed in greater detail in Chapter VI), increasing the chances that a statistical difference will be discovered if it exists.

Study-abroad (SA) contexts also provide researchers with another area in which to examine the relationship between interaction and SL settings. According to the US Department of Education (2013), over 80,000 students study abroad each year, with numbers increasing by two percent annually. Of this number, 70% are associated with formal programs sponsored by US universities or colleges, providing an educational environment similar to their at home institution. Although SLA research in study-abroad is growing (Lafford, 2006), there is little research focusing on interaction within this SL setting. Although SA is an area that both FTF and SCMC researchers could exploit, SCMC offers the unique opportunity to pair FL and SL learners together, a composition that would not be logistically possible in FTF contexts. Lastly, researchers could empirically compare interaction in SL and FL contexts. Previous research has noted the complex relationship between setting and L2 development (Segalowitz & Freed, 2005), suggesting that although there may be advantages for one context over the other on certain features, the relationship between the individual learner, the setting, and the overall learning outcomes is complex. In order to gain a better understanding of the influence setting might have on L2 development, researchers should consider directly comparing the efficacy of interaction in SL and FL settings in SCMC and FTF contexts.
5.5. Research Question 4b: How Does the Type of Research Setting (Classroom vs. Laboratory) Impact the Effectiveness of Interaction in SCMC?

Some researchers have argued that interaction may not be as effective in classroom contexts as in laboratory contexts, arguing that ‘uncoached negotiation for meaning’ may not occur naturally in an authentic classroom (Foster, 1998). Although Foster did not find evidence of negotiation in her study, her role during data collection was more akin to researcher than teacher, comprising the validity of the classroom context. Nicholas, Lightbown, and Spada (2001) have also pointed out that the benefits attributed to interaction in lab contexts may be caused more by the intense, focused, and consistent application of the treatment rather than the treatment itself. Although lab-based environments may present a variety of advantages due to their controlled and calculated nature, previous research has indicated that interaction is also facilitative of L2 development in classroom environments (Gass et al., 2005; Mackey & Goo, 2007; Russell & Spada, 2006). In order to examine the efficacy of SCMC interaction in both research settings, comparisons examining laboratory and classroom settings were conducted. Average effect size was larger for SCMC in laboratory and classroom contexts, although neither comparison was statistically significant. Although these results mirror those found in FTF contexts (Mackey & Goo, 2007), the minor advantages for SCMC may be due to a number of factors. For example, interaction in laboratory settings may be more controlled, providing increased opportunities for consistent feedback and fewer distractions from target tasks than in classroom-based contexts. The heightened control and consistency of lab-based treatments may offer learners greater opportunities
to attend to input and focus on form, potentially leading to greater performance and scores on measures of learning outcomes. Although the studies conducted in classroom environments have sought to control for a variety of intervening variables, the classroom environment is unpredictable due to factors often beyond researchers’ control, such as insufficient numbers of computers for participants or disruptive students. Classroom-based SCMC, although still subject to some of the same distractions as FTF classrooms, may be more controlled due to the reduction of social factors found in oral interaction.

SCMC contexts, particularly research designs using text-chat as the mode of treatment, might further reduce distraction and improve learner focus by removing audio or gestural input. Because learners are interacting on a purely textual level, their focus may be more frequently drawn to the feedback or target features, possibly providing more opportunities to focus on form and meaning in absence of other forms of input. Payne and Whitney (2002) found that learners reported attending more closely to linguistic forms and corrective feedback during SCMC interactions, suggesting that the mode of interaction may have contributed to the small effect size found in favor of computer-mediated contexts.

5.6. Research Question 4c: How Does the Type of Interlocutor Impact the Effectiveness of Interaction in SCMC?

In light of research demonstrating that the type of interlocutor may impact the effectiveness of interaction in SCMC and FTF contexts, the influence of interlocutor status on L2 learning outcomes was investigated. Although not statistically significant, a small effect was found in favor of SCMC for interactions involving the researcher as
interlocutor. In addition, small effects for SCMC was found for native speaker interlocutors and non-native speaker interlocutors, while trivial effects were found for FTF interactions in which students interacted with other students. These differences across mode may be due to a variety of factors. For instance, the small effects associated with SCMC interactions in which the researcher or teacher serves as the interlocutor may be caused by the perceived power differential between the interlocutors. Robinson and Gilabert (2007) point out that type of interlocutor impacts the interactivity demands of the task, and student-teacher dyads may be affected by the power relationship between participants. In addition, researcher or teacher interlocutors may provide more consistent and greater amounts of corrective feedback, impacting the interactivity and therefore the possible learning outcomes of the task. Furthermore, researcher responses to learner errors, including the provision of corrective feedback and modified output opportunities, may be more scripted in SCMC contexts, thereby further increasing the consistency of the interaction. For instance, although interactions cannot be scripted word-for-word, researchers are likely to have guidelines to ensure the validity of the procedure and treatment. Text-based environments, such as those most frequently used in SCMC contexts, may promote the more careful following of such guidelines, as researchers are afforded the same additional processing and planning time noted for computer-mediated contexts in previous research.

The computer-mediated text-chat environment, which was the mode for the majority of studies in the current sample, also eliminates variables, such as rate of speech or volume, which might impact learners’ comprehension of input. Instead, learners are able to focus on the visual input. Although FTF interaction also includes more information,
such as intonation and stress, these features are more likely to affect the communication of meaning rather than the comprehension of form. The lack of paralinguistic and nonverbal information in SCMC, particularly for text-chat environments, may provide learners with more opportunities to focus on language and form. Indeed, previous research has demonstrated that learners in SCMC report focusing on form more than in FTF contexts (Payne & Whitney, 2002), as well as increased levels of noticing due to the ability to review chat logs or transcripts (Lai & Zhao, 2006; Shekary & Tahirian, 2006), suggesting that learners in SCMC may have had more opportunities to notice new features as well as the gap between their IL and the TL (Kelm, 1992).

Because noticing is a critical aspect of SLA (Schmidt, 2001), these increased opportunities in computer-mediated contexts have important implications for learners’ development. For example, Schmidt’s Noticing Hypothesis (Schmidt, 1990; Schmidt & Frota, 1986) posits that the conscious noticing of mismatches between a learners’ IL and the TL is a necessary condition of successful SLA, with empirical studies demonstrating a direct link between the noticing of a target feature and its’ subsequent intake (Izumi, 2002; Izumi, Bigelow, Fujiwara, & Fearnrow, 1999; Leow, 2001; Mackey et al., 2000). If computer-mediated contexts promote the noticing of forms and gaps between learners’ IL and TL, they may be more facilitative of L2 development due to the increased number of noticing opportunities. Indeed, a number of SCMC researchers have argued that this is a distinct advantage for computer-mediated contexts. For instance, Pellettieri’s (2000) results indicated that text-chat promoted learners’ use of negotiation of meaning and form-focused interaction, leading her to argue that the additional time for monitoring and processing in SCMC supported learners’ noticing of gaps, thereby facilitating their L2
development. The unique characteristics of SCMC, specifically within the written mode of text-chat, provide learners with visual saliency, chances to review the output of both interlocutors, and longer processing times (Smith, 2004), increasing the opportunities learners have to focus on form and providing improved chances for L2 development (Salaberry, 2000). In their study examining the relationship between learners’ verbalizations, gestures, and scrolling, Smith and Gorsuch (2004) found an increased attention to form, while Smith’s (2010) study used eye-tracking technology to examine the relationship between recasts, noticing, and uptake, providing tangible evidence of what learners’ attended to in the input and feedback.

In addition, empirical evidence seems to suggest that learners do in fact take advantage of the additional processing time afforded them in SCMC. For example, in their study of learners of German, Sauro and Smith (2010) found that learners use online planning time to monitor their production and that of their interlocutors, resulting in more careful production and complex language. Overall, the small effects for interaction in SCMC may be attributed to the additional processing time of SCMC, which may promote increased noticing and focus on form, potentially leading to increased opportunities for developmental progress.

