The Effect of Computer-Mediated vs. Face-to-Face Instruction on L2 Pragmatics: A Meta-Analysis

Marziyeh Yousefi, Hossein Nassaji

Abstract—This paper reports the results of a meta-analysis of studies on the effects of instruction mode on learning second language pragmatics during the last decade (from 2006 to 2016). After establishing related inclusion/ exclusion criteria, 39 published studies were retrieved and included in the present meta-analysis. Studies were later coded for face-to-face and computer-assisted mode of instruction. Statistical procedures were applied to obtain effect sizes. It was found that Computer-Assisted-Language-Learning studies generated larger effects than Face-to-Face instruction.

Keywords—Meta-analysis, effect size, pragmatics, computer-assisted language learnin, face-to-face instruction, comprehensive meta-analysis software.

I. INTRODUCTION

EXAMINING opportunities offered by technology for improving oral proficiency has been a focus of many studies on the integration of technology and language learning [1]. Despite this rise in Computer-Assisted-Language-Learning (CALL) in recent years, only a small body of research has tried to examine the possible correlation between technology use and pragmatic teaching. Pragmatic competence is defined as knowledge of communicative action, how to carry it out, and the ability to use language appropriately according to contextual factors [2]. Mode of instruction refers to whether instruction is provided through the computer or in face-to-face communication. CALL is one option for pragmatics instruction. In other words, any type of instruction including (explicit, implicit, task-based, etc.) can either be delivered through face-to-face instruction or through computer mediated instruction. Mode of instruction was considered as an independent variable in this study because technology is only a potential way for designing tasks and delivery of instructional material. In order to assess the efficacy of mode on instruction in the field of L2 pragmatics, a meta-analysis can be used to summarize and present conclusive results. Meta-analysis refers to a set of statistical procedures used to summarize and integrate many empirical studies that focus on one issue [3]. According to [4], findings from *all* available primary research studies are converted to comparable values by estimating the magnitude of an observed relationship or effect, typically referred to as the *effect size*. For calculation of effect sizes, researchers usually use Cohen's *d* [5], which can be interpreted as the magnitude of an observed difference in

standard deviation units. This index is simply the standardized mean difference for any contrasts made between two groups within a primary research study.

II. PROCEDURE

The present meta-analysis is based on exhaustive electronic and manual searches to locate and retrieve full-text papers on L2 pragmatics instruction published between 2006 and 2016. The electronic search was through the two of the most commonly used databases in applied linguistics, i.e., ERIC and PhycINFO as well as LLBA, ProQuest, Google, and Google Scholar. As for the manual search, major review articles and widely cited journals were searched for relevant studies. The searched journals included Applied Linguistics, Annual Review of Applied Linguistics, System, International Journal of Applied Linguistics, Second Language Research, and Journal of Pragmatics. After the electronic and manual search, a hundred and 10 related primary studies were located. All these studies, however, did not qualify for being included in a meta-analytic study. Unpublished literatures, including dissertations, were not included because of the difficulty to track and retrieve these studies. This exclusion might lead to publication bias toward mostly significant findings in published literature known as "file-drawer" problem, i.e., publications tend to accept studies yielding statistically significant results [6]. However, according to [7] since the problem of publication bias is well recognized in the literature of meta-analysis, forest and funnel plots were used to examine publication bias.

A. Criteria for Inclusion

For the next step, the eligibility criteria were set for inclusion and exclusion of the studies initially found. Based on the present study's questions, only the studies with the following criteria from all the initially collected studies in the present meta-analysis were included.

- The study employed an experimental or quasi-experimental methodological design that allows for identification of instructional effects.
- The dependent (learning target) variable had to involve a pragmatic feature (e.g., speech acts, implicature...), rather than a purely linguistic grammatical function (e.g. verb tense)
- The independent variables had to involve some type of well-described instruction of L2 pragmatic features.
- The target language of instruction was either a second or foreign language for the study participants. (e.g., English for Malaysians, SL, English for Algerians, FL).

Marziyeh Yousefi is a PhD candidate at the University of Victoria, Department of Linguistics, Canada (e-mail: yousefi@uvic.ca).

