

# English Article Errors in Taiwanese College Students' EFL Writing

Neil Edward Barrett\* and Li-mei Chen\*

## Abstract

The English articles, *the*, *indefinite a/an*, and *zero* can often be troublesome for English language learners to master, especially in longer texts. Thomas (1989) demonstrated that English as a second language (L2) learners from first languages (L1) that do not have the equivalent of an article system encounter more problems using articles. Ionin and Wexler (2004) found that such learners fluctuate between the semantic parameters of definiteness and specificity. This study examines English L2 article use with Taiwanese English learners to determine the potential factors influencing English article substitution and error patterns in their academic writing. This corpus-based analysis used natural data collected for the Academic Writing Textual Analysis (AWTA) corpus. A detailed online tagging system was constructed to examine article use, covering the semantic (specific and hearer knowledge) as well as the other features of the English article. The results indicated that learners overused both the definite and indefinite articles but underused the zero article. The definite article was substituted for the indefinite article in specific environments. Although no significant difference existed between specific and non-specific semantic environments in zero article errors, a significant difference emerged between plural and mass/non-count nouns. These results suggest that, in regard to writing, learners need to focus on the semantic/pragmatic relationships of specificity and hearer (or reader) knowledge.

**Keywords:** Definite Article, Indefinite Article, Zero Article, Hearer Knowledge.

## 1. Introduction

The use of cohesive devices in writing is a well-researched topic in second language acquisition research, taking on a greater significance in recent years as increasingly more

---

\* Department of Foreign Languages and Literature, National Cheng Kung University, 1 University Rd., Tainan, Taiwan. Telephone: (06)2757575 ext. 52231

E-mail: leemay@mail.ncku.edu.tw

The author for correspondence is Li-mei Chen.

students are being asked to present their work in English, thus pointing to the need for greater accuracy and cohesion. Errors within the article system (*i.e.*, *a*, *an*, *the*, and the zero article) have been noted in studies examining L2 learners' writing, and such errors are present in advanced learners' texts as well (Lee, 2007). To put this issue into perspective, a corpus study of 668 TOEFL essays from Chinese, Japanese, and Russian students found that 13% of sentences-or 1 in every 8 noun phrases-had article errors (Han, Chodorow, & Leacock, 2006).

In written discourse, the omission of an article or the use of the wrong article may cause some ambiguity for the reader, especially when the writer wants to identify a noun anaphorically/cataphorically or assume reader/writer familiarity. Halliday and Hasan (1976), in their time-honored investigation into cohesion, pointed out that, "Whenever the information is contained in the text, the presence of an article creates a link between the sentence in which it occurs and that containing the referential information; in other words, it is cohesive" (p. 74). Therefore, the use of articles creates an understanding between the writer and reader, enabling the reader to locate where a noun or noun phrase is located as well as identify if it is already understood as known by the reader.

In addition to the above, the use of the English article involves the integration of semantic, pragmatic, and grammatical functions, as no one-to-one form-function mapping exists for each article, creating a large number of rules for students to master. In terms of native Mandarin-speaking English learners, article errors have been found to be cohesive writing errors in research by Chen (2002), Chiang (2003), and Ting (2003).

Research into article errors has revealed that English article errors may be due to an inability to acquire the semantic feature of specificity (Ionin & Wexler, 2004; Snape, 2006), resulting in the overuse of the definite article in specific environments. On the other hand, it may be a pragmatic deficit (Diez-Bedmar & Papp, 2008) when learners overuse the definite article due to extra-linguistic features, such as world knowledge. Other studies (Goto-Butler, 2002; Snape, 2008; Yoon, 1993) have investigated noun countability in terms of its influence on article errors.

Although previous research has examined Mandarin English L2 article use in spoken discourse (Moore, 2004; Robertson 2000) or article use in a cloze test (Lee, 2007; Snape, 2009), only Diez-Bedmar and Papp (2008) have investigated texts from native Mandarin speaking English learners. They concluded that native Mandarin-speaking English learners have both a grammatical and a pragmatic deficit. Nevertheless, in their study, the essays were completed with time restrictions placing constraints on the writer, which may have resulted in more article errors.

The aim of this study is to identify the features that influence students' article use or misuse. We first noticed the frequency of article errors in undergraduate writing while tagging

cohesion errors for the Academic Writing Textual Analysis (AWTA) corpus, an online corpus of Taiwanese undergraduate writing. Although the article errors did not seriously impair communication, they interrupted the cohesion of the writing. Consequently, it was felt that the reasons for these errors deserved further attention.

To investigate the factors that influence article errors, this study asks the following questions:

1. What is the influence of specificity and definiteness on the English article substitution and error patterns in the academic writing of Taiwanese college students?
2. What other potential factors influence English article substitution and error patterns?

## 2. Literature Review

English has three articles, the definite, indefinite, and zero, which have a wide range of semantic and syntactic functions in discourse (Moore, 2004). A widely-used theory related to English article use is the semantic wheel (Bickerton, 1981). According to Bickerton, English noun phrases (NPs) can be classified according to two semantic features: specific reference [+/-SR] and hearer knowledge [+/-HK]. Table 1 illustrates the four NPs. Many studies have shown that the failure to recognize [HK] has led to article errors in article production tasks (Lee, 2007; Robertson, 2000) and cloze tests (Goto-Butler, 2002; Trenkic, 2008).