Previous research has also indicated that NS-NNS dyads may experience more anxiety (Lee, 2004; Satar & Ozdener, 2008) as well as fewer instances of negotiation, pushed output, or correction (Gass & Varonis, 1985; Iwashita, 2003). For example, Lee (2004) found that NNSs with low proficiency levels reported that they did not feel comfortable interacting with NSs, with one student even remarking that she did not want to interact with other NSs in the future due to her levels of anxiety and discomfort. The
small effect for SCMC found in the current research may be due to the potentially reduced anxiety attributed to SCMC interactions, particularly during text-chat (Satar & Ozdener, 2008). According to Beauvois (1997), computer-mediated contexts provide ‘an anonymous, less pressured environment that tends to lower the affective filter’ (p.171), potentially providing anxious or introverted students with additional interactional opportunities. In addition, NS interlocutors provide learners with the best-possible language input. SCMC, by offering a less stressful environment than FTF contexts, may provide shy learners with opportunities to receive this input by counteracting some of the disadvantages and anxiety that learners perceive in NS interactions. Overall, for less proficient or more anxious learners interacting with NSs, SCMC contexts, particularly text chat, may provide a more comfortable and safer environment in which learners can feel free to make mistakes and test their hypotheses about the target language. However, research examining the effects of NSs-NNSs dyad composition on learners’ production and performance remains limited. Researchers should consider conducting further experiments examining NS-NNS pairs, in conjunction with measures of anxiety and willingness to communication, in order to gain a better understanding the role of interlocutor plays in L2 development.

In addition, although some studies have found greater amounts of corrective feedback in NS-NNS dyads when considering learners’ non-target-like utterances (Mackey et al., 2003), others have suggested that NNS-NNS interactions result in more negotiation for meaning (Sauro, 2001; Sotillo, 2005). The current findings show larger effects for NNS interlocutors over NS interlocutors in both modes, suggesting that interacting with NSs may impact learners’ development. However, when comparing the impact of interlocutor
across the two modes, a small effect for SCMC interactions was observed. Interactions in SCMC may help to mitigate the impact of the disadvantages noted in previous research. For example, learners’ anxiety about working with a NS interlocutor may be reduced in a computer-mediated context due to the reduction of non-verbal cues (Warschauer, 1997). Without the pressure of the social or pragmatic cues present in FTF interaction, learners may be freer to focus on language and meaning (Yamada & Akahori, 2007). The absence of this input may provide learners with opportunities to test hypotheses about the target language without social pressure (Payne & Whitney, 2002).

Research has also indicated that learners in SCMC interactions may participate more often and more equally than learners in FTF environments (Fitze, 2006; Kern, 1995; Kitade, 2000; Warschauer, 1996), with studies showing SCMC leads to increased individual practice time (Bueno-Alastuey, 2011), thereby offering learners opportunities to share the control and direction of the interactive task. In addition, although research has suggested that NSs may ignore learners’ errors, studies in SCMC have also noted low rates of corrective feedback provision in student dyads (Bower et al., 2003), indicating that participants’ use of feedback may be due to factors other than language background.

However, although results indicated a small effect for SCMC, the difference was not statistically significant, suggesting that type of interlocutor has no substantial impact across modes on learners’ overall outcomes. The lack of significance may be caused by the small number of effect sizes available, as the majority of studies sampled in the current analysis involved NNS-NNS dyads (n = 12). The lack of studies using NS-NNS interactions highlights the need for additional research examining the effects of dyad composition on L2 development. In order to develop a clearer picture regarding the
impact of interlocutor characteristics on L2 acquisition, more studies examining a variety of dyad and group compositions are necessary.

For example, researchers should consider examining the role of pair or group proficiency on L2 performance and development. Although many studies consist of learners with relatively similar proficiency levels, for instance, a class of intermediate learners, there are likely to be differences in language skills across learners nonetheless with some learners excelling in oral production while others are better at reading comprehension. This area warrants further investigation, as research has demonstrated a variety of results when learners interact with an interlocutor with a different proficiency. For example, Leeser (2004) found that as the overall proficiency of a pair increased, learners produced a greater number of LREs, resolved more LREs, and improved their focus on form. Watanabe and Swain (2007) found similar results, with pairs involving higher proficiency interlocutors producing more LREs than pairs with low proficiency interlocutors. In addition, their results seem to indicate that interactional patterns, and therefore developmental opportunities, are also affected by learners’ proficiency level. In general, studies seem to suggest that grouping different proficiency learners is facilitative of L2 development (Ohta, 2001; Storch, 2001; Swain & Lapkin, 1998; Watanabe & Swain, 2007), although more research is needed to provide more conclusive findings.

One way to better our understanding of the effects of proficiency is to examine pairs consisting of different proficiency levels, such as high-high, high-low, and low-low proficiency. Repeated measures designs pairing the same learners with different interlocutors of varying proficiency would facilitate direct comparisons across levels, mitigating the effects of learners’ individual characteristics such as motivation, aptitude,
and anxiety. Examining the influence of pair proficiency on learners’ use and provision of interactional features and feedback would provide important information for instructors seeking to maximize the positive benefits of interaction and learning opportunities available to them. In light of the fact that classes often consist of learners with varying levels of linguistics and communicative competence, more research explicitly examining the impact of interlocutors with different proficiency levels on L2 learning outcomes is needed.

In addition, research has found that performance varies according to individual learner experiences (Sauro, 2012), suggesting that researchers should consider factors, such as learners’ length of stay in the target country, when forming groups for interactional treatments. Although the studies included in the current sample feature a range of L1s and target language, the majority of the studies in the current sample featured learners with alphabetic L1 writing systems. Due to the differences between alphabetic and character-based input typing systems, such as those used by speakers of Japanese or Chinese, studies examining pairs consisting of different L1s may help to tease apart the effect the typing systems may have on the quantity and quality of learner production.

Similarly, learners’ education levels may play a role in the efficacy of the interactions. For example, many studies consist of relatively homogenous samples using university or college level learners who may have received similar prior language instruction. This is particularly true of learners enrolled in university FL contexts, as programs tend to follow certain types of curriculum, such as communicative or task-based, across all levels of instruction. Other language programs may have participant
populations comprised of learners from a wide range of educational backgrounds, with some learners having received instruction focused more on grammar or drill than communicative language skills. Research has demonstrated that learners from certain Asian (Yamada & Akahori, 2007) or Middle-Eastern (Osailan, 2009) countries may have received instruction in primarily teacher-centered drill-based classrooms, possibly impacting learners’ levels of participation and production in computer or face-to-face interaction. In order to better understand the relationship between prior educational experience, interaction, and L2 development, research directly comparing pairs or groups of learners with diverse instructional backgrounds is needed.

5.7. Research Question 4d: How Does the Type of Target Feature Impact the Effectiveness of Interaction in SCMC?

Although previous research in interaction has found that the type of target form influences the effectiveness of the interactional treatment (Keck et al., 2006; Mackey & Goo, 2007; Mackey et al., 2000), due to the wide variation of target features in the current sample, a comparison examining specific linguistic forms was not possible. Further analyses were conducted with a subset of the current sample in order to examine whether effectiveness was affected by grammatical or lexical target items. Findings revealed a small effect size ($d = .06$) for interactional treatments in SCMC focused on grammar and a small effect size ($d = .11$) for interactional treatments in FTF focused on lexical items. Mackey and Goo (2007) found that interactional treatments were more effective for lexical items than grammatical items, and previous studies in SCMC have demonstrated that, as in FTF contexts, lexical items generate more negotiation than
grammatical features (Blake, 2000; Fernández-García & Martinez-Arbelaitz, 2002; Pellittieri, 1999; Smith, 2004; Toyoda & Harrison, 2002). However, research in SCMC has demonstrated that learners may have greater opportunities to self-repair grammatical features due to the increased monitoring and planning time provided by SCMC. These self-repair opportunities may have subsequently contributed to improved performance on posttest measures, resulting in the small effect for grammatical items in SCMC when compared to FTF.

