

ACQUISITION OF THE ARTICLE *THE* BY NONNATIVE SPEAKERS OF ENGLISH

An Analysis of Four Nongeneric Uses

Dilin Liu

Oklahoma City University

Johanna L. Gleason

University of Oklahoma

Linguistic analyses suggest that the nongeneric use of the English definite article *the* falls into four major categories: cultural, situation, structural, and textual. This study aims to determine whether these uses present different levels of difficulty for ESL students and whether they are acquired at the same time. The instrument consisted of 91 sentences containing 60 deleted obligatory uses of *the* (15 per category) and, as control items, 40 zero articles (10 per category) where *the* is not allowed. The participants (41 low-, 49 intermediate-, and 38 advanced-level students) were instructed to read the sentences and insert *the* wherever they deemed necessary. Statistical analyses of the participants' performance indicate the following: (a) The four nongeneric uses pose different levels of difficulty, which suggests that ESL acquisition of *the* is use dependent and follows a natural order; (b) The participants' performance on the suppliance of *the* in obligatory contexts for all four uses improved significantly with proficiency level, whereas the overuse of *the* followed a different pattern: an initial worsening followed by an improvement as the subjects' proficiency level increased. Pedagogical implications, including instructional sequence and strategies for the various uses of *the*, are discussed.

Address correspondence to: Dilin Liu, TESOL Department, 271 Walker Center, Oklahoma City University, Oklahoma City, OK, 73106; e-mail: dliu@okcu.edu.

Because of its high complexity and frequent use, the English article system, which includes the indefinite article *a(n)*, the definite article *the*, and the zero (or “null”) article,¹ is one of the most difficult structural elements for ESL learners. In fact, it has often been considered hard grammar, very difficult if not impossible to teach (Dulay, Burt, & Krashen, 1982). A survey conducted by Covitt (1976) ranked the teaching of English article usage first among difficult tasks for ESL instructors. Quite a few ESL educators have explored different approaches and techniques for teaching article usage and examined the effectiveness of such instruction (Master, 1988a, 1988b, 1990, 1994, 1995; McEldowney, 1977; Pica, 1983b; Whitman, 1974); a few others (Huebner, 1983; Master, 1987a; Parish, 1987; Tarone & Parish, 1988; Thomas, 1989) have investigated the process of L2 acquisition of English articles, an issue that we believe deserves more attention. Because of the extreme complexity of the English article system, this study attempts to examine only one aspect of its acquisition—namely, the various nongeneric uses of the definite article *the*. We chose to focus on the use of the definite article because of its wide variety of usage and its higher frequency of use than the indefinite article *a(n)*.² Being limited to the nongeneric uses of the definite article and drawing largely from pedagogical research in ESL, this study is narrow both in its theoretical frame and the research methodology employed. The narrow scope of the study does not, however, diminish its importance because a better understanding of the acquisition process of *the* should in turn lead to more effective teaching and learning of this difficult article.

REVIEW OF LITERATURE

The English article system has long been a subject of interest for linguists, given its complex usage and the difficulty involved in analyzing it. Although some earlier studies (Christophersen, 1939; Jespersen, 1949) made significant contributions to our understanding of the issue, Bickerton’s (1981) work is arguably the most important and enlightening, as it renders a new and unique systematic approach to the analysis of the use of the English article system. According to Bickerton, the use of the English articles—*a*, *the*, and zero—is governed by the semantic function of the noun phrase (NP) in discourse. The classification of the semantic function of an NP is then determined, he argued, by two binary discourse features: (a) whether a noun is a specific referent (\pm SR), and (b) whether the hearer knows the referent (\pm HK). Based on such an analysis, NPs fall into four major semantic types.

