ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 10, No. 12, pp. 1558-1565, December 2020
DOI: http://dx.doi.org/10.17507/tpls.1012.07

# An Examination of Corrective, Reflective, and Rule-based Feedback in Chinese Classifier Acquisition in a CALL Environment

Yanhui Zhang
School of Education and English, University of Nottingham Ningbo China

***Abstract*—To facilitate effective learning in a computer-assisted language learning (CALL) environment, it is essential for the system to aid learners to not only pinpoint correct answers, but also identify the right process of learning so as to efficiently overcome various levels of difficulty with optimized practicing items. This study investigates how and to what extent different types of feedback from the CALL system may promote the grammatical knowledge learning for L2 learners of Chinese. Students in the Elementary Chinese program at the Carnegie Mellon University participated in the experiment of the computer assisted language tutoring for learning Chinese classifiers. Three kinds of feedback, namely corrective feedback, reflective feedback, and rule-based feedback, were designed and the relative effectiveness of each on the learning of the planned grammatical knowledge was assessed with pre and posttests. The results show that participants in the rule-based feedback group surpassed those in reflective and corrective feedback groups in an immediate posttest, but participants in the reflective feedback group outperformed the other two groups in a two-week delayed posttest. It is concluded that reflective feedback can more effectively promote self-explanation and memory retention in Chinese classifier acquisition in a CALL environment. The findings provide important insights for the construction of a dynamic, interactive Chinese learning courseware with adequate task design and optimal feedbacks.**

***Index Terms*—Chinese classifier, corrective feedback, rule-based feedback, reflective feedback, learning retention, optimal instructional design, computer assisted tutoring**

## I. Introduction

Design of courseware in a CALL environment has drawn increasingly more attention with the emergence of mobile apps and cloud computing. The CALL platforms have rapidly developed from basic drill and practice programs to today's more dynamic and interactive teaching and learning interfaces accessible from anywhere with internet or wireless coverage. The advancements of technology have also made it possible for CALL to embrace a pedagogy dedicated approach incorporating possibly the learners' metacognitive factors in the courseware design (Colpaert, 2004; Ma & Kelly, 2006; Fischer, 2007; Fan & Ma, 2018). In spite of the multiplying manifestations, the effectiveness of any CALL system is still largely determined by how well it incorporates the theories of second language acquisition (see Chapelle, 1998, 2001, 2005; Colpaert, 2004; Fischer, 2007; Shintani & Ellis, 2013). In particular, feedback from the system plays a significant role in optimizing learning achievement (see, Dick & Carey, 1996; Chapelle 1998, 2001; Mackey, 2006; Vasquez, 2010; Tanaka-Ellis, 2010 for instance). Chapelle (1998) cautioned that CALL research must take learner variables into account. Vanlehn (2006), Tanaka-Ellis (2010) and Cook (2015) also stressed the importance of adding cognitive variables in developing artificial intelligence applications to education. Developing effective CALL materials requires understanding how students learn so that the system's comments and feedbacks will prompt students to construct their own understanding of the subject matter (Vanlehn, 2006). But a critical issue is how the learner may take advantage of feedbacks to refine their cognitive skills.

In a computer learning environment, effective feedback directs learners' attention to the key components of linguistic features (Chapelle, 1998; Hattie & Timperley, 2007), so that such features can be further transferred into long-term memory and learning effectiveness can be enhanced. Ineffective feedback could, on the other hand, distract attention, and hence disrupt the learning process. From the metacognition's point of view (see Moses et al., 2001; Chamot, 2005, for instance), mere exposure or declarative instruction is not enough to help the learner master the knowledge. The learning result would be maximized when the learner is engaged as much as possible in the learning activity. And adequate feedback is one of the most powerful means to enhance the learners' engagement (Mackey, 2006; Brunit et al., 2000). For recent studies addressing the significance of engagement and adequate feedbacks in language acquisition, one can refer to Toyoda and Jarrison (2002), Sun (2009), and Yang (2011), for instance.