Although the current research indicates promising potential for interaction in SCMC for the development of a variety of linguistic features, including lexical and grammatical items as well as pronunciation and pragmatic development, more research examining specific types of features is needed. For instance, to the best of my knowledge, there are no studies examining the effects of interaction and corrective feedback in SCMC on the development of simple and complex features, an area that remains somewhat underexplored in FTF contexts as well (although see Ayoun, 2001; de Graaff, 1997; DeKeyser, 1995; Ellis, 2007; Lyster & Saito, 2010; Robinson, 1996). In general, findings indicate that different types of feedback affect the learning of simple and complex target structures in unique ways. For example, Ellis’s (2007) classroom study on the effects of recasts and metalinguistic corrective feedback on simple and complex grammatical structures found that there was no difference in the performance of the control and recast groups. In addition, there were no differential effects on complex and simple structures for implicit corrective feedback. Although the explicit feedback group, which received metalinguistic information about the source of the non-target like structure, performed better than the control and recast groups, the initial effects were greater for the complex
form and the delayed effects were greater for the simple form on the oral imitation test, suggesting that the impact of different types of feedback varies across target structures.

Although Ellis’ (2007) results seem to support the effectiveness of metalinguistic feedback over recasts, the assessments used in this study may have been biased towards explicit measures, with only one test of implicit knowledge used in the battery of three measurements. Additional assessments of implicit knowledge would have provided a more complete understanding of the differential effects of feedback on the development of simple and complex linguistic forms. These caveats highlight the need for additional investigation into how the effects of feedback vary depending on the target structure, as well as the short-term and long-term effects of explicit and implicit feedback on second language development. Although some studies have found similar advantages to metalinguistic correction over recasts in SCMC (e.g. Yilmaz, 2012), no research to date examines the effects of different types of feedback on simple and complex forms. Researchers should explore this area further, as additional investigations will provide important information regarding the impact of mode of interaction on the development of simple and complex L2 structures.

The studies collected for the current sample targeted a wide range of L2 features, which, due to the small sample size of lexical and grammatical items, prevented a reliable statistical analysis of the relationship between mode of interaction, target item type, and L2 learning outcome. Trends indicate positive benefits for SCMC on the development of grammatical forms, however, research has shown that interaction in SCMC is facilitative of lexical development (e.g. Smith, 2004, 2005). In order to better understand the relationship between SCMC and the development of lexical and grammatical target
features, more research is needed. In addition, based on the data presented here, these findings highlights the need for further research into specific areas, such as the development of learners’ pragmatic and pronunciation skills. Only one study in the current sample (Sykes, 2005) examined the development of pragmatic knowledge, while careful auditory and acoustic analyses of learners’ pronunciation were often absent from studies examining oral proficiency (e.g. Sequiera, 2010).

Furthermore, findings from the current research revealed that studies may focus on measures of general proficiency rather than on the acquisition of specific target forms (e.g. Sequeira, 2010; Sullivan & Pratt, 1996). Although it is important to understand the effects of interaction in computer-mediated contexts on learners’ overall oral or written development, it is also critical that researchers understand how the type of target structure determines the potential benefit. Based on the data analyzed in this dissertation, the current findings underscore the need for research examining specific target features, as this will allow us to tease apart the effects of interaction in SCMC on specific forms, including the impact of the development of simple and complex forms, providing a local as well as global perspective on the potential role of SCMC in SLA. In addition, it has been argued that the salience of a target structure may influence the benefit derived from the interaction (Long, 2007). Although SCMC has been argued to enhance the salience of target forms (Smith, 2004, 2010), research in this area could be expanded (although see Yilmaz, 2012), emphasizing the need for further investigations into the role of salience and complexity of target items at enhancing or mitigating the positive effects of interaction.
Overall, according to the data included in the current study, findings have identified under-researched areas in need of further exploration. Pursuing a research agenda examining more specific target features, as well as investigating the impact of structural salience and complexity, will allow more finely-grained analyses regarding the potential of SCMC to support learners’ development across a range of L2 items and skills. Although these results provides encouraging evidence for the overall positive benefits of interaction in SCMC, in order to better understand the impact of type of target form on the effectiveness of interaction in SCMC, more research examining the development of specific linguistic features is needed.

5.8. Research Question 4e,f,g: How Does the Mode of Interaction, Type of Feedback Provided During the Interaction, and Presence or Absence of Modified Output Opportunities Mediate the Relationship Between Interaction in SCMC and L2 Learning Outcomes?

5.8.1. Mode of interaction

Although the use of SCMC is dependent on a variety of factors, the development of technology is one of the most influential. Although the introduction of new technologies does not guarantee improved learner outcomes, it does provide new environments in which to examine language learning. Text-chat has traditionally been the most prevalent form of SCMC (Levy & Stockwell, 2006), with researchers examining virtual online environments, such as MOOs, and chat rooms as contexts for L2 learning. However, although text-chat was the most common form of SCMC noted in the search and retrieval phase of the current study, there is a growing number of studies examining audiovisual
forms of communication through the use of video-chat programs such as Skype or Google Hangouts becoming more common (Warschauer, 2004). Despite this increasing interest in oral SCMC, which includes both audio and video modes, the number of studies using these modes remains small overall (e.g. Jepson, 2005; Lee, 2007; Yanguas, 2010), with only three studies in the current sample using an oral communication mode (Bueno-Alastuey, 2011; Sykes, 2005; Yanguas, 2012). In addition, none of the studies included in this analysis were conducted within a virtual environment, such as SecondLife, or within a gaming context. Interest in the use of video games to teach a second language has grown rapidly (Aldrich, 2009; Thorne, Black, & Sykes, 2009), with the now popular modality providing interesting benefits, such as lowered affective filters, multiple practice opportunities in different contexts, and concentrated samples of targeted input (Reinders, 2012). Overall, results are promising and suggest that gaming provides a supportive environment for interaction and its associated benefits. For instance, Suh, Kim, and Kim (2010) found that participation in synchronous game-based interaction led to higher scores in listening, reading, and writing, while the results of Rankin et al. (2006) indicated improved vocabulary knowledge and target language output following synchronous game-play. However, more research is needed in order to better understand the potential contributions of game-based interaction on L2 development. Overall, the various modes of SCMC seem to provide learners with a variety of opportunities to benefit from interaction, although more research must be done before future meta-analytic work can come to any firm conclusions.

5.8.2. Type of interactional feedback
Interaction research has fostered a rich debate regarding the effectiveness of different types of corrective feedback, with some researchers arguing in favor or one type of feedback over another and others arguing comparisons are unreliable as different types of feedback are thought to impact different types of knowledge and learning (Ellis et al., 2006). Although corrective feedback is generally considered to be beneficial to second language acquisition (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), the effects of implicit and explicit types of feedback on learners’ interlanguage are not as clear. According to Ellis et al. (2006), explicit feedback provides learners with a clear indication that an error has been made, thus drawing attention to the error, and is often operationalized as either metalinguistic feedback or explicit correction. Implicit feedback, on the other hand, is designed to overtly draw the learners’ attention to his or her error (Ellis et al., 2006), and is often operationalized through clarification requests, elicitation, repetition, and recasts (Li, 2010; Lyster & Ranta, 1997). Results examining the differential effectiveness of the two types of feedback vary across studies, with some findings supporting the use of explicit feedback over implicit feedback (Carroll & Swain, 1993; Carroll, 2001), while others found implicit corrective feedback to be more effective (Leeman, 2003). Other studies indicated no significant difference between the effects of explicit and implicit feedback on L2 development (Kim & Mathes, 2001). Ellis et al. (2006) suggest that this variation in results may be due to the different operationalizations of implicit and explicit feedback, as well as the lack of valid measurements of both implicit and explicit knowledge in the testing phase. Because many of the studies used assessments favoring explicit knowledge, such as untimed grammaticality judgments and translation tests, results did not accurately represent gains.
in implicit knowledge. In the studies that did employ measures more likely to assess implicit knowledge (Muranoi, 2000), comparative measures of explicit feedback were not provided. These issues in testing methodology prevent clear, generalizable conclusions regarding the effectiveness of implicit or explicit feedback from being made.

Overall, research seems to indicate that explicit forms of corrective feedback are more effective than implicit forms of feedback on both immediate and short-term delayed measurements of second language development, although the effects of implicit feedback may be greater or more enduring in long-term improvement. Findings also suggest that explicit corrective feedback can lead to gains in both implicit and explicit knowledge, though implicit corrective feedback may be more effective in developing implicit knowledge due to its greater long-term effect. Furthermore, the target structure itself may also be a significant influence on the effectiveness of different forms of corrective feedback (Ellis, 2007).