Type 1 is [$-$ SR, $+$ HK], also known as “generics,” where the indefinite, the definite, or, if the noun is plural, the zero article is used. For example, *a/the tiger is a fierce animal*. Type 2 is [$+$ SR, $+$ HK], where the definite article is required. It includes four subcategories: (a) unique referent or conventionally assumed unique referent, such as *the Pope*; (b) referent physically present, as in the example *Pass me the pepper please*; (c) referent previously mentioned in the discourse; and (d) specific referent assumed to be known to the hearer

(e.g., a resident in a small village with one church tells another resident, “My wife is at the church”). Type 3 is [+SR, -HK], where the indefinite or, if the noun is plural, zero article is called for. There are two subcategories in this type: (a) first mention in the discourse of a [+SR] NP assumed to be unknown to the hearer, as in *Tom bought a car*; and (b) first mention of a [+SR] NP that follows existential *have* and is assumed to be unknown to the hearer, as in *My computer has a new sound card*. Type 4 is [-SR, -HK], where the indefinite or, if the noun is plural, zero article is required. This type consists of four subcategories: (a) equative NPs, as in *She is a single parent*; (b) NPs in a negation statement, as in *I don't have a car*; (c) NPs in an interrogative sentence, as in *Do you have a pen?*; and (d) NPs in hypothetical statements, as in *If I had had more money, I would have bought a new car*. This classification illustrates clearly that the four semantic types of NPs each have different discourse and referential constraints and thus call for the selection of a specific article or articles from among *a*, *the*, and zero to mark these constraints. It is this knowledge of the semantic types of NPs and the article(s) used with each type that enables English speakers and writers to mark the NPs in discourse with the appropriate articles.

If the grasp of the use of the English article system entails a command of the discourse and referential constraints on NPs, the acquisition of the article system must in turn involve the learning of these constraints—a task that research has shown, directly or indirectly, to be especially challenging, albeit possible, for L2 learners. Research on article acquisition in ESL falls into two areas: pedagogy and its effectiveness on the one hand, and the process of acquisition on the other. In the first category, some scholars (Covitt, 1976; Dulay, Burt, & Krashen, 1982; Grannis, 1972) have pointed out the complexity and the seemingly insurmountable difficulty of the English article system, whereas a few others (Master, 1994, 1995) have managed to show that systematic teaching of the article system can help students to grasp it more effectively. The results of Master's (1994) 9-week pedagogical intervention study showed that, whereas ESL students who received systematic instruction on the use of articles significantly improved their performance on article usage, those who did not receive such instruction did not. In another study involving 19 advanced ESL students enrolled in a Master's degree-level Applied Linguistics–TESOL course, Master (1995) had the students write a series of reading summaries, corrected their errors in article usage as feedback, and conducted brief classroom discussions on the most frequent errors. The course was divided into four successive periods of three to four weeks for the purpose of data analysis. The results showed a significant decrease in the number of article errors the students made between periods 1 and 4.

Concerning research on the ESL article acquisition process, some studies (Hakuta, 1976; Huebner, 1979, 1983; Tarone, 1985) were not specifically on article acquisition but on acquisition of grammatical morphemes in general. Only Master (1987a), Parish (1987), Tarone and Parish (1988), and Thomas (1989) studied the acquisition of articles exclusively. Research in both categories has

yielded some significant findings. The early studies of ESL use of articles looked mostly at the presence or absence of articles in obligatory contexts. In scrutinizing a Japanese child's acquisition of English morphemes, Hakuta (1976) found two types of errors in the subject's article usage, termed "error of omission" (no use of an article in an obligatory context) and "error of commission" (use of an article in a context where the zero or null article is used). In a longitudinal study of a Laotian ESL learner, Huebner (1979) employed a dynamic paradigm method in analyzing the changing functions of his subject's use of the definite article. The results showed that the learner's grasp of the article was gradual, first in one linguistic environment or function and then spreading to other environments and functions. In other words, the learner's use of articles, though systematic, varied from one linguistic environment to another and such variation led to change and development in the grasp of the system. Also investigating L2 variation, Tarone (1985) examined ESL students' grammatical accuracy on different tasks, such as grammaticality judgment, oral interview, and oral narration, whereby she found that the subjects' accuracy in the use of morphemes (including articles) varied significantly from task to task.