**Table 1. Bickerton's noun phrase environments (Goto-Butler, 2002, p. 478)**

Noun phrase environment	Example
[-SR, +HK], ( <i>the, a, zero</i> ): Generics.	A cat likes mice. <i>The</i> whale is a mammal. ( <i>zero</i> ) Language is a great invention of human kind.
[+SR, +HK], ( <i>the</i> ): Unique, previously mentioned, or physically present referents.	When I found a red box in front of my house, it was too late. <i>The</i> box blew up with a terrific explosion. This book did not sell well even though <i>the</i> author was a famous writer.
[+SR, -HK], ( <i>a, zero</i> ): First-mention NPs or NPs following existential "has/have" or "there is/are."	There is <i>a</i> new version of the I-phone. Did you see it? I keep sending ( <i>zero</i> ) messages to him.
[-SR, -HK], ( <i>a, zero</i> ): Equative NPs or NPs in negation, question, or irrealis mood.	He used to be <i>a</i> lawyer. ( <i>zero</i> ) Foreigners would come up with a better solution to this problem.

A more recent development in article system research was proposed by Ionin and Wexler, (2004). Based on their studies of Russian and Korean-two languages that do not have an article system-these authors proposed that articles are governed by semantic parameters. Their theory is known as the article choice parameter.

Evidence for this comes from languages such as Samoan, which has different articles to indicate if a NP is specific or non-specific. English does not have the [+/-specific] setting, but instead has the definiteness setting [+/- definiteness]. Samoan uses the article *le* with specific noun phrases and *se* with non-specific, but does not mark definiteness (Ionin & Wexler, 2004).

The Samoan data analyzed by Ionin and Wexler demonstrate that definiteness may be irrelevant in languages like Samoan. Thus, the authors proposed the article choice parameter, which states that, “A language that has two articles distinguishes them as follows: The Definiteness Setting: Articles are distinguished on the basis of definiteness; The Specificity Setting: Articles are distinguished on the basis of specificity” (Ionin & Wexler, 2004, p. 12).

For [-] article languages, the authors proposed the fluctuation hypotheses, which states that learners fluctuate between the two parameter settings until they have enough input and the settings stabilize. Moreover, L2 learners may adopt parameter settings not found in their L1 or their L2 because, if an L2 learner lacks articles in his/her L1, no language transfer should occur as there should be no parameter preference (Ionin & Wexler, 2004). Thus, if languages, such as Mandarin Chinese, are seen as having neither articles for definiteness nor specificity, learners should fluctuate between the two settings for definite and specific reference. Based on this, Ionin & Wexler (2004) made specific predictions for [-] article L2 learners (see Table 2).

**Table 2. Definite and Indefinite Fluctuation Hypothesis Predictions (Snape, 2009, p. 32)**

Semantic type	+ definite	-definite
+ specific	Correct use of <i>the</i>	Overuse of <i>the</i>
-specific	overuse of <i>a</i>	Correct use of <i>a</i>

Although studies indicate that the fluctuation hypothesis correctly predicts L2 output (Snape, 2009), it has been criticized for several reasons. First, the fluctuation hypothesis does not take the zero article into account. For many first mention mass and plural nouns, specificity-as in first mention singular nouns-can be a semantic feature of zero article NPs, so the fluctuation hypothesis should also be able to predict these error types. Furthermore, Snape (2008) pointed out that, in both his and in Ionin and Wexler’s studies, individual patterns among participants do not fit into either the definiteness pattern used by L1 English or the proposed fluctuation patterns. Instead, individual learners showed a miscellaneous pattern, whereby article errors occur in all four semantic types [+/-definite, +/-specific].

## 2.1 The Definite Article in English

Hawkins (1978) initially based his location theory on previous article studies and subsequently revised his theory (Hawkins, 1991). Hawkins identified eight different types of definite articles. By using *the*, a writer or speaker asks the reader/listener to locate the referent using knowledge that is available in the text (anaphoric and associative anaphoric use), can be sensed in the vicinity (visible and immediate situation use), or is available from local or general knowledge (immediate and local situation use). The other types of use-what Hawkins (1978) called 'structural information,' which refers to prepositional phrases, relative clauses, or adjectives-help locate the referent.

In 1991, Hawkins revised his location theory based on theories of pragmatics developed by Grice (1989). According to Hawkins, the referents are located in pragmatic sets (p-sets) that are available to the speaker/hearer via discourse sets that contain information about a certain situation or event. These p-sets are associated knowledge shared by the discourse participants and can be accessed from present or prior discourse, the local environment, shared knowledge, or general knowledge. The main point of the p-sets is that they allow the hearer or reader to accept information as definite.

## 2.2 Definiteness in Mandarin Chinese

A major difference between English and Mandarin Chinese is that English is a language that uses articles to show that a noun phrase is definite/or indefinite whereas Mandarin generally lacks articles (Snape, 2009). In Mandarin Chinese, a bare noun (with no classifier, demonstrative, or numeral) can be definite, indefinite, or generic. Classifiers can be defined as: "One of a set of specialized grammatical form constituents of certain types of noun phrases, especially those containing numerals, the choice of classifier being determined by the semantic characteristics of the head noun" (Trask, 1995, p.44). For example, classifiers include *ge* 個 and *ke* 棵 and they are a salient feature of Mandarin Chinese.