Although it is clear that corrective feedback has an effect on second language acquisition (see Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006, for reviews), additional research is necessary in order to more clearly understand the differential effects of implicit and explicit forms on learner development, particularly on the development of different target structures. As Mackey and Goo (2007) point out, the overall effectiveness of one type of feedback over another remains unclear. This lack of clarity exists in both FTF and computer-mediated contexts, as there is little research in computer-mediated contexts comparing the efficacy of different forms of feedback. For example, a relatively small number of studies have compared feedback within SCMC (Loewen & Erlam, 2006; Sauro, 2009) or examined the relationship between mode and
implicit and explicit forms of feedback on L2 development (Yilmaz, 2012; Yilmaz & Yuksel, 2011). In order to better understand how the effectiveness of different types of feedback might be enhanced within SCMC, more research directly comparing implicit and explicit forms in both FTF and computer-mediated contexts is needed. In addition, researchers should be more transparent regarding the type of feedback taking place during interactional treatments, as this will help contribute to the existing body of work on feedback, development, and computer-mediated contexts.

Within the current analysis, the majority of studies did not explicitly mention the individual types of feedback provided during the course of the interactions. Rather, it is assumed that interlocutors used a range of feedback types during the interactional treatments. This might range from implicit forms, such as recasts and confirmation checks, to explicit forms of meta-linguistic correction. Although this diversity in language use is a natural phenomenon of human interaction, the lack of direct investigation of specific types of feedback across the sampled research precluded any statistical analysis. For instance, only two empirical studies examined the effects of recasts (Baralt, 2010; Petersen, 2010) and two directly mention investigating the impact of negotiation (de la Fuente, 2003; Yanguas, 2012), prohibiting any calculation of effect sizes regarding the advantage of one type of feedback over another in promoting L2 development.

Although research examining the differential effects of types of interactional feedback in SCMC has grown during the last decade (Bower et al., 2003; Loewen & Erlam, 2006; Sauro, 2001; Yilmaz, 2012; Yilmaz & Yuksel, 2011), based on the current data, it remains an under explored area in need of further development. One of the goals
of this analysis was to assess where interaction through SCMC was more beneficial to learners than interaction in FTF, and although overall results suggest there are no significant differences between the modes, the effect of the communication mode through feedback is provided remains unclear. Although some researchers have suggested that certain types of feedback may be more effective in one communication mode over another (Yilmaz & Yuksel, 2011), without more research examining the influence of individual feedback types on L2 development, questions regarding which type of feedback is most effective in SCMC remain. For example, because SCMC has been argued to enhance salience and attention to form (Payne & Whitney, 2002; Smith, 2004), implicit forms of feedback may have more positive benefits in computer-mediated contexts over FTF contexts. In order to address these types of questions, future studies should directly compare different forms of implicit and explicit feedback in both FTF and SCMC on learners’ development and performance, as one mode may be more facilitative of certain types of feedback.

5.8.3. Presence or absence of modified output opportunities

Swain’s output hypothesis (1985, 1995, 2005) argues that opportunities for learners to push their language production and modify their output during interaction are beneficial to L2 development. This production not only provides practice, potentially promoting fluency and automatization, but also allows learners to test out hypotheses regarding the target language, directs them to attend to target language forms, and to notice any gaps between their interlanguage and the target language (Swain, 1995). However, very few studies in FTF contexts have directly examined the effects of modified output
opportunities on L2 development (Adams, Nuevo, & Egi, 2011; Ellis, 1993; McDonough, 2004, 2005), with the majority of studies providing indirect evidence of the facilitative effects of modified output on L2 development (e.g. Mackey et al., 2003; Pica, 1988; Pica, Holliday, Lewis, & Morgenthaler, 1989; Shehadeh, 1999, 2002; Swain & Lapkin, 1995; Van den Branden, 1997). There appears to be a similar lack of direct investigations in SCMC (although see Smith, 2008, 2009, 2010 for investigations of self- and other-initiated modified output and L2 development), highlighted by the fact that none of the studies included in the current analysis directly examined the efficacy of modified output opportunities on learners’ acquisition. Furthermore, modified output opportunities were not included as part of any of the sampled studies’ research designs, suggesting that this factor was uncontrolled for within each of the individual studies. Therefore, despite the theoretical importance of modified output to learner development, the control of modified output opportunities was not included as a variable in any of the analyzed studies, preventing the examination of any mediating effect within the current sample.

Overall, definitive conclusions regarding if and how learners’ output impact L2 learning are still difficult to make (Shehadeh, 2002), due to the lack of studies investigating the direct link between output and development. In addition, it is difficult for researchers to accurately measure whether learners have in fact acquired the form, or are simply repeating a reformulation of the target item received via input without any actual changes to their IL having taken place (Mackey, 2012; Mackey & Philp, 1998; Long, 2007). Studies in FTF contexts have indicated that immediate repetition of a target item is not necessarily a strong predictor of later learning (Mackey & McDonough, 2006), suggesting that modified output cannot be equated with actual learning or
development. Previous research has also demonstrated that the efficacy of modified
output may be mediated by other factors, such as task complexity (Nuevo, Adams, &
Ross-Feldman, 2011; Révész, Sachs, & Mackey, 2011). Research in SCMC has also
produced varying results regarding learner uptake, although some of the variation may be
attributed to common methodological limitations, mainly that approaches are
retrospective/introspective (relying on self-report) or product-oriented (indirect
measures). However, although research has demonstrated both similarities (e.g. Blake,
2000; Pellittieri, 2000; Tudini, 2003) and differences between FTF and SCMC interaction
(e.g. Smith, 2003; Toyoda & Harrison, 2002), to the best of my knowledge, there is no
research directly examining the relationship between modified output, L2 development,
and mode of communication.

In order to obtain more information regarding the direct link between modified output
and L2 development, more research is needed. SCMC offers researchers opportunities to
empirically examine instances of uptake in relation to what they attend to in the input
using technology such as eye-trackers (Smith, 2010). Although it is possible to assess
what learners attend to in FTF contexts through self-report or retrospective measures, the
unique environment of SCMC provides researchers with more concrete assessments of
learners’ noticing and subsequent use of target forms. More studies examining self- and
other-generated modified output using eye-trackers will help researchers make valuable
contributions to our knowledge of the role of modified output opportunities in learners’
acquisition of L2 forms. Furthermore, direct investigations of the efficacy of modified
output opportunities in SCMC on L2 development are needed to deepen our
understanding of their potential impact when compared to FTF contexts, particularly as
delayed modified output (McDonough & Mackey, 2006) may be more prevalent in text-chat where learners have more processing and planning time to produce responses (Beauvois, 1992). In addition, the effects of split-negotiation routines on the production of modified output remains unclear, underscoring the need for more research directly examining the impact of SCMC specific features on modified output opportunities and learners’ subsequent L2 development. Overall, the additional processing and planning time present in SCMC may lead it to be more facilitative of the production of modified output, but until more studies are conducted directly examining the role of mode on learners’ output, any benefits remain speculative.

The results of the current study have made clear the need for further investigations of the direct link between modified output opportunities, production, and L2 development, not only in FTF but in computer-mediated contexts. Research which controls for and compares learners’ opportunities to produce modified output will help to tease apart the role that modified output plays in development, providing a valuable contribution to the field of interaction research, while comparative studies across mode will empirically identify any potential advantages unique to SCMC. In addition, research should examine the short and long-term impact of opportunities for and production of modified output, as effects may also vary over time.
6.1. Limitations

Although the current meta-analysis makes valuable contributions to the field, as with most research of this type, there are some important limitations that must be noted. Due to the application of rigorous eligibility criteria, the sample of individual effect sizes (n=37) calculated from the 14 primary studies is relatively small. Although there are ample examples of meta-analyses published with similar sample sizes in L2 research (e.g. Keck et al., 2006, $k = 14$; Russell & Spada, 2006, $k = 15$), the reliability and external validity of the results would be improved with a larger sample size. The reader is also cautioned to interpret the results of moderating variables with caution, as the small number of effect sizes obtained for certain features, such as SL contexts or NS-NNS interactions, prevents the drawing of firm conclusions.