Huebner (1983) opened up a new avenue of research on ESL article acquisition by employing Bickerton's (1981) noun classification system discussed earlier. Unlike the traditional research that looked only at the presence or absence of articles in obligatory context, Huebner's analysis covers two more important issues: first, the various semantic functions or types of NPs and the article(s) used with each semantic type; and second, the development of ESL learners' grasp of the article + NP function relationship. Using the two binary features that Bickerton developed, Huebner classified the semantic functions of the NPs in his data. With this classification of nouns, one can examine the article(s) that an ESL learner uses with each type, thereby understanding the learner's use of articles in semantic context. Huebner's method of analysis thus allows researchers to gain a more in-depth understanding of ESL article usage than the method of examining only suppliance of articles in obligatory contexts. Tarone and Parish (1988) applied this new method in reanalyzing Tarone's (1985) data, whereby they discovered, among other things, an additional factor that may cause L2 language variation (i.e., variation in article usage): the different communicative functions that language plays.

In another study, Parish (1987) used Huebner's (1979, 1983) practice in combination with two other methods in studying a Japanese ESL learner's acquisition of the English articles over a 4-month period. One of the other two methods was an adaptation of Huebner's system. In addition to looking at which articles were used with each type of noun, the adapted method also examined the accuracy of the articles used. The third method Parish used was the traditional suppliance-in-obligatory-context analysis. With these three methods, especially the adapted method, Parish was able to ascertain whether the three articles (indefinite, definite, and zero) were acquired at the same time. Her data analysis indicated that the zero article was acquired first,

followed by the definite article, and finally the indefinite article. In one sense, Master's (1987a) study seemed to suggest the same sequence, that is, if one only looks at the percentage of articles correctly supplied in obligatory contexts. However, if one includes the overuse of articles in considering the accuracy of article use, one would have to revise the conclusion about the sequence, for the overuse of the zero article (therefore an underuse of *a* or *the*) by his subjects whose L1 did not contain an article system continued even at the advanced level. In fact, as Master (1997b) pointed out, the problems that such advanced learners have with the articles are limited mostly to the overuse of the zero article, particularly with a clearly identified referent (p. 220). This overuse of the zero article and the underuse of *the* at the advanced stage would suggest that the two articles are acquired rather late, a hypothesis indirectly supported by Master's (1995) study.

Although the focus of Master's (1995) investigation was the effectiveness of instruction on students' acquisition of articles, his study yielded some findings on the sequence of article acquisition. His analysis of the errors that the subjects made revealed that the most frequent error was omission of the definite article, and, more importantly, this particular error "tended to increase as a proportion of total errors over time" (p. 183). Given that the subjects were very advanced ESL learners (Master's degree-level TESOL students), it would suggest that *the* was perhaps acquired rather late in the ESL development of the article system. Another interesting finding from Master's (1995) study is that the frequently missed *the* was largely the result of the subjects' not knowing that certain NPs were unique in the speech community. In other words, most of their errors were related to what we consider unique referent or cultural use of the definite article, a concept we will explain shortly. This finding would imply that certain uses of the definite article might be more difficult than others, an issue that none of the existing studies seem to have investigated. Research so far has only examined the use of *the* as a whole, in comparison with the zero and indefinite articles, without looking at the learners' grasp of the various uses of the definite article. To do so, an informed discussion of the usage types of *the* is in order.

The Use of *The*

It is generally accepted that the use of *the* first falls into two major categories: generic and nongeneric use (Celce-Murcia & Larsen-Freeman, 1999; Hawkins, 1978; Quirk, Greenbaum, Leech, & Svartvik, 1985). The generic use of *the* refers to cases where it is used to mean either a species, a race, or people of a nation. It is usually used with a singular noun, as in *The German is very athletic*, although it may also be used, as some grammarians (Celce-Murcia & Larsen-Freeman; Christophersen, 1939) suggested, with plural nouns, as in *The Germans are very athletic*.³ All other uses of *the* are considered nongeneric, which makes its use much wider and more frequent than the generic use. In fact, except in the scientific register, the generic use of *the* has been found to