According to Cheng and Sybesma (2005) this semantic reading is dependent on the predicate. The following examples illustrate this (Cheng & Sybesma, 2005):

1. *Hu2fei1 mai3shu1 qu4 le* 胡飛買書去了 Hufei buy book go = Hufei went to buy a book/books (indefinite).
2. *Hu2fei1 he1 wan2-le tang1* 胡飛喝完了湯 Hufei drink-finished soup = Huefei finished the soup (definite).
3. *Wo3 xi3huan1 gou3* 我喜歡狗 I like dog = I like dogs (generic).

In preverbal position, bare noun phrases receive a definite or generic interpretation. Noun phrases with a classifier but no numeral only receive a nonspecific interpretation.

4. *wo3 xiang3 mai3 ben3 shu1* 我想買本書 I want buy CL book = I would like to buy a book

(any book, nonspecific).

Noun phrases with both a number and a classifier can have either a specific or non-specific reading.

5. *wo3 xiang3 mai3 yi1-ben3 shu1* 我想買一本書 I want buy one-CL book = I would like to buy a book. (non specific)

6. *Ta1 he1-wan2-le yi1-wan3 tang1* 他喝完了一碗湯 He drink-finished one-CL soup = a finished one/a bowl of soup (specific). (Cheng & Sybesma, 2005).

Definiteness in Chinese can be marked by a demonstrative and a numeral (Li & Thompson, 1981), which also gives the noun phrase a deictic function (Wu & Bodomo, 2009). The following examples illustrate this.

22. *Nei4 ben3 shu1 wo3men dou1 du2gou4* 那本書我們都讀過 That CL book we all read = as for/the/that book, we have all read it. (Wu & Bodomo, 2009).

Definiteness is also marked in Mandarin by word order, as Mandarin is a topic-prominent language. This means the topic appears sentence initial and shows either known information or generic uses, such as referring to an entire class of objects. The second part of the sentence is the comment, which contains new information (Moore, 2004). If a noun is preverbal, it is usually definite regardless of the use of the demonstratives *na4* 那 (“that”) or *zhe4* 這 (“this”). In addition, nouns that take the classifier *yi1* 一 (“one”) usually do not appear in the topic position, making the sentence indefinite. If a subject is post verbal and without the demonstratives, it is indefinite; therefore, if a subject appears before the verb without a demonstrative, it is perceived to be definite (Moore, 2004).

The issue of noun countability for many Chinese dialects has divided scholars, with some claiming that all Mandarin nouns are mass nouns (Chierchia, 1998; Wu & Bodomo, 2009), while others argue that Mandarin Chinese has both mass and count nouns. Chierchia (2008) argues that all nouns are treated as mass nouns; therefore noun countability would have to be learnt. Others (Cheng & Sybesma, 2005; Zhang, 2007) argue that Mandarin has both count and mass nouns with count and mass classifiers.

### 2.3 English Articles in Second Language Acquisition Studies

Numerous studies in second language acquisition (SLA) research have examined English articles, starting with Brown (1973). Research has indicated that both young L1 children and L2 learners tend to associate the definite article with specific contexts rather than hearer/discourse knowledge. This became known as *the* flooding, whereby a beginning learner overuses the definite article in all article contexts. Chaudron and Parker (1990) found evidence that English learners misused articles in specific, discourse-first locations. Using Huebner’s (1983) noun types, Thomas (1989) investigated whether L2 learners overused *the* in [+SR-HK]

(first mention) contexts. Interestingly, unlike earlier L2 article acquisition studies (Huebner, 1983), Thomas's participants did not show any significant signs of 'the flooding,' but the learners did overuse the definite article. The learners also over-generalized the zero article, although it was not clear whether the learners had failed to use this article or had made an explicit article choice, as the difference was impossible to detect without interviewing the participants. Master (1997) investigated how acquisition differed between English L2 learners from article-less L1s (Japanese) and L1s with articles (Spanish). He found that *the* flooding was more dominant in the Japanese subjects; *a/an* acquisition was also delayed for these subjects. These studies further demonstrated that L1 had an effect on article acquisition and that learners with L1s lacking articles had more difficulty acquiring the English article system. This has been confirmed in studies by Trademan (2002) and by Diez-Bedmar and Papp (2008).

In terms of native Mandarin-speaking English learners, Moore (2004) found that intermediate/advanced learners tended to overuse the indefinite article in both a cloze test and an oral narration task. Most of the indefinite errors occurred in *a for the* errors during the cloze test, but *zero/the* accuracy was almost the same during the narration task. Lee (2007) investigated advanced English L2 Mandarin speakers studying at the PhD level in the United States. Unlike other studies, Lee's research looked at error patterns in an online forum and a cloze test based on the findings of the online forum. The learners tended to omit rather than overuse the indefinite and definite articles in the online forum, but overused the definite article in the cloze test. For definite article error types, *the for ∅* errors were more common in front of unique common nouns and in specific contexts. More recently, (Diez-Bedmar & Papp, 2008) carried out a corpus study into article acquisition in Spanish and Chinese English L2 learners. The definite article was overused in specific contexts, but the zero article was also overused, demonstrating issues related to noun countability. They suggested that the overuse of the definite article was a pragmatic problem, as the writers did not consider the readers knowledge, while noun countability was seen as a grammatical problem.