Based on the existing models such as connectionism and metacoginition (see, for instance, Broeder and Plurkett (1994) and Basser (1990), for connectionism; and Moses et al. (2001), for metacognition) justifying the importance of feedbacks in enhancing the learning process, we sought to compare the effectiveness of three types of feedbacks in terms of facilitating the learning of Chinese classifiers under CALL environment. One of the key findings is that

reflective feedbacks are more effective for enhancing and retaining the learning achievements in the long run, although rule based feedbacks could be more efficient in intensive learning in short periods. Thus our results will have particular pedagogical significance for dedicated Chinese learners such as college students majoring in Chinese language or those who take regular Chinese courses. The rest of the paper is organized as follows. In section 2, we briefly discuss our research questions and purpose of study. In section 3, we introduce the research methods. In section 4, we present our data analysis and results. In section 5, we summarize the current study and discuss possible future research directions.

## II. Research Objective and Questions

This study attempts to examine whether active feedback could promote second language learners' long-term retention of Chinese grammatical knowledge. An effective feedback in CALL should cater to the learner's cognitive pattern, engage the learner through various learning tasks so as to assist the learner to internalize the linguistic knowledge. It is hypothesized that learners treated with direct right-or-wrong feedback and rule-based instructional feedback (passive feedback) would be surpassed by learners treated with question-and-answer feedback (active feedback) in the posttests when feedback is withdrawn. One heuristic explanation has to do with the role of working memory: didactic right-or-wrong feedback and rule-based feedback do not successfully attract the learner's attention, while the reflective feedback more actively engages the learner in the learning process which facilities knowledge retention in the long-term. Three types of feedback (reflective feedback, rule-based feedback, and corrective feedback) were investigated in terms of how effective they are in promoting Chinese classifier learning. Specifically, the following research questions were proposed:

1. Which type of feedback will optimize the learning result of Chinese classifier learning in a CALL learning environment?
2. Does reflective feedback yield better learning results over rule-based and corrective feedback in the long run?
3. Which type of feedback could contribute to improve the learner's metacognitive awareness?

## III. Methodology and Procedures

### *A. Subjects*

Eighty-two adult learners enrolled in the Elementary Chinese course in the Carnegie Mellon University participated in the study. By the time of test, all subjects had successfully completed the first eight chapters of the Elementary Chinese textbook (Wu, et al., 2006). The grammatical topic of classifier is part of the curriculum for the elementary level learners. All subjects had learned, by the time of experiment, the basic concept of how classifiers should be used in Chinese, but the sets of classifiers used by the computer training program had not been taught to the subjects yet.

The experiments were run in a computer cluster in the Psychology Department at the Carnegie Mellon University. Participants were all registered students in the course Elementary Chinese. But since the data collector could not know in advance the exact number and detailed background information of the participants, they could not be evenly assigned to different types of treatment groups prior to the experiment. As a result, group assignments were done on the spot right before the experiment. The students were randomly and evenly assigned to three groups treated with different types of feedback.

The total number of participants was 82. To eliminate the influence of prior knowledge acquired through heritage background or other Chinese language acquisition sources, not all subjects' results were included in the ANOVA analysis. Based on language background survey and pretest results, the final numbers of subjects distributed to each type of treatment groups are: the corrective-feedback group ($N$ = 14), the reflective-feedback group ($N$ = 20), and the rule-based-feedback group ($N$ = 20).

### *B. Tasks and Materials*

#### *1. Grammar topic – Chinese Classifier*

The Chinese classifier, which is highly frequently used, is one of the special features of the Chinese language. When denoting the number of entities, the number alone cannot function as an attributive but must be combined with a classifier inserted between the number and the noun it modifies. For example,

| Numeral | Classifier | Noun |
|---|---|---|
| yī | běn | shū |
| 一 | 本 | 书 |
| (one) | | (book) |

English also makes limited use of classifiers, such as "pieces" in "three pieces of paper". However, in Chinese, their use is much more pervasive (Sun, 1998, Zhang, 2007).

The Chinese classifier is chosen for the study because this grammar point is unfamiliar to Chinese learners with various language backgrounds so that syntactic transfer from learners' first language can be avoided. Similar to the usage of English preposition, there are certain rules governing which specific classifier to apply before certain category of nouns, although sometimes the correlation between them is not as strong enough as one might straightly deduce ( e.g.

一塊黑板, a blackboard). The great number of details are hard to be mastered by a second language learner because the seemingly random associations between the classifier and nouns indeed have internal rules. For example, the measure words 張, 片, and 塊 are used to modify flat-shape objects, and 個, 顆, and 粒 are used to modify rounded objects. The use of classifiers in Chinese language, to a large extent, reflects human's innate cognitive abilities of categorization and generalization. It can be also explained by and understood through the cognitive-based functional model (Tai & Wang, 1990; Loke, 1996; Wu, 1998). In practice, which classifier to use is determined by the perception of the physical attributes of the noun to be classified. Accordingly, the grammatical rules governing the use of the Chinese classifiers are accredited to represent, at maximum likelihood, such perceptions.