be very rare (Parish, 1987; Tarone & Parish, 1988; Whitman, 1974). Furthermore, the generic use of *the* in most instances can be replaced by the indefinite article *a(n)* if the noun is singular or substituted by the zero article if the noun is plural. The nongeneric use of *the*, on the other hand, cannot be replaced (in the case of a singular noun) or deleted (in the case of a plural noun). Furthermore, the nongeneric use of *the* is much more complex and hence more problematic for ESL students than the generic use.⁴ Quite a few scholars (Celce-Murcia & Larsen-Freeman; Christophersen; Grannis, 1972; Hawkins; Quirk et al.) have wrestled with the difficult issue of classifying the complex uses of *the*. Hawkins's work deserves our special attention for the purpose of this study.⁵

Drawing on the work of previous scholars such as Christophersen (1939) and Jespersen (1949), Hawkins (1978) developed a rather comprehensive theory known as the Location Theory to explain the various uses of nongeneric *the*. He identified a total of eight types of nongeneric use (pp. 106–149), listed here in 1–8:

1. *Anaphoric use*: use of *the* when something is mentioned a second time and subsequently (e.g., *Bill was working at a lathe the other day. All of a sudden the machine stopped running*)
2. *Visible situation use*: use of *the* with a noun mentioned the first time to refer to something that both the speaker and the listener can see (e.g., *Pass me the bucket*)
3. *Immediate situation use*: very similar to type 2, the only difference being that the thing referred to may not be visible (e.g., *Don't go in there, chum. The dog will bite you.* [Hawkins, p. 112])
4. *Larger situation use relying on specific knowledge*: use of *the* with a first-mention noun because it is known in the community (e.g., people from the same village talking about *the church, the pub*, and so forth)
5. *Larger situation use relying on general knowledge*: use of *the* with something that one can assume people from a country or around the world should know (e.g., *The White House* referring to the U.S. government, *the moon*)
6. *Associative anaphoric use*: basically the same as type 1, the only difference being that the first-mention *the* is used with a noun that is related to a previously mentioned noun, rather than being the same noun (e.g., *We went to a wedding. The bride was very tall.*)
7. *Unfamiliar use in NPs with explanatory modifiers*: use of *the* with a first-mention noun that has an explanatory or identifying modifier in the form of a clause, prepositional phrase, or noun (e.g., *The movies that are shown here now are all rated R; There was a funny story on the front page of the Guardian this morning; I hate the name Algernon.* [Hawkins, pp. 139 and 147])⁶
8. *Unfamiliar use in NPs with nonexplanatory modifiers*: similar to type 7, the only difference being that the modifier does not provide explanatory information (e.g., *My wife and I share the same secrets*, where the modifier *same* does not inform us as to what the secrets are but “only points to an identity between the two sets of secrets, my wife’s and my own” [Hawkins, p. 148]). Here *same* is used as a unique adjective that always requires *the*. There are a few adjectives that can be used this way, such as *only* and *sole*.

Through these examples, Hawkins (1978) illustrated how his Location Theory operates. Basically, when an individual uses *the*, he or she invites the listener or reader to locate the referent by using provided or assumed known cultural, situation, structural, or textual information. In both anaphoric and associative anaphoric use (types 1 and 6, respectively), the listener or reader relies on textual information. In visible and immediate situation uses (types 2 and 3, respectively), the person makes use of information readily available within his or her sensory reach. Similarly, in type 4 (larger situation use relying on specific knowledge), people in a local community rely on information locally available. In type 5 (larger situation use relying on general knowledge), one identifies the referent by resorting to information considered unique and shared by all the people who speak the language, information we call cultural knowledge. In types 7 and 8, the person locates the referent by using structural information because such information lies in a modifier such as a prepositional phrase, a relative clause, or certain adjectives. Based on this theory, we believe that we can combine some of his categories and classify the nongeneric use of *the* into four major types. The first is cultural use, where *the* is used with a noun that is a unique and well-known referent in a speech community. The second is situation use, where *the* is used when the referent of a first-mention noun can be sensed directly or indirectly by the interlocutors or the referent is known by the members in a local community, such as the only dog in a family or the only bookstore in a town. The third is structural use, where *the* is used with a first-mention noun that has a modifier. The fourth is textual use, where *the* is used with a noun that has been previously referred to or is related to a previously mentioned noun.