## 2.4 Noun Countability and English Articles in SLA Studies

Noun countability has been an issue in article acquisition, especially for languages that do not use an article system (Goto-Butler, 2002; Hua & Lee, 2005; Lee, 2007; Master, 1997; Moore, 2004; Snape, 2008; Yoon, 1993). Using a cloze test, Yoon (1993) found that Japanese learners had problems with *indefinite for zero* errors, especially with mass nouns. Goto-Butler (2002) found that noun countability was also a source of errors with Japanese participants. Lower proficiency participants encountered problems with mass and count nouns, but higher level participants also had problems with countability-especially in nouns where the countability was context-dependent. Goto-Butler (2002) suggested that these errors with context-dependent

nouns often cause definite article errors, as the listener depends on noun countability to determine if a noun is unique within a set. The noun *culture* is an example. Goto-Butler (2002) pointed out that culture is often indivisible, so it can be seen as uncountable. Nevertheless, the NP *old culture* belongs to a set of old cultures that need to be identified. When introducing the NP *old culture*, it must be introduced using the indefinite article. Goto-Butler's participants often introduced the phrase "Japan has **an** old culture" with the definite article (*i.e.*, "Japan has **the** old culture"), believing that Japanese culture was identifiable as definite (Goto-Butler, 2002). This problem with abstract nouns may lead to problems with definite article use.

Snape (2008) found that native Japanese-speaking English learners made more errors with the definite article within plural and mass contexts compared to singular contexts. For Mandarin L1s, Hua and Lee (2005) found that participants were able to distinguish between countable and uncountable nouns in English L2 and were more accurate with abstract nouns. Lee (2007) did not find any relationship between definite article errors and noun countability with her Mandarin-speaking high level participants, but did find errors with indefinite articles and noun countability. The learners often failed to use an indefinite article with countable singular nouns and failed to judge if a noun had a countable or uncountable reading.

### 3. Methodology

A total of 30 students participated in this study. The subjects were third-year university students who had attended writing class with the same instructor for four semesters. These participants were chosen for several reasons. Participants who had taken a writing class with the same instructor were needed in order to avoid the effect of differing writing instruction. In addition, all participants had received the same length of writing instruction. Although an earlier pilot study had shown no longitudinal effect, some of the essays may have been too short to provide an adequate amount of tokens; thus, it is possible that longitudinal changes could affect article accuracy. In order to control for this, the participants had to be students who had a similar amount of exposure to writing instruction.

The corpus consisted of 30 argumentation essays, with a total of 28,020 words. Only 30 essays were coded due to time limitations. The article types and error types had to be coded manually, as no automatic parsing had been developed to deal with the multiple functions of the article system. The pilot study revealed that shorter essays did not contain enough articles and article errors. Therefore, argumentation essays were coded, because their lengths ranged from 789 to 1,449 words, resulting in a mean of 980 words per essay. The original drafts of the essays were coded because they had not been corrected by the student, instructor, or peers.

To explore article use and error patterns in Taiwanese students' EFL academic writing, a coding scheme was developed to annotate the data with linguistic information. The coding scheme was based on a modified version of the one used by Moore (2004). Other



corpus-based article coding schemes were examined, such as Han *et al.* (2006), who used the syntactic position of the NP in an automated system. Although Han *et al.*'s approach provided a general account of the errors over a large corpus; it did not include the semantic environment of the noun phrases, making it unsuitable for the current investigation. Neff *et al.* (2007) used the Spanish International Corpus of Learner English (SPICLE) corpus to investigate definite, indefinite, and zero articles, and this effectively described the general differences between the three articles; however, it was not comprehensive enough for the current study because it did not provide information on the semantic and pragmatic features of the English articles. Diez-Bedmar and Papp (2008) used Huebner's (1983) semantic environments to investigate article use in Spanish and Mandarin speakers' English writing; although their study bears some similarities to the present one, it did not investigate the use of the eight definite article types described by Hawkins (1978), which were needed for a related study into English article accuracy.

After investigating these other schemes, Moore's taxonomy (2004) was found to be the most comprehensive system, as it was based on article research conducted by Hawkins (1978) and Robertson (2000). This coding scheme has many advantages over the other schemes used in SLA article research because it combines the semantic environments, the definite article types identified by Hawkins (1978), and the language transfer features described by Robertson (2000). Although this scheme follows the procedure described by Moore (2004), it was sometimes necessary to make some adaptations or collapse some of Moore's categories.

In terms of coding, Figure 1 shows a brief diagram of the actual data as they would appear in the window of the corpus. The tagging system and AWTA corpus are described in detail in Kao and Chen (2009). The first pair of brackets indicates the meta-linguistic tag used in the corpus, and the annotation shows either the article type or the error type after the equal sign. The original text is in the arrowed brackets, followed by the meta-linguistic information to make the tags clear in the reviewing process. The tagging system works as follows. Inside the brackets is the name of the article (*e.g.*, the semantic or article type); information regarding whether it is used correctly is indicated by the letters Y or N, which represent correct and incorrect use. This is followed by a number indicating the general error type. For example, in <tag D PN N annotation="2">, the D is a definite article, PN stands for plural noun, which is the error type, N indicates an article error, and 2 is the code number for *definite for zero specific* errors. In this way, the article error can be identified first and meta-linguistic information can be added afterward. Figure 1 is an extract taken from the AWTA corpus. The tag <tag D IA N annotation="5"> indicates a definite for indefinite article substitution. The D is a definite article, IA stands for indefinite article, which is the error type, and 5 is the code number for *the for specific indefinite a/an* errors.