In this study, three types of feedback were adapted to examine how feedback may facilitate learners to establish prototypes for various classifiers. Nouns used in this study are all objects commonly found in daily life. They are also objects of various assortments of shapes, sizes, and functions. Four sets of classifiers are selected. They are listed as follows.

| Objects | | | |
|---|---|---|---|
| long-shape | 條 [tiáo] | 根 [gēn] | 枝 [zhī] |
| flat-shape | 張 [zhāng] | 片 [piàn] | 塊 [kuài] |
| round | 個 [gè] | 顆 [kē] | 粒 [lì] |
| constructions | 座 [zuò] | 間 [jiān] | 所 [suǒ] |

In the experiment, learners watched a picture of a noun, and tried to select a classifier that modifies the object in the picture. While the learner selects a classifier different types of feedback are prompted to suggest if an appropriate classifier is chosen. These different types of feedback are to help learners to identify the mental representations of different objects, build up prototype for objects under the same property, so as to acquire the understanding of different kinds of Chinese classifiers.

*2. Background Survey*

Participants all filled in a Background information survey. The survey collects mainly the learner's language background information such as nationality, other foreign languages learned before Chinese, family language environment, time of exposure to Chinese. Data of Chinese heritage participants are not included in following data analysis of the study.

*3. Tests and Training Tasks of Chinese Classifiers*

The training treatment task is preceded by a brief tutorial that demonstrated and explained how to operate the program and all the online features (i.e., functions of different buttons and the built-in dictionary). The learner got used to the program by practicing to select a classifier that modifies the following noun. The program offers a glossary for Chinese and a countdown timer. When placing the cursor on a Chinese word, the English translation can be shown by the built-in online glossary; when pressing the "Control" key as placing the mouse over a Chinese word, the corresponding Pinyin (phonetic annotation) can be shown. A dynamic countdown timer on the upper right corner of the screen shows the remaining time of training. After 30 minutes, a window pops up to remind that time is up, and the learner is forced to end the training session and proceed to the post-test section.

A pretest measuring prior knowledge of Chinese classifiers was conducted before the participants were trained by the computer. The pretest consisted of 20 items of classifiers. They were presented as 8 items of multiple choice exercises and 12 items of drag-and-drop exercises. The purpose of the pretest was to screen off participants with prior knowledge of Chinese classifiers. In the data analysis process, data of participants who correctly answered more than 50% in the pretest (>10 items) were not included into our statistical samplings.

The treatment tasks consisted of three sets of phrases with classifiers indicating one-dimensional (條，根，枝), two-dimensional (張，片，塊), three-dimensional (個， 顆，粒) objects in different features and sizes, and one set of phrases with classifiers indicating constructions with different features (座, 間, 所). Each set of phrases consisted of three classifiers. And each classifier was presented six times by modifying six different nouns and with feedbacks suggesting whether an appropriate classifier is selected or not. In order to standardize the experimental condition, the treatment period was confined to 30 minutes. All subjects had to finish the tasks within the time given. Otherwise, the learner would be directed to posttest if time was up.

The computer program offered three feedback options to indicate the grammatical accuracy of an answer, namely, reflective feedback, rule-based feedback, and corrective feedback. The reflective feedback suggested whether the learner's selection was correct, and promoted the learner to think about the correct answer by asking them questions related to features of the classifiers used. For example, if a choice was incorrect, a feedback was prompted with "Not quite… Please consider, is …?". Then the button "More answer?" invited the learner to discover more features about the classifier. If the students responded affirmatively, another question was shown to elicit more thinking about the question. If the answer was correct, the computer would then affirm the answer by prompting a message: "Correct! The

grammatical phrase is …". The rule-based feedback indicated if the learner made a correct choice, and displayed the rule for each classifier chosen. For example, if a choice was incorrect, a feedback was prompted with "Not quite… X is used to …"; and if the answer was correct, the computer demonstrated a prompt: "Correct! The grammatical phrase is …". The corrective feedback stated whether the learner's choice is correct and presented the correct phrase. For example, if the learner's answer contained any mistakes, the program prompted the message in English: "Not quite… The grammatical phrase is …"; and if the answer was correct, the computer would display on the screen: "Correct! The grammatical phrase is …".