Hossein Nassaji is Professor of Applied Linguistics in the Department of Linguistics at the University of Victoria, Victoria, BC, and currently the Department Chair, Canada e-mail: nassaji@uvic.ca).

World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:12, No:8, 2018

- Tasks constituting the instructional treatment are adequately described in terms of pedagogic features congruous with the fields of L2 pragmatics, in particular, and SLA, in general (input-based activities, meta-pragmatic discussion, etc.).
- The study is written in English and published between 2006 and 2016.

These inclusion criteria led to the exclusion of studies which adopted qualitative and ethnographic. Moreover, studies based on the same sample that appeared in more than one journal or book, were counted as one unique sample study. Finally, despite my effort, some potentially synthesizable studies could not be retrieved. Of the 50 potentially relevant studies that were located in the initial review, 39 studies remained after the inclusion and exclusion step.

B. Coding

After the relevant studies have been identified (n = 39), a coding scheme was developed based on the previous studies and meta-analyses. This scheme helped to sum up the features of individual studies. However, before coding each study's main methodological features, key terms had to be operationalized. In order to determine the strength of the effect, the relationship between the dependent (effect sizes) and independent (instruction type) variables of the study, was examined. Studies employing any form of computer assisted programs and virtual settings at any phase of instruction were coded as CALL. Any other study where instruction was delivered in any form of face to face communication was coded as face-to-face mode (FF).

Finally, in order to create a reliable coding scheme, around 10% of the studies (N = 5) were randomly assigned to a second coder who was a PhD student in linguistics. Each study was independently coded for such features such as publication, participants, methodological, type of instruction, and outcome measures. The results were then compared, and disagreements were discussed until both coders reached an agreement. The average Pearson correlation coefficient was 97%. The final coding scheme was used to collect and summarize comparison categories for all the studies included.

III. ANALYSIS

A total of 116 unique sample studies were included in the meta-analysis from the primary 39 studies, as some experimental studies included several independent experiment samples.

A. Calculations of Effect Sizes

In a meta-analysis, once the data from related studies have been standardized, the study findings can be combined to produce an average effect size, which summarizes a treatment's effectiveness (or any relationship among study variables) across studies [4].

To obtain Cohen's *d*, the mean of one group is subtracted from the mean of the other and is then standardized by dividing by s, which is the sum of squared errors (i.e., take the difference between each score and the mean, square it, and then add all of these squared values up) divided by the total number of scores:

$$d = \frac{M1 - M2}{\sigma}$$

Effect size $= \frac{[mean\ of\ experimental\ group] - [mean\ of\ control\ group]}{standard\ deviation}$

B. Software

For the statistical calculations of the present study, the Comprehensive Meta-Analysis (version 2) software by [8] was used. Although there are several other free and commercial software available to run meta-analyses (e.g., STATA, Revman, MetAnalysis, MetaWin, MIX, RevMan, and WEasyMA, Excel, and SPSS), CMA was more powerful than other programs due to the following 10 reasons:

1. The program helps in coding and organizing data from the literature.
2. It computes and converts effect sizes through over 100 formulas.
3. According to the developers' website [8], CMA allows entering data for each study in its own format. For instance, data can be entered as the number of events and sample size for one study or means and standard deviation for another study, and so on [8].
4. The program allows multiple study designs to be used in the same analysis. Data can be entered from studies that used paired designs, pre-post designs, post-test-only designs as well as or crossover trials [8].
5. All forms of data including correlational, continuous, and binary data can be entered and compared within the same analysis [8].
6. It gives both computational models for one group: Fixed effect, and Random effects.
7. Provides other effect size measures including bias-corrected standardized mean difference (Hedges's g), and raw mean difference.
8. It performs moderator analyses, i.e. determining whether a particular treatment is actually effective or whether there is indeed an association between variables. For example, it is conceivable that the effectiveness of a treatment observed in a particular study depends on the treatment duration or intensity, the characteristics of the sample, the study setting, or the type of outcome measure used.
9. It determines publication bias, and creates forest plots (a graphical representation of a meta-analysis, usually accompanied by a table listing references) and funnel plots (a graph designed to check for the existence of publication bias).
10. Finally, it scored highest on usability criteria in a recent study comparing all meta-analysis software [9].