Although the current study was able to include 14 primary studies in the final sample, a number of comparisons were negatively impacted by the lack of sufficient data. Important constructs, such as the differential efficacy of individual types of interactional feedback or modified output opportunities, could not be analyzed because there were not enough studies examining these features. These features, although identified in interaction research as important variables, have not been systematically examined in a sufficient quantity in SCMC to allow for a statistical synthesis of the findings. Furthermore, despite promising research comparing various modes of computer-mediated research that suggests there are important differences in efficacy across modes (Yanguas, 2010, 2012), not enough research met the eligibility criteria to warrant inclusion in this
meta-analytic assessment. Until more research is conducted in these areas, firm synthetic conclusions regarding their impact on L2 learning cannot be made.

As with any study, there are also limitations inherent to the design. Some researchers have criticized meta-analyses for ‘mixing apples and oranges’ by combining different kinds of studies into one aggregate sample. These criticisms maintain that the average effect size will ignore potentially important differences across the collected studies (Borenstein et al., 2009). However, in a field as diverse as CALL and applied linguistics, it is unavoidable that studies will differ across a number of characteristics, including participant population, research setting, educational context, and target language. In addition, treatments vary according to type of treatment task, exposure time, and testing instruments. However, an important caveat to the lack of similarity across studies is that meta-analyses are designed to ask broader research questions than the primary studies on which they are based. In other words, meta-analysts may be more concerned with fruit in general, a classification to which apples and oranges correctly belong. However, not every type of ‘fruit’ is included in the current analysis, a limitation which is discussed in detail below.

Another important limitation of this research is that some key studies in the field have not been included in the final sample. The inclusion criteria for this study were rigorous, causing a number of studies to be eliminated for failing to meet the standards set forth in the methods section. For example, there are studies that make important contributions to the field of interaction research in SCMC (e.g. Blake, 2000; Chun, 1994; Kern, 1995; Rosa & Leow, 2004; Smith, 2004, 2005), but were not included in the current sample because they did not utilize a two groups pretest-posttest or posttest only design.
Although the eligibility criteria were driven by the research questions, it is important to acknowledge that this limited the current sample to quasi-experimental and experimental studies using a comparison groups design. Furthermore, the studies in the current sample used an independent groups design, restricting comparisons to between-groups contrasts. In addition, not all studies in the current analysis controlled for between group differences at the posttest, suggesting that learners’ individual differences may have affected the individual and therefore summary effect sizes. In order to better understand the impact of interaction in SCMC on L2 learning outcomes, future analyses should include repeated measures within-groups designs in order to account for possible influencing differences across learners. Analyses were also limited to posttest measures only and did not provide an assessment of the effect of mode on L2 gain scores. A comparison of gain scores across FTF and SCMC modes would provide evidence for differences in the rate of acquisition, a finding with potentially interesting pedagogical implications. In order to deepen our understanding of the differential efficacy of interaction on the rate of development over time, future meta-analytic work may wish to consider this line of research.

In general, although these studies were systematically eliminated from the final sample based on a careful set of substantive and methodological criteria, their exclusion limits the generalizability and external validity of the current findings, as not all studies examining the effects of interaction in SCMC were included. Future meta-analytic work may wish to expand the current line of research by examining the effect sizes of studies using a one group or within groups designs, or by investigating the role of mode on L2 gain scores, thereby providing a complementary analysis to the current findings.
In addition, there are observational and descriptive studies that have had a positive impact on SCMC-based research (e.g. Beauvois, 1992, 1998; Kelm, 1992; Salaberry, 2000), but due to their design, were not included in the current sample. These studies have provided researchers with important information regarding interactional features, such as the most common triggers of negotiation (e.g. Blake, 2000; Pellittieri, 2000; Tudini, 2003) and the occurrence of negotiation in FTF and SCMC contexts. However, because they did not include the necessary statistical information, such as means and standard deviations, they were not meta-analyzable. The exclusion of this research from the current sample does not imply that they are not important to the field, but it does highlight the need for more rigorous reporting of statistical measures. In order to facilitate future meta-analytic work, it is critical that studies report sample size, means, and standard deviations for all outcome measures.

One of the goals of the current study is to make suggestions regarding the improvement of methodological quality in SCMC interaction research (please see the following chapter for a more thorough discussion). Although the current study notes a number of areas for improvement in research methods, the suggestions regarding quality are somewhat limited in scope due to the small number of journals represented by the sample. SCMC research is published in a wide variety of journals targeted towards specific audiences within the fields of applied linguistics and CALL, and the number of journals included here provides only a small slice of the multiple publications available. In order to better understand the methodological progress and limitations in our field, a more complete methodological review of SCMC research should be conducted with a larger sample. As mentioned in the previous limitation, a number of studies were
excluded because they failed to meet the eligibility criteria for the substantive features that were the target of the current research. By focusing more closely on methodology, a systematic review encompassing both experimental and descriptive designs might provide a more global perspective on the design aspects of SCMC research over the past few decades. An empirical comparison across, as well as within, a wider range of SCMC research in CALL and SLA journals would provide important information regarding the methodological quality of published studies in the field.

Although careful steps were taken to control for the possibility of publication bias in the current research, the results may have been impacted by bias nonetheless. Studies were not retrieved from conference proceedings, collections of working papers, or masters’ theses, which may have contributed to the likelihood of bias. In addition, the invisible college was not surveyed for research meeting the inclusion criteria. Crane (1969) defines the invisible college as an informal group consisting of researchers working on similar issues who are usually aware of one another’s research. Because these avenues were not pursued during the search and retrieval phase, eligible studies may have been missed, thereby contributing to publication bias. The impact of these missing studies may have resulted in an upward bias caused by unreported non-statistically significant findings. Furthermore, the studies in the current sample were limited to publications in English, a substantial limitation considering the worldwide interest in CALL. Future research should consider expanding the source materials, as well as including primary studies published in other languages.

Despite these limitations, the current study has deepened our understanding of the relationship between interaction and SCMC as well as providing a structured framework,
which, as the body of relevant work continues to grow, can be used for future meta-analytic investigations. The eligibility criteria, while strict, provide a foundation on which later studies can elaborate or expand in order to address new research questions. Grounded in synthetic and meta-analytic work, the development of these research questions and the methodologies used to answer them will push the field forward into new areas at the intersection of CALL and SLA.

6.2. Expanding the Research Agenda

During the last few decades, interaction research conducted in computer-mediated contexts has evolved from descriptive and observational studies of interactional features to complex empirical examinations of the direct link between interaction in SCMC and L2 development. In addition, interaction researchers in general have improved their reporting of important validity and reliability issues (Plonsky & Gass, 2011; Mackey & Goo, 2007), with many CALL researchers answering Chapelle’s (2001) call for more theoretically grounded and methodologically sound research. However, despite these improvements, based on the data presented here, a deficit remains regarding the use and reporting of critical methodological features in computer-mediated interaction research. In addition, many of the studies examined in the initial search and retrieval phase failed to report the statistical analyses necessary for a meta-analysis or other statistical comparison. Early studies in particular, such as those conducted in the early to mid-1990s, provided only observational information without the support of descriptive statistics, such as the mean or standard deviation. Studies that may have met the other
inclusion criteria were eliminated because they lacked this critical statistical information, failing to report more than the sample size in many cases.

Many of these early studies used a pretest-posttest design, but did not report the scores for the treatment and comparison groups, instead reporting correlations of test scores with attitudes or participation measures. Although these reporting conventions may have been driven by the research questions, which at the time were more concerned with describing the interaction rather than empirically investigating a link between interaction and learning, they not only prevent a more comprehensive meta-analysis, but provide the reader with only a partial interpretation of the impact of the treatment on learner outcomes, due to the omitted information. Although more recent research demonstrates that reporting conventions have improved, with the majority of studies sampled reporting descriptive statistics, the reporting is still sometimes incomplete, with studies selectively reporting means and standard deviations only for statistically significant comparisons (e.g. Sykes, 2005).