An immediate posttest without feedback was administered immediately after the training tasks to the three groups of learners. The posttest consisted of 20 items of phrases, 10 of which were learned during the training phase, and 10 of which were new items. They were presented as 8 items of multiple choice exercises and 12 items of drag-and-drop exercises. The purpose of applying new objects in posttest was to force learners to consciously or subconsciously exercise their judgment in the identification of the object, and test whether the learners were able to generalize the rules just learned.

Participants were encouraged to take notes during the entire experimental period, and they were requested to write comments about the training program, e.g. how they would evaluate the training program in general, how helpful they felt the feedback was, how confident they were about the correctness/accuracy of their immediate posttest answers, and their suggestions on how to teach and learn Chinese classifiers. In the data analysis procedure, these self-report and comments were used to investigate the reasoning process and metacognitive awareness the learners applied in the training.

Two weeks after the data collection through computer, a paper-and-pen delayed posttest was administered during the Chinese class instructional time. The delayed posttest was composed by 20 items of phrases, with all classifiers which had been instructed two weeks earlier. The paper-and-pen test was presented similarly to the computerized test: students were instructed to select an appropriate classifier for a noun, where pictures of nouns were placed above. In addition, both English translation and Pinyin of different Chinese characters were showed on the test package. Purpose of the delayed posttest was to check which type of feedback has long-lasting training effects on learners.

### *C. Data Collection Procedures*

Before the data collection process, this study had been approved by the university IRB office. In addition, immediately prior to the data collection in computer, all the subjects read and signed an assent form, agreeing to participate in the study. The study was administered in the CMU Psychology Laboratory, where there were 16 Dell PCs and 2 Macintosh computers. It was explained in detail to all participants about the tasks prior to the start of each section, by using a projector in front of the classroom. Three methods were used to collect the data: a computer record, a paper-and-pen multiple-choice test package, and a paper-and-pen open-end notes and comments. The program maintained a record of students' feedback selections, and glossary help. Furthermore, the amount of time spent with a particular phrase or time spent between feedback selections was also documented. After finishing the computer exercise, the students were required to write their individual comments on the back of the background survey form.

## IV. Data Analysis

### *A. Coding and Screening*

There were altogether 82 students participated in the study, but 28 subjects were screened off from the sample pool in accordance with the following criteria. The first criterion is pretest score. Subjects who selected more than 10 correct classifiers for the nouns were screened from the data pool, since, according to research custom, 50% is the threshold rate indicating whether a score is the result of random guessing or logical thinking. The second criterion is language background. Data from participants who indicated their mother tongue were Chinese or Chinese dialects were not used in ANOVA analysis to avoid the transfer of syntax. The main reason is that the classifier system is similar across Chinese dialects although there are some variations. According to Downing (1996), Japanese language also boasts a rich classifier system. Ideally subjects who had learned Japanese in the past should not be included. But since most of such subjects indicated that their Japanese were at beginner level only, in this study subjects with Japanese language background stayed in the pool unless their pretest score were above 10. After subject screening, the reliability test is run over scores of immediate and delayed posttest. The Cronbach's alpha is positively 70.3%.

### *B. Empirical Result*

#### *1. Inferential Statistics*

A mixed analysis of covariance (Mix ANOVA) was performed on correct classifier scores as a function of feedback types (reflective, rule-based, and corrective) and times of tests (immediate posttest, delayed posttest). The assumption of homogeneity of variance was met by Mauchly's test of sphericity and Box's M test of homogeneity of covariance matrices $F$ (6, 32757.06) = 0.204, $p$ =.976. The assumption of normality was met (Table 1).

TABLE 1
TEST OF NORMALITY

| Tests | Feedback Type | Shapiro-Wilk W | *df* | *p* |
|---|---|---|---|---|
| Immediate Posttest | Reflective | .985 | 20 | .981 |
| | Rule-based | .923 | 20 | .112 |
| | Corrective | .965 | 14 | .808 |
| Delayed Posttest | Reflective | .959 | 20 | .521 |
| | Rule-based | .963 | 20 | .611 |
| | Corrective | .930 | 14 | .309 |

However, there were no significant differences on the correct classifier scores among the three feedback types $F$ (2, 53) = .183, $p$ = .833, partial $\eta^2$ = .007, and there were no significant differences on the scores between immediate and delayed posttests $F$ (1, 53) = .564, $p$ = .456, partial $\eta^2$ = .011.