C. Developing Contrast Categories

In order to calculate Cohen's *d* for each unique sample study, samples were coded either as a control or experimental group. However, there were studies included in the present meta-analysis that did not have a control group or had

World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:12, No:8, 2018

different designs considering pre- and post- tests. The following contrasts were made based on formats available in the CMA:

1. For studies reporting data on a true control group *(n = 23)* and one or more treatment groups, *d* was calculated by contrasting each experimental group with the control group on the immediate post-test.
2. If a study did not involve control group but that reported pre-test and post-test values on a dependent variable *(n = 16)*, effect-size contrasts were drawn between the post-test and pre- test data for each experimental group
3. If the study did not involve a control group or a pre-test but reported posttest values for all experimental groups *(n = 4)*, CMA raw differences were used for both independent and paired groups. However, some previous research such as [4] have considered the instructional condition with the least attention-focused treatment as the baseline comparison group (i.e., treatments involving the processing of experimental input under largely incidental conditions) and each experimental group was therefore contrasted with this baseline group on the immediate post-test.

However, the design of the studies (having a control group or a pre- and posttest) did not guarantee that the authors had reported all the data necessary. Therefore, each study was examined based on the data reported and selected the best formula that CMA offered. Ten different formula formats to calculate Cohen's d were eventually utilized for the 116 sample studies; the most frequent of which was based on means and SDs in each group *(n = 52)* with 44.82% of the calculations. This format was used when the study had a control group and reported pre- and post-data.

The following strategy was adopted for calculating and combining effect sizes:

1. Average effect sizes were calculated for instructional treatment categories identified across studies, focusing specifically on explicit, implicit, CALL, and face to face instruction types. Before averaging, each experimental group was considered as a unique sample study with independent effect size.
2. For studies comparing the effectiveness of two versions or subtypes of the same instructional type, an average effect size was calculated.
3. For all studies reporting pre-test levels on dependent variables, in order to investigate the amount of change observed within studies, average pre- to post-test effect sizes were calculated for instructional treatments and for control/ comparison groups.

D. Publication Bias

Another threat to the reliability of a meta-analysis is publication bias. According to [14], papers with relatively large treatment effects are more likely to publish but studies with non-significant findings or small treatment effects may not be published. As a result, the body of the published research is biased, in that it does not reflect the true magnitude of the treatment effect [8]. In other words, although meta-analysis provides an accurate synthesis of available data, the pool of data may be biased because significant studies were more likely to be published than non-significant studies [8]. In a meta-analysis publication bias is presented in a funnel plot. According to [7, p.330]:

"In a funnel plot, studies with large sample sizes, because of their smaller sampling error and higher precision values, appear toward the apex of the graph and tend to cluster near the mean effect size. Studies with small sample sizes have greater sampling error and lower precision values, so they tend to appear toward the bottom of the graph and are dispersed across a range of values. If there is no availability bias, the studies will be symmetrically distributed around the mean; if availability bias is present, small studies will be concentrated on the right side of the mean. This would mean that small-scale studies with greater sampling error (lower precision values) and lower effect sizes are missing from the data".

The unique sample studies were plotted against the average sample size of study reports. Fig. 1 illustrates the publication bias after the exclusion of outliers. As the funnel plot of this meta-analysis shows, larger sample studies (those with higher precision values) were evenly distributed around the mean and appeared even toward the bottom part of the funnel. However, at the bottom of the plot, there were only a few effect sizes and there were more effect sizes on the right side of the mean than on the left side. This indicates that there was a lack of small-scale studies in the data, and studies with small sample sizes and small effect sizes were not available, i.e., publication bias is present in this meta-analysis with medium and large sample sizes being well represented in the data, but small sample studies being unrepresented. Fig. 2 shows a trim-and-fill analysis performed to search for the missing values that would change the mean effect size if these values (nine darker spots on the left of the funnel) were imputed. It was found that under the Random Effect (RE) model, nine values should be added to the left side to make the plot symmetrical, and imputing these values would change the mean effect size from 1.081 (95% CI = 0.846, 1.316) to 0.76 (95% CI = 0.53, 0.99). In Fixed Effect (FE), model mean effect sizes change from 0.526 to 0.448