Many studies have showed that it would be better for the hearing disabled to have <tag D IA N annotation="5">the</tag> cochlear implant at an early age. Also, if implanted the cochlear implant at the age one to two, their language learning could come out of great improvement. However, the situation now seems that the elder people who are more than 55 years old, are not suitable to have the cochlear implant. They are usually told only to use <tag D PN N annotation="2 ">the</tag> hearing aids for that most people think it would be too late for them to have the implantation.

***Figure 1. Annotating meta-linguistic information.***

To deal with the repetition of an NP, which is often necessary in writing due to its cohesive function (Trademan, 2002)-although some overuse or repetition can be interpreted as an immature writing style-a types/token distinction was used. Here, token counts refer to the frequency of a particular word or phrase whereas type refers to the occurrence of a distinct word or phrase in a text. In terms of errors, token counts would record the same error throughout the text, whereas type frequency would only record a mistake once. Therefore, if tokens were classified as errors, it would present an inflated picture. This paper only coded the types to avoid inflating the number of errors.

Once the coding procedures were decided, the data were coded for errors, as article errors are often discourse-dependent, making it necessary to read the essays first without the distraction of tagging every English article. All of the errors were highlighted and subsequently coded according to their error type. Next, the essay was coded for article use, starting with the definite article, followed by the indefinite, and finally the zero article. This was done to collect information for related research into L2 English article use. The annotation system consisted of two main parts: the semantic and pragmatic relations of each article and a description of the common error patterns.

Article error types can tell a researcher a lot about what kind of articles the participants were using in their writing (Lu, 2001). The most important contribution is that they can indicate if any patterns of underuse or overuse exist or if the errors are purely random. Altogether, 37 possible error types were identified. Article errors in the text that could not be tagged according to the error system were labeled “unclassified”; these included definite and indefinite articles that were erroneously used outside the NP, meaning they were general errors, not errors within the article system. Furthermore, it was presumed that these were writing mistakes, as there was no pattern to the errors.

Cohen’s Kappa analysis was used to measure inter-rater reliability. In the inter-rater procedure, only two coders were used due to time and financial restrictions. Both coders were linguistics graduate students and experienced English teachers. The coders were trained to use

the corpus over three essays. If agreement was not reached, the two coders discussed the coding problems, and extra training was provided when necessary. In this study, 20% of the data was randomly selected from the argumentation essays and coded by the two raters. The Kappa statistic was calculated to be 0.332, which indicates a fair level of agreement between the two raters.

#### 4. Results

This section explains the rationale and formulas for reporting accuracy and presents the accuracy of the three articles. Following this is a description of the distribution patterns of each article, including the semantic and structural functions. After the essays had been tagged, the data was checked for inter-rater reliability, and the raw frequency counts for each error type were computed.

In order to report the frequency of the article errors, the data had to be normalized to allow data from different texts to be accurately compared. As the lengths of the essays differed between participants, reporting the raw frequency counts would not present an accurate account of the errors. In a longer text, there are more opportunities for errors to occur, so 'normalization' is a formula that adjusts the raw frequency counts so texts of different lengths can be compared (Pica, 1983). In normalization, the raw frequency counts are divided by the number of words in the text then multiplied by the mean essay lengths for the 30 essays, which are 980 words per essay. The following example illustrates the normalization formula:

*definite for zero errors*  $26 / 1020 \times 980 = 24.98$  *definite for zero errors per 980 words.*

In this formula, there are 26 *definite for zero* errors in one essay. This is divided by the total number of words in the essay then multiplied by the mean essay length, giving a total of 24.98 errors per 980 words.

Table 3 reports the distribution of the article types and article errors throughout the corpus. It is presented as a matrix table and it is read from left to right. The article type *the* on the horizontal axis shows the definite article, and reading the column from left to right indicates where the definite article is substituted for another article. If the table is read from left to right, starting with the definite article, it indicates where the definite article is being substituted in place of another. For example, reading the matrix from left to right indicates that 9.16% *the for a* substitution errors occurred. The highest frequency is *definite for zero* errors at 28.45%. Countability errors occurred when the indefinite article was substituted for the zero article or vice-versa. The results illustrate that 6.77% *zero for a/an* errors occurred, and 2.45% *a for zero* errors occurred. The number of *zero for the* and *a for the* errors are low at 5.33% and 0.79%, respectively, indicating that the frequency of definite article underuse is low. Table 3 indicates that overuse of the indefinite and zero article is low, but more errors are

made with the definite article, while countability errors are relatively lower. In other words, there are far more semantic or pragmatic errors than grammatical errors. Grammatical errors are due to noun countability errors where the writer must assign the indefinite article to singular nouns and the zero article to plural or mass/non-count nouns.

**Table 3. Article error distribution**

Article	The		Indefinite a/an		Zero	
	freq	%	freq	%	freq	%
The	922.71	<b>93.87</b>	<b>48.41</b>	<b>9.16</b>	<b>120.36</b>	<b>28.45</b>
Indefinite a/an	<b>7.81</b>	<b>0.79</b>	443.95	<b>84.07</b>	<b>10.4</b>	<b>2.45</b>
Zero	<b>52.43</b>	<b>5.33</b>	<b>35.65</b>	<b>6.77</b>	292.21	<b>69.13</b>
Total	982.95	100	528.01	100	422.97	100

The next section analyzes the influence of semantic NP environments and countability in order to determine their impact on article errors, as the effects of specificity and countability have been well-documented as factors that influence L2 learners' article errors.