*2. Descriptive Statistics*

Although due to the small size of data pool and uneven number in each feedback group, the mixed ANOVA results turn out to be not significant, the descriptive statistics in immediate posttest and delayed posttest show very interesting trend of the three feedback treatments. As seen from the comparison table (Table 2), although the mean score in reflective feedback (M = 9.60) was the weakest among the three groups in immediate posttest, its mean score became the strongest one in the delayed posttest (M = 10.25) two weeks later. It is even more interesting to find out that the reflective feedback group earned an even higher mean compared with the result in immediate posttest.

TABLE 2
DESCRIPTIVE STATISTICS OF FEEDBACK TYPES IN DIFFERENT POSTTESTS

| | Feedback Type | N | Mean | Std. Deviation |
|---|---|---|---|---|
| Immediate Posttest | Reflective | 20 | 9.60 | 3.676 |
| | Rule-based | 20 | 10.35 | 3.884 |
| | Corrective | 14 | 10.00 | 3.374 |
| | Total | 54 | 9.98 | 3.626 |
| Delayed Posttest | Reflective | 20 | 10.25 | 3.567 |
| | Rule-based | 20 | 9.80 | 3.350 |
| | Corrective | 14 | 8.86 | 3.348 |
| | Total | 54 | 9.72 | 3.412 |

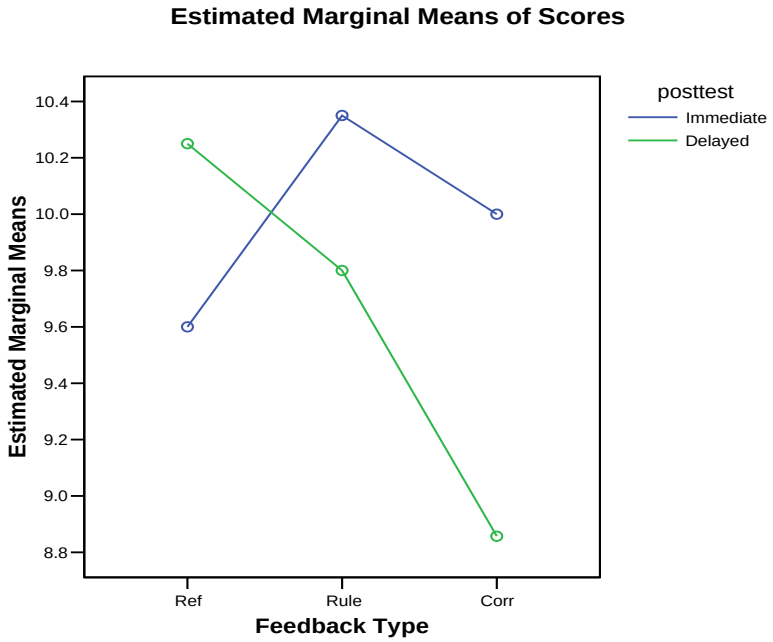


Figure 1: Means of Scores on Different Feedback Types in Two Posttests

The result to some extent shows that reflective feedback does promote long- term retention in the task of Chinese classifier acquisition. Rule-based feedback dogmatically teaches the syntactical rules to students. Since the learner did not actively involve in the learning process, the rules are forgotten relatively soon. However, compared with the

reflective and corrective feedback groups, participants in the rule-based feedback group still show relatively stable performance in the immediate posttest and delayed posttest. Reflective feedback promotes the learner to actively scaffold the syntactic knowledge thus the knowledge is internalized and retained in the memory for a longer time. In contrast, performance of the corrective feedback group dramatically declines. As one possible explanation, students in the corrective feedback group might have relied heavily on rote memory in the immediate posttest. But because the syntactic rules were not well-formed yet, memory of classifiers showed a rapid decay in the delayed posttest.

Correlations between the time participants spent on classifier training and scores of immediate posttest and delayed posttest shed a deeper insight from the data. As indicated by the results from the reflective feedback group, there are significant correlations between the time students spent in the training and the achievements, respectively, in the immediate posttest ($r$ = .467, $p$ = .038), and the delayed posttest ($r$ = .464, $p$ = .04). This shows that students who are actively engaged in the problem solving process have more successfully inferred and internalized the rules, and thus showed more achievement than others. There also shows a strong correlation between achievement scores and time spent in training under the corrective feedback ($r$ = .59, $p$ = .026). The reason should be that although corrective feedback did not provide rules, correct forms of classifiers were supplied to the learner whenever an error occurred. This type of feedback provides learner more exposures of correct usage of classifiers. As a result, rote memory did boost sizable gains in the immediate posttest. But since the rules were not internalized, such gains also decayed in the delayed posttests as memory waned.