Although Hawkins (1978) included what we call cultural use in the situation category, we believe it should be a separate type because, although such use shares some characteristics with situation use, it differs from the latter in two important ways. First, whereas situation use (what Hawkins calls immediate situation use [type 3] and larger situation use relying on specific knowledge [type 4]) refers to instances where the referent is within the view of the interlocutors or is known to everyone in a community, cultural use (larger situation use relying on general knowledge [type 5] in Hawkins's terms) does not have that luxury. Instead, the interlocutors have to assume and resort to a shared knowledge in the entire language community to make such use functional. Second and more importantly, such use of *the* is often not framed by situation but is determined, to a large extent, by conventional practice. For example, we use *the* with some but not all disease names. Similarly, we place *the* before some geographical names (such as rivers) but not others (such as most lakes). We also use *the* with the musical instruments we play but not with the sports we play, for we can say *play the piano* but not *play the basketball*. Of course, the use of *the* in these circumstances is not completely arbitrary but often governed by rules.⁷ Yet unlike in the other uses where the rules are simple (e.g., one rule for textual use is that the referent must have been previously mentioned directly or indirectly; one rule for situation use is

that the interlocutors need to sense or know the referent), the rules in cultural use are often too complex and numerous, and, more importantly, they are shaped largely by convention, a point that can be made more obvious if we compare the rules with those of some other languages (cf. the use of the definite article before names of provinces or states in French but not in English). The term “cultural” thus best captures the nature of *the* so used.

Halliday and Hasan’s (1976) discussion of the meanings of *the* seems to support our classification system:

Its [*the*’s] meaning is that the noun it modifies has a specific referent, and that the information required for identifying this referent is available. . . . It does not say where the information is to be located. It will be found somewhere in the environment, provided that we interpret “environment” in the broadest sense to include the structure, the text, the situation, and the *culture*. (p. 74, emphasis added)

Therefore, in Halliday and Hasan’s view too, culture is a source distinct from situation. Huebner’s (1983) classification of the use of *the* with the four subcategories of type 2 NPs is also similar to ours except for four differences. Huebner’s first use in his system—unique referent or conventionally assumed unique referent—belongs to what we call cultural use because, as pointed out earlier, what is assumed to be unique is often culturally based. Huebner’s second type—referent physically present—falls into our situation use. The third—referent previously mentioned in the discourse—equals our textual use. His last one—specific referent assumed known to the hearer—is in fact part of situation use, where something is well known in a community (e.g., *The furniture store [first-mention] was robbed yesterday*). Huebner did not include in his analysis what we call structural use, a use that has a fairly high frequency of occurrence. Thus our classification of *the* is both more complete and concise and should therefore serve as a useful system for helping us examine ESL learners’ use of *the*. In this study, we aim to determine if the four types of use of *the* specified here are equally difficult for ESL students and are acquired simultaneously. Our hypothesis is that, because the four types of use vary considerably in context and rule complexity, they will not be equally difficult for ESL students and hence not acquired at the same time.

THE STUDY

Participants

The participants included 41 low-, 49 intermediate-, and 38 advanced-level ESL students. The 41 low-level students were from intensive English programs at several universities in the Oklahoma City metropolitan area. Most of them were college bound, and none of them had a TOEFL score above 500, as required by most colleges. Although most of them had been in America for less than a year, many of them had studied some English in their home country.

Their average length of English study was 4.41 years. The intermediate and the advanced students were undergraduate and graduate students attending a university in Oklahoma. They all had a TOEFL score of 500 or above and were given a cloze test to demonstrate their current English proficiency. The cloze test consisted of 60 blanks with each blank worth one point. Those who scored above 45 were placed in the advanced group, 71% of whom were graduate students. The rest were placed in the intermediate group. The average length of English study was 6.31 years for the intermediate students and 9.95 years for the advanced students.