Table 4 illustrates the frequency of the main article errors according to NP environment and countability. The highest frequency of errors can be found in *definite for zero specific plural* errors followed by *the for specific indefinite a/an* errors. These descriptive results suggest that specificity influences the frequency of the *for indefinite a/an* errors, as more errors occur in specific NPs. The frequency of *zero for a* errors is low at 10.58% of total errors, but suggests that some participants have trouble using the correct article with singular and plural nouns. The number of *a for zero* and *zero for the* errors was not reported as their frequencies were very low, indicating that this was not a problem for the participants. The frequency of *definite for zero* errors in both specific and non-specific environments suggests that specificity may not be the only influence on *definite for zero* errors.

Further statistical analysis was needed to investigate the influences on error types. It has been predicted that, for English L2 learners with no article system in their L1, more errors are found in specific indefinite noun phrase environments. To determine the effect of specificity on definite for indefinite errors, a paired sample *t*-test was conducted. As there are only two independent variables, a *t*-test could show if the difference between specific and non-specific *the for indefinite a/an* errors is significant. It revealed a significant difference between the two groups ( $t(29) = 6.94, p < .001$ ). The mean of the specific indefinite errors was significantly higher ( $m = 1.36, sd = 1.03$ ) than the mean of the non-specific errors ( $m = 0.25, sd = 0.46$ ), indicating that specificity influences definite article errors in indefinite specific environments. In other words, the definite article is being substituted for the indefinite article in specific environments, as predicted by Ionin and Wexler's (2004) fluctuation hypotheses. The

implications of this are discussed in Section 5.

**Table 4. Error types across the corpus per 980 words**

Error type	Freq. of errors	% of total errors*
<b>Zero for A</b>	<b>35.65</b>	10.58
<b>Definite for zero</b>		
Non-count specific	23.98	9.62
Plural specific	<b>45.12</b>	<b>18.10</b>
Plural non-specific	<b>35.12</b>	<b>14.09</b>
Definite for zero non-count non-specific	16.14	6.47
<b>The for A</b>		
Specific	<b>41.09</b>	16.48

Note. N = 30 (N = shows the size of the data pool which is 30 subjects.)

\* Other error types are not included in this table.

Table 5 presents the descriptive statistics for definite article for zero article errors, where the four independent variables are *definite for zero specific plural* errors, *definite for zero non-specific plural* errors, *definite for zero non-count/mass specific* errors, and *definite for zero non-count/mass non-specific* errors. Some researchers (Goto-Butler 2002; Yoon, 1993) believe that, in addition to semantic environments, the difference between count and mass/non-count nouns may have an influence on article errors. Due to this, more errors are expected with mass/non-count nouns than with plural nouns. Also, due to the fluctuation hypothesis (Ionin & Wexler, 2004), which states that specificity influences article errors, more errors are expected in specific NPs. It was suggested that a repeated measure ANOVA would be able to show any significant differences between NPs environments and would also reveal any differences between plural and mass noun errors.

**Table 5. Descriptive statistics for definite article for zero article errors**

Substitution type	M	SD
Definite for zero specific plural errors	1.50	1.67
Definite for zero specific non-count/mass errors	0.79	0.99
Definite for zero non-specific plural errors	1.19	1.51
Definite for zero non-specific non-count/mass errors	0.53	0.73

Note: N = 30

Table 6 shows the repeated measure ANOVA results for the definite article for zero article errors. A significant effect was found ( $F(3, 87) = 5.66, p < .005$ ). Follow-up protected *t*-tests revealed a significant difference between *definite for zero plural* ( $m = 2.70, sd = 2.67$ ) and *definite for zero non-count/mass* substitution errors ( $m = 1.33, sd = 1.54$ ), showing an effect with noun countability on *definite for zero* errors. In other words, more *definite for zero* substitution errors are found with plural nouns indicating that, for these participants, mass/non-count nouns do not have a significant influence on definite article errors. The follow-up protected *t*-tests between *specific definite for zero* ( $m = 2.37, sd = 2.31$ ) and *non-specific definite for zero* errors ( $m = 1.73, sd = 2.09$ ) revealed no significant difference between specific and non-specific zero, indicating that specificity is not a significant influence in *definite for zero* article errors. The implications of this are discussed in Section 5.

**Table 6. ANOVA results for definite article for zero article errors**

	df	F	$\eta^2$	<i>p</i>
<i>Between subjects</i>				
Definite for zero subs	3	5.66	.003	.001**
Within-group error	87	(0.96)		

Note: N = 30; \*\* $p < .001$

## 5. Discussion

The results indicated that the participants in this study had problems using the English article in terms of distinguishing between a definite and indefinite noun phrase. Correct article use in terms of noun countability was not a major problem for these writers. This section discusses the influence of specificity on article error patterns. First, the indefinite article is discussed, followed by the zero article.