In sum, my main comparisons were made based on the effect sizes obtained from 116 unique studies. To judge the magnitude of effect sizes, [10] suggested that effect size of 0.20 is considered to signify a small effect, of 0.50 signifies a medium effect, and of 0.80 indicates a large effect size. Besides the average effect sizes, in order to decide whether the null hypothesis can be rejected or not, the probability level was set at $p < 0.05$ level. Moreover, to decide about the trustworthiness of these calculated effect sizes, the confidence interval around each average effect size was computed through random effect size model. According to [6, p. 187], "if the claim of the effect size for the population falls within the 95% confidence interval, the claim will be correct 95% of the time". The reason for choosing random-effect model was also that the assumption under fixed-effect model (that the true effect size is the same in all studies) could hardly be met (see

World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
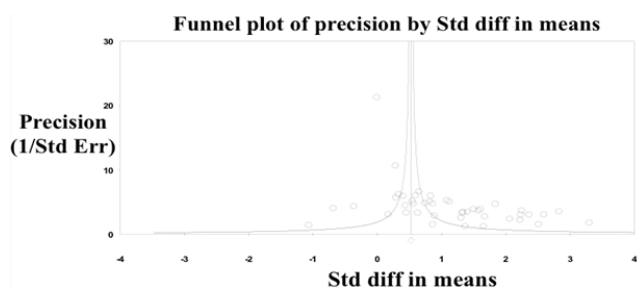Vol:12, No:8, 2018

[11]).

Fig. 1 Funnel plot of precision by effect sizes (of observed studies In Fixed Effects model)
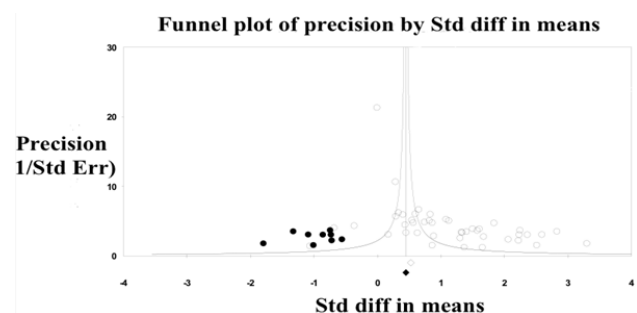

Fig. 2 Funnel plot of precision by effect sizes (of observed and imputed studies in fixed effect model)

IV. RESULTS

Comparing instruction modes resulted in larger effects for CALL (mean d = 1.172) than FF instruction (mean d = 0.965). As mentioned earlier for estimating effect sizes of CALL instruction, CALL was not considered as a main type or method of instruction, rather a mode of instruction. This is due to the fact that technology is only a potential way for designing tasks and delivery of instructional material. Nine studies reported using technology in different ways in instruction which led to 30 unique sample studies. The large effect size (1.172) obtained in this analysis suggests its importance in the current meta-analysis. Table I summarizes the overall effect sizes for FF and CALL instruction mode.

TABLE I
EFFECTS OF CALL VS. FF INSTRUCTION

Instruction	n	Mean ES	SE	Lower CI	Upper CI
CALL	30	1.172	0.154	0.870	1.475
FF	80	0.965	0.097	0.775	1.156

N = number of sample studies, Mean ES = mean effect size, SE=standard error, CI = confidence interval.

The effects of this mode of instruction was even larger than FF mode of instruction (d = 0.965) despite fewer number of sample studies in the analysis (N (CALL) =30, N (FF) = 80). This difference was not statistically significant after Q-test (Q (1) =1.29, p = 0.256). There were also studies on L2 pragmatic instruction that employed other types of instruction, for example [12] examined the effects of task based instruction on pragmatic awareness, or [13] investigate the effects of output instruction on comprehension and production of requests. Fig.
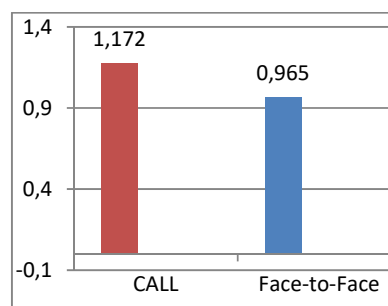
3 illustrates the same results.