In addition, despite the numerous calls for improved methodology in the field of interaction, and second language research in general, (Mackey & Goo, 2007; Plonksy, 2011; Norris & Ortega, 2000), none of the collected studies published confidence intervals or standard error. Confidence intervals have been identified by researchers in psychology and applied linguistics as a measure critical to improving researchers’ understanding of statistical testing, and some have argued they are more informative overall and useful for further post-hoc comparisons (Larson-Hall, 2010; Kline, 2004; Wilcox, 2001, 2003). Confidence intervals provide a range of potential values in which the sampled population or parameter is likely to fall (Kline, 2004). For example, a
confidence interval of 95%, one of the most commonly used levels (Hu, Zhao, & Yang, 2010), indicates that 95% of the time the parameter or population of interest will be found within the identified range. In addition, a 95% confidence interval implies that there is a 5% chance that the same results would not be found were the study to be replicated. In other words, these intervals indicate the level of confidence the reader might have in the results. Furthermore, because confidence intervals provide a range, they offer information regarding not only the statistical significance of the difference between groups, denoted by a range which includes zero, but an indication of the effect size by showing how far from zero the confidence level can be found. However, confidence intervals that do not include zero, and are therefore significant, should also be interpreted cautiously. For instance, a study might report confidence levels with a wide range that does not include zero, indicating a statistically significant difference. The wide range, however, indicates that the precision of the findings is low, and that any estimates are limited. If the confidence interval is narrow, this suggests that the sampling error was relatively small and the results can be expected to be accurate estimates of the effects of the independent variable. Overall, despite the crucial information confidence intervals can provide to the reader, it is quite rare for researchers in CALL or applied linguistics to report them.

Neglecting to report these features obscures, or in some cases even conceals, important information necessary for accurate interpretation of the data.

Power analysis, although also an important aspect of statistical reporting, was reported in only one of the studies (Baralt, 2010) collected for the current research. Power provides information regarding the probability of obtaining a statistical result regarding the differences between groups. Studies conducted with adequate power
ensure that accurate results will be obtained and valuable findings will not be lost due to Type I and Type II errors, respectively. Within the social sciences, power analysis is often interpreted as the probability that the significance test will lead to correct conclusions regarding the null hypothesis (Murphy & Myors, 2004). Studies lacking in power are unlikely to uncover true effects, potentially causing researchers to come to unreliable conclusions regarding their research questions. Indeed, Plonsky (2011) notes that the average power in L2 research is .57, falling below the level of .80 deemed adequate in social sciences and psychology (Murphy & Myors, 2004). A power level of .57 indicates that L2 researchers are willing to accept a probability only slightly better than chance that they will be able to confirm their research hypotheses, indicating the possibility that much of the published research may be unreliable in its conclusions.

Despite the problems regarding power, it is important to acknowledge that one of the factors that affecting power is the trend in applied linguistics and CALL research of small sample sizes. Many researchers have pointed out the negative impact small sample size may have on power (Larson-Hall, 2010; Norris & Ortega, 2006; Oswald & Plonsky, 2010; Plonsky & Gass, 2011), and because researchers may lack access to large participant pools or the financial resources to recruit a critical sample size, there may be a reticence to calculate a test that would discourage an experiment with the sample size available. Classroom researchers in particular may be at a disadvantage in collecting larger samples, as they must attempt to preserve the ecological validity of the classroom-based environment while still collecting data from enough participants to obtain a sufficient level of power. Many FL and SL programs are limited in size, and many programs advertise small class sizes of 15 students or less as an attractive feature to
potential students. These small class sizes may place logistical and methodological limitations on researchers seeking a larger sample, and for researchers examining the impact of treatments on certain proficiency levels or L1s, the number of qualified participants is likely to be further reduced.

As Plonsky (2011) points out, the failure to report statistical power is a chronic issue in second language research. After surveying more than 600 L2 studies, Plonsky’s (2011) results indicate that only 1% of the sample conducted power analysis, and of these, a number were post-hoc analyses. However, power analysis is most meaningful when conducted a priori (Wilkinson and the Task Force for Statistical Inference, 1999), when the researcher still has the opportunity to improve it by collecting a sample size appropriate for obtaining statistical results. According to Kline (2004, p.43), post-hoc power analyses are ‘more like an autopsy than a diagnostic procedure,’ and although small sample sizes seem to be a defining characteristic of the field, it is necessary for researchers to acknowledge the methodological and interpretive issues arising from the lack of power analyses.

Overall, power indicates the probability of detecting the impact of the independent variable on the dependent variables. Without understanding the importance of power, researchers finding insignificant results may relegate these studies to the file drawer on the potentially erroneous assumption that there is no effect due to the treatment, when perhaps the issue was there wasn’t enough power to identify the impact of the treatment in the first place. Important discoveries may be lost due to the lack of knowledge regarding power and its implications on research, and if researchers examining interaction in SCMC hope to provide reliable results, it is necessary to take the
appropriate steps to remedy these issues. For those researchers restricted to working with small sample sizes, inferential statistics should be used judiciously to avoid the dissemination of potentially misleading results, and perhaps most importantly, journal editors and reviewers should carefully reconsider the perceived requirement of statistically significant findings for publication.

Dissertations, although still failing to report some features such as confidence intervals or statistical power, provided the most complete reporting of statistical measures, although even these were lacking, with only 67% reporting features such as a priori alpha levels and effect sizes. Although Mackey and Goo (2007) note the improvement in reporting of effect sizes, the data presented in the current analysis demonstrate that there is still room for improvement, as only one journal article included in the present study reported effect sizes (Yanguas, 2012). The absence of the reporting of this feature is particularly disturbing as effect sizes, which have been called the ‘most important parameter of all’ (Murphy & Myors, 2004, p. 106) indicate the amount of the variance that can be attributed to the impact of the independent variable. In other words, effect sizes indicate whether the difference between groups is substantial or negligible. Because effect sizes provide the reader with information regarding how much of an influence the treatment may have had on the outcome, providing readers with the information needed to critically assess the reliability of results across various populations, research designs, and analyses (Wilkinson and the Task Force for Statistical Inference, 1999), they are crucial to the interpretation of the findings, and therefore the development of the field. Many CALL researchers report p values and the results of significance testing in their results sections. However, although useful for null hypothesis testing,
these measures are to a large extent dependent on group size, while effect size measures do not change regardless of the sample size. This feature of effect sizes provides valuable information regarding treatments, as results are able to indicate the potential effectiveness of a treatment that may not reach statistical significance due to small sample size or low power. In addition, in tests reaching statistical significant, effect sizes provide information indicating whether the difference is largely unimportant (for small effects), or if it is something warranting further investigation (for large effects). Studies reporting a statistically significant finding without the effect size prevent the reader from being able to fully interpret the results, and do a disservice to the field by obscuring the true impact of the treatment on the dependent variable.

Another inadequately reported statistical measure in the current sample was the reporting of a priori or exact p-values. Twenty-eight percent of the current sample reported an a priori level for statistical significance, a finding similar to that of Plonsky & Gass’ (2011) examination of interaction research, where 25% of the studies reported a predetermined value of statistical significance, and L2 research in general, where 22% of the studies examined reported an a priori level (Plonsky, 2011). Only about half of the studies collected here reported exact p-values, and considering that the p-value provides an index of group size and power, thereby impacting the interpretation of the results, this information should be readily provided. A number of researchers in the social sciences advocate for the reporting of exact p-values rather than simply indicating if the value falls above or below alpha, as this number may be misleading (Larson-Hall, 2010; Kline, 2004; Maxwell & Delaney, 2004; Murphy & Myors, 2004). For instance, a significant p-value does not necessarily indicate an important result, as effect sizes provide more
information regarding the impact of the independent variable on the dependent measure. A small $p$-value also might indicate higher power, suggesting that similar results would be found were a replication to be conducted. Given the recent interest in applied linguistics and CALL on replication, as indicated by a number of publications (Abbuhl, 2011; Mackey, 2012; Porte, 2012) and upcoming special issues (Smith & Thorne, email communication), this type of information is vital to researchers interested in replicating research. Based on the present data, researchers failing to report confidence intervals and exact $p$-values, thereby withholding information that would indicate the likelihood that the same results would be found in a replication, are failing to highlight research questions in need of further investigation and potentially stifling growth within the field.