Instrument

The instrument (see Appendix A) consisted of 91 sentences. In 51 of the sentences, there were a total of 60 deleted obligatory uses of *the* (15 per category), with some sentences containing one and others containing more. The remaining 40 sentences were included as distracters or control items (10 per category). Because there had been no known instrument of this nature, serious efforts were made to ensure the instrument's validity and reliability. To attain validity, we followed authority and used great care and precision in developing the items and testing the instrument in a pilot study. We will discuss these validity procedures in detail after we briefly explain the instrument's reliability. As was the case with validity, the care and precision used in developing the instrument and the pilot study employed to test it were crucial in ensuring the instrument's reliability. To this end, we conducted a Kuder-Richardson 20 reliability test on the instrument—that is, on the subjects' accuracy performance on the instrument's 100 items. The test yielded a K-R 20 reliability of .843, a result indicating that the instrument indeed had good reliability.

In developing the instrument, we first took great pains to make sure that the items were clear and appropriate. In writing the deleted obligatory *the* items, we first created sentences by consulting and following example sentences in Celce-Murcia and Larsen-Freeman (1999) and Hawkins (1978), and then simply deleted the obligatory uses of *the* we wanted to test from these sentences. This task, though demanding, was not too complicated compared with that of writing distracter items, especially those for situation and textual use. This is because writing distracters for cultural and structural uses was rather straightforward but finding distracters for situation and textual uses was not. As explained earlier, cultural use of *the* is often conventional. Not all names of places and diseases require the definite article. Therefore, for cultural distracters, we simply included names of diseases or places that do not take *the*, such as *polio* and *Yellowstone Park*. Similarly, not all NPs that have an explanatory modifier call for *the* (see note 6), so for structural-use distracters we used NPs that contained a modifier but did not require the use of the definite article (e.g., *Children growing up with both parents are healthier than children growing up with only one parent*).

Unlike in cultural and structural uses, the definite article is always required

in situation and textual uses—that is, it is always required with a specific referent in a known situation or with a referent mentioned again. To develop distracters for situation use, we had to settle for sentences that involve a scenario whereby something is referred to in an immediate situation but cultural practice makes *the* unacceptable. For example, in sentence 89, *The game show host says to the contestant, “What’s behind door number one?”*, *the* is not allowed before *door number one* because such NPs take the null article, whereas phrases like *first door* take *the*. The definite article therefore becomes unacceptable here not because the referent is not known in an immediate situation but because of cultural practice. For textual-use distracters, we made use of sentences in which a noun appears twice but, in its second appearance, is used as a general reference rather than a reference specifically to what has been mentioned earlier (e.g., *At the zoo I saw several tigers. I think that tigers are beautiful animals*).

As for the test format, we did not leave blanks for the missing obligatory uses of *the* nor for the unnecessary *the* distracters. We simply asked the subjects to read the sentences and insert *the* wherever they deemed it necessary. Our rationale for not including blanks was that if we did, some of the students, especially the low-level students, might fill every blank with *the*, making the data very unreliable. Our decision turned out to be a good one for another reason, for some students placed *the* in places we had not expected (i.e., not in the distracters). It thus provided us with some additional useful data. After the instrument was completed, it was first given to 30 native speakers of English as a pilot test. Although none of the native speakers produced *the* unnecessarily, a few did miss *the* in 21 places where we expected it. Of those 21 places, 9 were each missed by only one subject and 2 by two subjects. A scrutiny of these latter 11 items indicated that the most likely reason for their omission was the subjects’ carelessness, as can be seen in the following examples: *I look after a little girl and a little boy on Saturdays. The little girl is smart but [-] little boy isn’t*; *My mother has a white dog and a black dog. The white dog is taller than [-] black one*. Because they were deemed careless errors and missed by less than 6.7% of the subjects (nine of them by only 3.3%), we kept the items with just a few minor changes to some of them. The other 10 items were each missed by between 3 and 5 participants. A close analysis of these items suggested that, although carelessness on the part of the participants could also have caused these omissions, two more likely reasons were item ambiguity and some participants’ tendency to omit *the* in certain obligatory contexts, a problem already discussed in note 4. For example, four participants did not place *the* before *mumps* in *John nearly died of mumps when he was a little boy*. This could have resulted from what Rastall (1995) described as some speakers’ tendency to omit *the* before certain disease names that have historically taken *the*. In another example, five participants did not furnish *the* before *sun* and *moon* in *We may not have much but we have sun in the morning and moon at night*. The omission might have resulted from the fact that the two words *sun* and *moon* here could mean sunlight and moonlight