Fig. 3 Effects of instructional types and modes on L2 pragmatics (all post-tests included)

V. CONCLUSION

While teachers can follow explicit, implicit or other methods of instruction, they can look for technological tools that best assist them reach instruction goals. However, it is important to know whether this tool makes a difference in pragmatic learning as compared to face to face instruction. Although more sample studies (N (FF) = 80 vs. N (CALL) = 30) had applied FF mode of instruction, CALL was still superior. The large effect size of 1.172 shows that implementing technology in instruction can improve L2 pragmatic development. Technology-informed instruction provides more authentic contexts and materials to the learners. However, dealing with technology as a mode of instruction rather than a method (as the current meta-analysis tried to do) points toward the fact that if activities are managed appropriately, using the technology in L2 pragmatic classes leads toward success.

REFERENCES

[1] T. Khaerudin, "Envisioning the use of technology to teach pragmatics," *IJ-TEFL*, vol.1(1), pp. 60-69, 2012.
[2] M. Farahian, M. Rezaee, and A. Gholami, "Does direct instruction develop pragmatic competence? Teaching refusals to EFL learners of English," *Journal of Language Teaching & Research*, vol. 3(4), pp. 814–821, 2012.
[3] S. L. Chow, "Meta-Analysis of Pragmatics and Theoretical Research: A Critique," *The Journal of Psychology*, vol. 121(3), pp. 259-271.1999.
[4] J. M., Norris, and L. Ortega, "Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis," *Language Learning*, vol. 50, pp. 417–528, 2000.
[5] A. D. Cohen, "Teaching and assessing L2 pragmatics: What can we expect from learners?" *Language Teaching, vol.* 41, pp. 213–235, 2008.
[6] R. Rosenthal, "Writing meta-analytic reviews," *Psychological Bulletin, vol. 118*(2), pp. 183-192. 1995.
[7] S. Li, "The effectiveness of corrective feedback in SLA: A meta-analysis," *Language Learning,* vol. 60, pp. 309–365, 2010.
[8] M. Borenstein, L. V. Hedges, J. P. T., Higgins, and H. R. Rothstein, "*Comprehensive meta-analysis (Version 2)"*. Englewood, NJ: Biostat. 2005.retrieved from https://www.meta-analysis.com/pages/specs.php.
[9] L. Bax, L. Yu, N. Ikeda, and K. Moons, "A systematic comparison of software dedicated to the meta-analysis of casual studies". *BMC Med Res Methodol.* vol. 7(40) doi: 10.1186/1471-2288-7-40. 2007.
[10] J. Cohen, "*Statistical power analysis for the behavioral sciences (2nd ed.)"*. Hillsdale, NJ: Lawrence Earlbaum Associates. 1988
[11] M. Borenstein, L. V., Hedges, J. P. T Higgins, and H. R. Rothstein, (2009). "*Introduction to meta-analysis,"* Hoboken, NJ: Wiley. 2009.
[12] Z. Tajeddin, M. H., Keshavarz, and A. Zand Moghaddam, "The Effect of Task-Based Language Teaching on EFL Learners' Pragmatic Production, Metapragmatic Awareness, and Pragmatic Self-

World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:12, No:8, 2018

Assessment". *Iranian Journal of Applied Linguistics (IJAL), vol.* 15(2), pp.139-166, 2012.

[13] S. Li, (2013). "Amount of practice and pragmatic development of request-making in L2 Chinese", In N. Taguchi & J. M. Sykes (eds.), *Technology in interlanguage pragmatics research and teaching.* Amsterdam/Philadelphia: John Benjamins, pp. 43–70, 2013.

[14] H. R. Rothstein, A. J. Sutton, M. Borenstein, "*Publication Bias in Meta-Analysis: Prevention, Assessment, and Adjustment,*" Wiley Publications. 2005. Retrieved from https://www.meta-analysis.com/pages/pub_bias.php.