TABLE 3
TIME FOR TRAINING

| Feedback Type | Mean of Time | Std. Deviation | N |
|---|---|---|---|
| Reflective Feedback | 15.5655 | 5.92498 | 20 |
| Rule-based Feedback | 13.0550 | 5.08283 | 20 |
| Corrective Feedback | 12.2729 | 3.89280 | 14 |

TABLE 4
CORRELATIONS BETWEEN TIME FOR TRAINING AND POSTTESTS

| Feedback Type | | | Time for training |
|---|---|---|---|
| Reflective Feedback | Immediate Posttest | Pearson Correlation | .467(*) |
| | | Sig. (2-tailed) | .038 |
| | Delayed Posttest | Pearson Correlation | .464(*) |
| | | Sig. (2-tailed) | .040 |
| Rule-based Feedback | Immediate Posttest | Pearson Correlation | -.154 |
| | | Sig. (2-tailed) | .516 |
| | Delayed Posttest | Pearson Correlation | .119 |
| | | Sig. (2-tailed) | .619 |
| Corrective Feedback | Immediate Posttest | Pearson Correlation | .590(*) |
| | | Sig. (2-tailed) | .026 |
| | Delayed Posttest | Pearson Correlation | .364 |
| | | Sig. (2-tailed) | .201 |

## V. DISCUSSION AND CONCLUDING REMARKS

The ability to provide feedback on individual responses is one of the main advantages of CALL. But as described by Cohen (1985) Vasquez (2010), and Tanaka-Ellis (2010) for instance, feedback is one of the most instructionally powerful and least understood features in instructional design. The rapid growth of computer and artificial intelligence technology allows courseware designers to incorporate more feedbacks into their programming. Through computer assisted language tutoring as well as paper-and-pencil tests, the study compared and contrasted three kinds of feedback, namely corrective feedback, reflective feedback, and rule-based feedback, and examined the impact of these types of feedback on the effectiveness of Chinese syntax learning.

The direct instruction of syntactic rules can quickly fill the void of concept in the learner's mind with the correct forms. This didactic way of instruction, however, may not draw sustained enough attention from the learner for successful knowledge retention. Compared with the other two types of feedback, reflective feedback adapts question-and-answering format to guide the learner to discover the shared property of objects, and thus infer the rules by themselves. Obviously the self-explaining way of learning carries more cognitive load to learners during the learning process. But it provides an avenue to have the learner actively involved in the task. Results show that learners in the reflective feedback group did not perform as well as those in the rule-based feedback and corrective feedback groups in

the immediate posttest, but outperformed the other two groups in the delayed posttest. Furthermore, the reflective feedback group shows better achievement in delayed posttest compared with the immediate posttest. The findings highlight the importance and effort-taking of knowledge internalization in language learning. But once the knowledge is adequately assimilated by the learner, it would retain in the memory and won't easily fade away. The pedagogical implications drawn is that, in a CALL teaching environment or in a second language classroom, courseware designers should differentiate the levels of difficulties of the learning materials, and provide more adequate self-explaining conditions so as to foster the learner's metacognition and motivate more self-initiative engagement in the learning process.

One limitation of the current study is that participants were randomly assigned to each treatment group according to the positions they sat in the Psychology Lab. Among them 25% of the original subjects were Chinese heritage students and were not included in the data pool. This sample screening resulted in an unbalanced group size, which may partially impede the significance of the statistical analysis. As for the variable of metacognitive awareness, to our best knowledge, there is not yet a standardized questionnaire to measure individual metacognition since it is rather difficult to classify and quantify the very broad spectrum of the constituents of metacognition. As a result, the relationship between types of feedback is qualitatively assessed based on participants' self-report.

One explorable future direction will be to add more variables to the present study, examining various facets of feedback and taking more cognitive and social variables into account. Cognitive parameters such as working memory span and megacognitive awareness of individual learners can be included in the experiment design. Quantity and frequency of feedback in terms of immediate vs. delayed feedback, and frequent vs. intermittent feedback are also of interest for further investigation. Finally, it would be beneficial to investigate the role of language background, general language proficiency, and learning attitudes on second language acquisition in a CALL environment.