Although the sample size of the current analysis is admittedly small and the scope somewhat narrow, as additional studies reporting effect sizes, exact $p$-values, power, or confidence intervals may have been omitted for not meeting the strict inclusion criteria, the sample is relatively representative nonetheless, suggesting that there is room for improvement in the overall reporting found in SCMC interaction research. In order to facilitate future meta-analyses, reporting conventions should be more rigorously followed, with all empirical research providing descriptive statistics for all statistical analyses at the minimum. However, in order to provide research consumers with a complete understanding of the impact of the treatment on learners’ development or performance, researchers should consider reporting effect sizes and confidence intervals.

In addition, much of the research in interaction in SCMC involves complex coding of interactional features, such as instances of negotiation or corrective feedback. In order to demonstrate that the coding has been reliably conducted and confirmed, interrater or
intrarater reliability should be reported. Within the current sample, only half of the studies reported any reliability measures. Although this is similar to reporting levels found in previous analyses (Plonsky, 2011; Plonsky & Gass, 2011), the overall number of reports is still low, indicating the need for improvement in this area. Without information regarding the reliability of the coding, the reader may be inclined to question the validity of the results. Lack of reliability reporting harms the credibility of the research, and with the increase of materials sharing through organizations such as the IRIS project, researchers using similar instruments are restricted in their ability to compare materials and samples. Furthermore, many studies often include survey measures designed to obtain information on learners’ perceptions or individual differences. Although coding is not necessarily a factor in the reporting of survey measures, particularly those using Likert type scales, few studies report the internal reliability or validity of the survey measures. These statistics are critical as they demonstrate that the instrument measures what it was designed to assess, allowing the reader to draw reliable conclusions from the findings. Studies lacking these measures provide somewhat inconclusive findings, as interpretations must be made cautiously due to the inadequate reporting conventions.

Overall, based on the data examined in the current research, results demonstrate that the reporting of statistical tests, as well as reliability and validity measures, in computer-mediated interaction research could be improved. In order to facilitate future meta-analytic work, studies should regularly report descriptive statistics, including the standard deviation for every reported mean. To increase the contributions of research, as well as better support the theoretical and methodological decisions of future studies, researchers should consider reporting effect sizes, confidence intervals, and *a priori* alpha and power.
Journal reviewers and editors can help encourage this process by examining the reporting of statistical procedures and the appropriateness of research methods of all studies having passed the first round of review. As Magnan (1994) suggested, journals should consider expanding the review process to explicitly included a statistical and methodological review for empirical studies that have passed the first round of reviews. Such an approach would highlight the importance of accurate testing and reporting, and would encourage better practices within the field.

6.3. Implications

6.3.1. Theoretical implications

The results of this study also have important theoretical implications for the interaction approach (e.g. Gass & Mackey, 2007; Long, 1996; Mackey, 2012). Second language researchers generally believe that interaction is beneficial for L2 development because it provides learners with comprehensible input, opportunities to use the second language, and opportunities to receive feedback (Long, 1996; Keck et al., 2006; Mackey, 2012; Mackey, Abuhl, & Gass, 2011; Mackey & Goo, 2007). The results of the current research have demonstrated that synchronous computer-mediated technology not only offers learners similar interactional opportunities as FTF interaction, but that there are no significant differences between the two communication modes on L2 learning outcomes. These findings provide additional support for the efficacy of interaction in both FTF and SCMC contexts as summary effect sizes indicated both modes of communication resulted in positive benefits for overall L2 learning outcomes.
However, although these findings provide encouraging evidence for the positive benefits of interaction, this dissertation has also highlighted a number of areas in need of improvement if we are to deepen our understanding of the role interaction plays in SLA. One of the goals of meta-analytic research is to identify the relationships in the primary literature that have not been sufficiently addressed, and to call on researchers to fill these deficiencies through future exploration (Plonsky, 2011). The results of the current study have identified a number of research topics warranting further exploration, including the examination of video and audio forms of SCMC, the differential effects of individual types of interactional feedback and target feature on L2 development, as well as the role of modified output, in computer-mediated contexts. This is particularly important in light of the split-negotiation routines typical of text-based SCMC, as learners may produce modified output in later rather than adjacent turns, a result first identified in FTF research by McDonough & Mackey (2006). In addition, based on the data in the current study, research examining the effects of learner characteristics and research settings on learning outcomes, as well as the efficacy of interaction in gaming or virtual settings, remains underexplored. The various computer-mediated contexts identified here (e.g. oral SCMC, virtual environments, multi-player gaming) provide a new environment in which researchers can test the hypotheses proposed by the interaction approach. By investigating the relationship between these novel contexts and interactional features, including the role of feedback, negotiation for meaning, and modified output, researchers will be able to contribute important information to our understanding of the efficacy of interaction in different communication modes.
Overall, this analysis identifies areas within SCMC and interaction research needing further exploration, thereby suggesting possible future research agendas in CALL-based SLA interaction research. By highlighting under-investigated areas of research, as has been done with previous syntheses (e.g. Lai & Li, 2011; Liu et al., 2002; Sauro, 2011), it is hoped that the present findings will help guide future work examining interaction in computer assisted environments, encouraging future researchers to continue to move the field forward and expand the scope of SCMC interaction research by addressing the gaps in the literature identified here.

6.3.2. Methodological implications

This dissertation sought to make valuable contributions to the fields of CALL and SLA by providing a replicable, systematic review of the relationships between interaction, computer-mediated contexts, and L2 development. In order to continue the advancement of current research methodologies, the present study identified areas of potential reform in the methodology of SCMC research conducted from an interactionist perspective, providing suggestions for improvements in the design and execution of future empirical studies in CALL.

For instance, individual researchers conducting experimental or quasi-experimental designs should consider the importance of using a true control group. Although many of the studies retrieved for the current study, as well as those included in the current sample, stated that a control group was part of the design, very few were true control groups. A true control group receives a placebo activity in place of the treatment activity, thereby ensuring that results are due to the treatment rather than simple task participation. Studies
within SCMC and FTF interaction, however, often define the control group as a group that does not receive any treatment and instead completes only the pretest and posttest measures. Although such a design may also contribute important information to the field, it does not isolate the direct effects of the treatment on the dependent variable and any results may be due to the Hawthorne effect.

The current results also highlighted the lack of long-term or longitudinal experimental research. Very few studies retrieved and selected for inclusion in this dissertation used delayed posttests in their design, thereby limiting conclusions to immediate and short-term effects. Longitudinal data will provide crucial information regarding the relationship between interaction, development, and communication mode, as without data from delayed posttests, our understanding of the durability and delayed impact of interaction in SCMC contexts is restricted.

In addition, this study highlighted the need for improved reporting of statistical measures. Reporting practices varied widely across studies, with some early studies failing to report basic descriptive statistics such as means and standard deviations (e.g. Warschauer, 1996), measures that are common to most studies published after 2000. Most studies in the current sample also failed to report a number of critical features, such as effect sizes and exact \( p \)-values, and all but one (Baralt, 2010) lacked power analysis. Future researchers should consider power when preparing to collect data to ensure that any differences between groups will in fact be discovered. Furthermore, consumers of research should be skeptical of studies reporting only relative \( p \)-values, as any difference between groups will attain statistical significance if the sample size is large enough. In order to provide better interpretations of results, exact \( p \)-values should be provided in
addition to confidence intervals and effect sizes. The reporting of effect sizes is crucial to understanding the true impact of the independent variable on the dependent variable, and should be included in the reporting of all statistical analyses. Lastly, reliability measures should be conducted and reported for materials, as this will provide readers with the knowledge necessary to accurately interpret any relevant findings. By examining the methodology of multiple primary studies within the interactionist SCMC paradigm, this study makes a valuable contribution by identifying these areas where researchers might improve the methodological quality of studies.