Errors with specificity may stem from some participants' identification of a specific noun clause as definite, as predicted by the fluctuation hypothesis (Ionin & Wexler, 2004). The results of this study support the view that the definite article is overused in specific noun phrases with indefinite *a/an*, as the results of the *t*-test show a significant difference between *the for indefinite a/an* errors, with more errors occurring in specific NPs. Nevertheless, the fluctuation hypothesis also predicts overuse of the indefinite article with definite non-specific nouns (*i.e.*, *a for the generic* errors). No such errors were evident in the results of this study, although only 54 generic indefinite noun types were counted in the data. This is a result of the low frequency of generic indefinite noun types in the writing samples.

Zero articles not taking a generic, proper noun, or idiomatic reading can be specific or non-specific, in accordance with Lu's (2001) specifications. The repeated measure ANOVA

and follow up protected *t*-tests revealed no effect of specificity on *definite for zero* errors, as no significant difference was found between specific and non-specific errors. Thus, unlike the indefinite article, specificity was not the only influence on the overuse of the definite article with zero articles. A misrepresentation of the pragmatic functions of the definite article is a possible reason for these errors, and this will be discussed below.

The results demonstrate that the learners in this corpus lacked accuracy with regard to the zero article, regardless of semantic type. As a result, the participants often compensated for this by using the definite article. The indefinite article cannot be used for plural nouns or mass or non-count nouns due to countability rules. Thus, a writer has two article options: the zero or the definite. Although the fluctuation hypothesis may explain the errors in specific environments, it cannot explain definite article overuse in non-specific environments; thus, the effects of other influences need to be considered-particularly mass/non-count nouns or the hearer knowledge [HK] feature of definite articles.

In English, the context-namely, the speaker's and hearer's knowledge of the context-determines whether an NP can be located by both participants. If the writer believes that the hearer is aware of the noun, the definite article is used. In other words, as Diez-Bedmar and Papp (2008) pointed out, a writer often takes the readers' knowledge into account when using the definite article.

According to Hawkins (1991), using the definite article enables the hearer to access the NP in a *p*-set (a set of knowledge known by the hearer/reader as being definite). The speaker/writer should use the definite article when he/she is confident that the other party knows that the NP is definite. A communication breakdown will occur if the speaker/writer uses the definite article erroneously or mistakenly believes that the hearer has such knowledge. The writers in this corpus have not been falsely assuming that the reader had definite knowledge-this would signal a lack of pragmatic awareness-but the writers may not have acquired how the definite article signals this knowledge. Thus, errors with the definite article could be classified as errors regarding the acquisition of the pragmatic functions of the English definite article.

The results reveal that participants made significantly more errors with plural nouns than with mass/non-count nouns. Errors involving the definite article with mass/non-count nouns have been found in other studies with Japanese L1s (Goto-Butler, 2002; Snape, 2008), although the results in this study reveal plural errors have a greater effect on error patterns. A *t*-test indicated a significant difference between mass/non-count nouns and plural nouns, indicating that definite article errors with mass/non-count nouns are less frequent than errors with plural nouns. This differs from what Goto-Butler (2002) found with their Japanese participants, who made more errors with mass/non-count nouns. In other words, for the participants in this study, the influence of mass/non-count nouns is not a significant factor in

English definite article errors.

Although noun countability has been seen as a problem for English L2 learners, especially learners whose L1 does not have an article system, for the participants in this study, the number of errors in *zero for a* and *a for zero* contexts was relatively low (9.51% and 2.87% of the total errors, respectively). A total of 18 *zero for a* and *a for zero* errors occurred with count nouns, indicating that the writers may be influenced by their L1.

Example 1. Every citizen is suitable by the law. No one is exception if he or she committed *crime*.

Example 2. For some losers may bankrupt and then rob *bank* in order to win back.

These examples indicate that zero articles were substituted for the indefinite article. One reason for this is that the writer applied his/her L1 rule instead of using an article with singular nouns because, in Mandarin Chinese, nouns do not always need a classifier, demonstrative, or numeral.

## 5.1 Pedagogical Suggestions

This section will offer suggestions to the language teacher based on the results of this study. It has already been pointed out the English articles are extremely difficult words to teach for two reasons. First, the definite article stacks multiple functions onto one word, making it cognitively more demanding for a learner to process. Second, as article errors do not cause communication breakdowns in daily conversation, they may be subject to fossilization in a learner's interlanguage (Brender, 2002). Although many researchers have looked at ways to teach all of the articles under one system (Bitchener, 2008; Master, 1990; 1994), the results of the current study demonstrate that the most frequent errors occur with the definite article in two main areas: *the for zero*, and *the for specific indefinite*. As most of the errors involved the definite article, the semantic environment of [+/-HK] and [+/- SR] are effective parameters for helping learners determine whether an NP needs the hearer's knowledge element or whether it is just a specific noun. In this way, both specificity and hearer knowledge can be brought into focus, as this study found it was the influence of both factors that resulted in more than 80% of the article errors.

Research on teaching article use (Bitchener, 2008; Brender, 2002; Master, 2002) has shown that explanations in the form of mini-lessons-along with group work and meaning-focused activities-are more suitable for this type of language feature. As much of the information about hearer knowledge is found in discourse or is non-linguistic, activities that incorporate the communication aspect of definiteness would also be beneficial for article errors.

Finally, this study helped with our understanding of the influence of specific knowledge,



hearer knowledge and noun countability on English article errors in writing. Given that the participants were all undergraduate English majors, it would be beneficial to design a cross-linguistic study involving higher level and lower level learners to observe the changes as learners' writing improves with ability and exposure to academic reading and writing. This would allow the researcher to design article teaching systems for all levels of learners based on the frequency of error types for each level.