## REFERENCES


[1] Broeder, P., & Plunkett, K. (1994). Connectionism and second language acquisition. In N. Ellis (Ed.) *Implicit and explicit Learning of Languages.* London: Academic Press, 421-454.

[2] Brunit, S., Huguet, P., & Monteil, J. M. (2000). Performance feedback and self-focused attention in the classroom: When past and present interact. *Social Psychology of Education*, 3, 277–293.

[3] Chamot, A. U. (2005), Language learning strategy instruction: current issues and research, *Annual Review of Applied Linguistics*, 25, 112-130.

[4] Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, 2 (1), 22-34.

[5] Chapelle, C. A. (2001). Computer applications in second language acquisition. Cambridge University Press, Cambridge.

[6] Chapelle, C. (2005). Computer-assisted language learning. In E. Hinkel (Ed.), *Handbook of second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum, 743-755.

[7] Colpaert, J. (2004). Design of online interactive language courseware: Conceptualization, specification and prototyping, research into the impact of linguistic-didactic functionality on software architecture. Antwerp, Belgium: University of Antwerp.

[8] Cook, J. (2015). Negotiation for Meaning and Feedback among Language Learners. *Journal of Language Teaching and Research*, 6 (2), 250-257.

[9] Dick, W., & Carey, L. (1996). The Systematic Design of Instruction. Harper Collins College Publishers, 4th. ed., New York.

[10] Downing, P. (1996), Numeral Classifier Systems: The Case of Japanese, John Benjamins Publishing Co., Philadelphia.

[11] Fan, N., & Ma, Y. (2018). The Role of Written Corrective Feedback in Second Language Writing Practice. *Theory and Practice in Language Studies*, 8 (12), 1629-1635.

[12] Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *Computer Assisted Language Learning*, 20 (5), 409-42.

[13] Hattie, J., & Timperley, H. (2007). The power of feedback. *Review on Educational Research*, 77 (1), 81-112.

[14] Loke, K. (1996). Norms and realities of Mandarin shape classifiers. *Journal of the Chinese Language Teachers Association*, 31(2), 1-22.

[15] Ma, Q., & Kelly, P. (2006). Computer assisted vocabulary learning: design and evaluation. *Computer Assisted Language Learning*. 19 (1), 15-45.

[16] Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27 (3), 405-430.

[17] Moses, L. J., & Baird, J. A. (2001). Metacognition. In Robert A. Wilson and Frank C. Keil (Ed.), *The MIT Encyclopedia of the Cognitive Sciences.* Cambridge: the MIT Press.

[18] Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286-306.

[19] Sun, C. (1998). The discourse function of numerical classifiers in Mandarin Chinese. *Journal of Chinese Linguistics*, 2, 298-322

[20] Sun, Y. C. (2009). Voice blog: an exploratory study of language learning. *Language Learning & Technology*, 13(2), 88-103.

[21] Toyoda, E., & Harrison, R. (2002). Categorization of text chat communication between learners and native speakers of Japanese. *Language Learning & Technology*, 6(1), 82–99.

[22] VanLehn, K. (2006). The behaviors of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16 (3), 227-265.

[23] Vasquez, C. (2010). Raising teachers' awareness about corrective feedback through research replication. *Language Teaching Research*, 14 (4), 421-443.

[24] Wu, S. (1998). The role of size and shape in three-dimensional classifiers in Mandarin Chinese. *Proceedings of the Ninth North*

*American Conference on Chinese Linguistics (NACCL-9)*, 310-323.
[25] Wu, S., Yu, Y., Zhang, Y., & Tian, W. (2006). Chinese link: Zhong wen tian di: elementary Chinese, New Jersey: Pearson Education, Inc.
[26] Yang, Y. (2011). Engaging students in an online situated language learning environment. *Computer Assisted Language Learning*, 24 (2), 181-198.
[27] Zhang, H. (2007). Numerical classifiers in Mandarin Chinese. *Journal of East Asian Linguistics*, 16, 43-59.

**Yanhui Zhang** is affiliated with School of Education and English, University of Nottingham Ningbo China. Dr. Zhang's research is in applied linguistics, with specializations in computer-assisted language learning, Chinese-English bilingual literacy development, and corpus linguistics and quantitative linguistics.