Finally, this study identified the wide variation in methods and materials across studies. In order to facilitate more direct comparisons, as well as build a body of research facilitative of future meta-analytic work, researchers should consider conducting more replications as well as sharing their instruments and methods with others in the field. These innovations are supported by recent developments in second language research, such as the IRIS project. Designed to encourage the sharing of data collection instruments for L2 research and teaching, IRIS provides free access to materials uploaded to a digital repository. IRIS expands access to materials used in published research, thereby improving transparency and facilitating replication and systematic reviews of primary research (IRIS, 2012). In order to improve comparability across studies, researchers should consider using similar testing and treatment materials, rather than creating custom-made items for each individual study. The sharing and use of similar tools across studies will provide data that can be more readily compared, thereby building upon previous research and strengthening the conclusions that might be drawn from any significant findings.
In general, it is hoped that these findings will lead researchers to design and conduct more rigorous studies, thereby producing more robust results that will help broaden our understanding of the effects of interaction in experimental and quasi-experimental research within computer mediated environments. It is also hoped that the improvement of statistical reporting will lead to improved interpretation and understanding of results by the readers of this research, as well as improving the general statistical knowledge of those new to the field. The results of this dissertation emphasize the importance of understanding, reporting, and interpreting statistical features, such as power, effect size, and descriptive statistics, and should encourage the field to foster this knowledge in novice researchers and graduate students.

6.3.3. Pedagogical implications

As the use of technology in the classroom continues to grow, methodologically sound, well-grounded, relevant research can be used to responsibly inform educational practices. By using a synthetic approach, the present dissertation has identified important findings for the second language classroom, namely that there are no significant differences in L2 learning outcomes for face-to-face and synchronous computer-mediated interaction. This finding is particularly important for distance learning programs, as it suggests that learners participating in computer-mediated contexts are likely to experience the positive developmental benefits associated with FTF interaction. In addition, no significant differences were found between the two modes on the development of learners’ oral and written skills or their productive and receptive skills. Overall, these results are highly encouraging, as they provide evidence for the wide-
ranging efficacy of interaction in SCMC. The positive benefits associated with interaction in SCMC in lexical, grammatical, and phonological skills, ranging from measures of overall proficiency to assessments of individual target items, illustrates the potential applications of SCMC in promoting L2 learning outcomes. This should provide instructors with encouraging evidence that computer-mediated communication can be successfully used independently or as a supplement to FTF within the classroom. Furthermore, when interpreted alongside other important descriptive or observational findings in the field, such as those indicating the SCMC may promote more equal participation (e.g. Chun, 1994; Kern, 1995; Warschauer, 1996) or reduce anxiety (Abrams, 2003; Kelm, 1992; Sullivan & Pratt, 1996), instructors may find that the integration of SCMC components into their classrooms are likely to provide learners with numerous linguistic, pragmatic, and communicative benefits.

Furthermore, because technology is a constantly evolving field, new and emerging media is continuously being integrated into the second language classroom. For instance, the use of Web 2.0 technologies have rapidly expanded within the classroom, with many educators experimenting with how best to support and facilitate L2 learning (Wang & Vasquez, 2012). Although the body of research is not extensive enough to warrant statistical meta-analysis, these tools, including multi-player games, virtual worlds, online collaborations, social media tools, and other user-driven technologies, provide new and exciting contexts for second language instruction and research. Because many of these tools require communicative interaction of participants, they provide exciting forums in which interaction is likely to play a role. As Wang and Vasquez (2012) point out, Web 2.0 technologies exploit the collaborative and participatory potential of the web,
highlighting their potential to be used as a context for interaction supported language learning. These technologies offer learners’ opportunities for increased student interaction, including negotiation and feedback provision, as well as opportunities to produce output in the target language (Baten, Bouckaert, & Kan, 2009; Chen, 2009; Lee, 2006; Peterson, 2006). Web 2.0 technologies offer researchers fertile grounds for examining interaction and L2 development, particularly regarding collaborative and community based learning, and since many of these tools are already an indispensable aspect of many learners’ daily lives (McBride, 2009), educators may find students more receptive or enthusiastic to second language instruction situated within these contexts.

The availability of technology provides educators and researchers with unique opportunities to not only integrate cutting-edge technology into the second language classroom, but to pursue research agendas that are continuously pushing the boundaries of CALL and how it might facilitate L2 development. Instructors seeking to integrate SCMC technology into their classrooms have many options, including video, audio, and text-chat, that are likely to appeal to a range of students. Web 2.0 tools, such as social media and forums, are interaction-driven, highlighting the important role that interactional features, including feedback, negotiation, and output, are likely to play in L2 instructional contexts utilizing these technologies. Overall, the current findings demonstrate the potential for interaction in SCMC to facilitate and support learners’ development in a diverse range of skills, as well as the medium’s potential to promote lower levels of anxiety and more equal learner participation. This suggests that SCMC need not be restricted to distance learning contexts, but could be integrated successfully
in to more traditional classroom settings to support and enhance learners’ L2 development.

6.4. Conclusion

This dissertation sought to examine the comparative effectiveness of interaction in SCMC with interaction in FTF contexts on the development of L2 learning outcomes. Through the quantitative systematic investigation of primary research conducted between 1990 and 2013, this study sought to not only provide a better understanding of the role communication mode may play in the efficacy of interaction at promoting L2 development, but to offer a critical review of methodological practices as well as suggest research areas in need of further exploration. Results demonstrate that although interaction in both face-to-face and synchronous computer-mediated communication leads to significant L2 development, thereby providing additional support for previous findings of the positive effects of interaction (e.g. Keck et al., 2006; Mackey & Goo, 2007), there were no significant differences between the two modes for overall L2 outcomes, for productive and receptive measures, or for oral and written measures. Trends in the data seem to indicate a small advantage for SCMC interaction for overall development as measured by immediate posttest scores, as well as small effects for written and productive skills, while small effects were found for FTF contexts for oral and receptive skills. Small effects were also identified for SCMC regardless of native-speaker status of the interlocutor, as well as for second-language and foreign-language settings, and classroom and laboratory contexts. Although these results have revealed promising trends in favor of SCMC, there were no significant differences between the
two modes on any of the measures, suggesting that mode of communication has no significant impact on L2 learning outcomes and the positive benefits associated with interaction.

In addition, this study identified a number of areas within SCMC interaction research in need of further investigation. Due to a lack of data, the role of modified output, type of interactional feedback, type of target feature, and mode of computer-mediated communication was not statistically analyzed. Based on the data presented here, this finding highlights the need for more research examining these topics, as not only will this provide the field with the primary studies necessary to facilitate a future meta-analytic review, but will deepen our understanding of the influence these factors may have on the efficacy of interaction across modes. Future researchers might use the inclusion and exclusion criteria in the current study as a framework from which to conduct such investigations, thereby complimenting the findings of this dissertation and filling the gaps in the literature identified here.

Furthermore, by quantitatively examining the primary research included in the current sample, this dissertation found a number of issues regarding methodological quality. Issues regarding study design and the reporting of reliability and statistical measures are in need of attention, as these concerns prohibit growth and advancement of research findings. In addition, by not encouraging and requiring the accurate use and reporting of appropriate statistical measures, researchers impede the precise interpretation of data and the identification of questions needing further development, potentially stifling growth within the field. These findings should encourage reflection within the field regarding current methodological practices, as the legitimacy of CALL research within mainstream
SLA will be limited as long as research practices lag behind in innovation, execution, and reporting.

Overall, this dissertation aimed to deepen our understanding of the role of interaction, including comprehensible input, modified output opportunities, negotiation for meaning, and corrective feedback, in synchronous computer-mediated communication, particularly in relation to face-to-face contexts. In addition, the synthetic and meta-analytic approach identified areas of potential improvement in the methodology and possible future research agendas in CALL-based SLA interaction research, hopefully inspiring methodological improvements and providing the means through which future SCMC empirical investigations that help to move the field forward.
References

Note: Studies included in the current meta-analysis are marked with an asterisk.


Language Teaching Research, 15, 201-229.


Amsterdam: John Benjamins.


Kormos, J., & Trebits, (2009)


Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom:


Marcos, TX: CALICO.


*Petersen, K. (2010). Implicit corrective feedback in computer-guided interaction: Does


Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language research and instruction. In G. Crookes & S. Gass (Eds.), *Tasks and
second language learning (pp. 9-34). Clevedon, UK: Multilingual Matters.


Smith, B. (2010). Employing eye-tracking technology in researching the effectiveness of recasts in CMC. In F. M. Hult (Ed.), *Directions and prospects for educational linguistics* (pp. 79-97). Springer.

24.


United States Department of Education U.S. Network for Education Information. Retrieved from