## References

- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma Press.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17, 102-118.
- Brender, A. (2002). *The effectiveness of teaching articles to (-ART) students in EFL classes using consciousness raising methods*. Doctoral dissertation, Temple University.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition*, 12(1), 43-64.
- Chen, Y. (2002). The problems of university EFL writing in Taiwan. *Korea TESOL Journal*, 5(1), 59-81.
- Cheng, L.-S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. In G. Cinque & R. Kayne (Eds.), *The Oxford handbook of comparative syntax* (pp. 259-292). Oxford: Oxford University Press.
- Chiang, S. (2003). The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System*, 31, 471-484.
- Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics* 6, 339-405.
- Diez-Bedmar, M. B., & Papp, S. (2008). The use of the English article system by Chinese and Spanish learners. In G. Gilquin, M. B. Diez-Bedmar, & S. Papp (Eds.), *Linking up contrastive and learner corpus research* (pp.147-175). New York: Cambridge University Press.
- Goto-Butler, Y. (2002). Second language learners' theories on the use of English article: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system. *Studies in Second Language Acquisition*, 24(3), 451-480.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. Hong Kong: Longman Group.
- Han, N. R., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article use by non-native speakers. *Natural Language Engineering*, 12(2), 115-129.

- Hawkins, J. A. (1978). *Definiteness and indefiniteness*. London: Croom Helm.
- Hawkins, J. A. (1991). On (in) definite articles: Implicatures and (un)grammaticality predictions. *Journal of Linguistics*, 27(2), 405-442.
- Hua, D., & Lee, H. (2005). Chinese ESL learners' understanding of the English count-mass distinction. In D. Laurant, R. A. Sprouse, & A. Liljestr nd (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)* (pp. 138-149).
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor: Karoma.
- Ionin, T., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12, 3-70.
- Kao, T. & Chen, L. M. (2008). Coherence in Chinese students' English writing: An initiative to a learner's corpus. In Y. Leung, & H. Chang (Eds.), *Selected Papers from the Seventeenth International Symposium on English Teaching* (pp. 167-175). Taipei: Crane Publishing.
- Lee, E. H. (2007). *English article usage in online graduate forums by non-native EFL teachers*. Doctoral dissertation, Indiana University.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese. A functional reference grammar*. Los Angeles: University of California Press.
- Lu, F. C. (2001). The acquisition of English articles by Chinese learners. *Working Papers in Second Language Studies*, 20, 1-36.
- Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly*, 24, 461-478.
- Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System*, 25(2), 215-232.
- Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System*, 25(2), 215-232.
- Master, P. (2002). Information structure and English article pedagogy. *System*, 30(3), 331-348.
- Moore, J. M. (2004). *Articles and proper names in L2 English*. Doctoral dissertation, Northwestern University.
- National Cheng Kung University. (n.d.). *Academic Writing Textual Analysis (AWTA) corpus*. <http://awta.csie.ncku.edu.tw/>
- Neff, J., Ballesteros, F., Dafouz, E., Martinez, F., Rica, J. R., Diez, M., & Prieto, R. (2007). A contrastive functional analysis of errors in Spanish EFL university writers' argumentative texts. A corpus based study. In E. Fitzpatriz (Ed.), *Corpus linguistics beyond the word. Corpus research from phrase to discourse* (pp.203-227). Amsterdam: Rodopi.
- Pica, T. (1983). The article in American English: What the textbooks don't tell us. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 222-233). Rowley, MA: Newbury House.

- Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, 16(2), 135-172.
- Snape, N. (2006). L2 acquisition of definiteness and specificity in English by advanced Japanese and Spanish learners. In A. Belletti, A. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari (Eds.), *Language acquisition and development. Proceedings of the Generative Approaches to Language Acquisition Conference* (pp. 591-596). Cambridge, UK: Cambridge Scholars Press/CSP.
- Snape, N. (2008). Resetting the nominal mapping parameter in L2 English: Definite article use and the count-mass distinction. *Bilingualism: Language and Cognition*, 11, 63-79.
- Snape, N. (2009). Exploring Mandarin Chinese speakers' article use. In N. Snape, Y. K. I. Leung, & M. Sharwood Smith (Eds.), *Representational deficits in SLA: Studies in honor of Roger Hawkins*. (pp. 27-51). Amsterdam: John Benjamins.
- Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics*, 10, 335-355.
- Ting, F. (2003). An investigation of cohesive errors in the writing of PRC tertiary EFL students. *SIETS Language and Communication Review*, 2(2), 1-8.
- Trademan, J. E. (2002). *The acquisition of the English article system by native speakers of Spanish and Japanese: A cross-linguistic comparison*. Doctoral dissertation, University of New Mexico.
- Trask, R. L. (1995). *A dictionary of grammatical terms in linguistics*. New York: Routledge.
- Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition*, 11(1), 1-18.
- Wu, Y & Adams, B. (2009). Classifiers#determiners. *Linguistic Inquiry*, 40, 487-503.
- Yoon, K, K. (1993). Challenging prototype descriptions: Perception of noun countability and indefinite vs. zero article use. *International Review of Applied Linguistics*, 31(4), 269-289.
- Zhang, H. (2007). Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics*, 16, 43-59.