and, so used, they were common nouns that could be treated either as unique or nonunique (cf. *I've been in the sun for too long today vs. I've had a little too much sun today*). Because these 10 items were missed most likely because of their ambiguity or the participants' disagreement, as well as because they were missed by at least 10% of the subjects, we deleted them. We then had two colleagues read the revised instrument to further ascertain its accuracy and clarity before using it in the study. The subjects were given 1½ hours to complete the test, including the cloze test, and they were not allowed to use dictionaries.

DATA ANALYSIS AND DISCUSSION

Because reporting the participants' total accuracy scores (the number of items they marked correctly) is not highly meaningful for the purpose of our study, we have chosen to calculate and report the number of obligatory uses of *the* that the subjects missed and the number of unnecessary uses of *the* (those in the null article context) they used. For the former category, we first counted every obligatory *the* they failed to provide and then computed the subtotal for each of the four types of use. Each subject therefore received four scores in this category, corresponding to each use type. The scoring and tabulating for the latter category, on the other hand, turned out to be more complex than we had expected. As explained earlier, some of the participants furnished *the* not only in the distracters but also in many unexpected places. We included these unexpected uses of *the* in the total number because, whether expected or not, an unnecessary *the* was an error and, more importantly, examining it would help us understand better where ESL students are likely to overuse this article. Another problem we encountered was how to classify the different types of the overuse of *the*. Although it makes sense to classify as cultural or structural those overuses of *the* in the cultural or structural distracters, such as *the Lake Michigan* (sentence 68) and *the children growing up...* (sentence 27), it does not seem logical to categorize as situation or textual those overuses of *the* in the situation and textual distracters. For example, in the test item *At the zoo, I saw several tigers. I think that (the) tigers are beautiful animals*, the definite article is unnecessary because the noun *tigers* in the second sentence refers to tigers in general (generic use), not to the tigers previously mentioned. Such overuse is thus not textual but overuse with general reference nouns. As stated earlier, if a noun truly has a previous textual reference, *the* will be necessary, hence no textual overuse per se. The same can be said of *the* placed in the situation use distracters because, in these distracters, the noun is sensed or well known by both interlocutors, though cultural practice disallows the use of *the*. For example, *Chapter Twenty* in sentence 61 and *door number one* in sentence 89 are both in situation use, but conventional rules stipulate that such NPs take the null article (as opposed to NPs like *twentieth chapter* and *first door*, which take *the*). Given that

Table 1. Results of MANOVA on missed uses of *the* across three levels in four use types

Use type	<i>df</i>	Multivariate		
		Λ	<i>F</i>	
All four	8	0.44	15.37*	
Use type	<i>df</i>	Univariate		
		<i>SS</i>	<i>MS</i>	<i>F</i>
Cultural	2	166.39	83.20	10.14*
Situation	2	354.74	177.37	25.27*
Structural	2	406.00	203.00	47.58*
Textual	2	364.86	182.43	62.06*

**p* < .05.

the reason we do not use *the* with these nouns is cultural, we considered *the* placed before these nouns as cultural overuse.

Concerning the unanticipated overuse of *the*, some examples fall into the above mentioned three categories: cultural, structural, and general reference overuse. Others fall into the category of ungrammatical use, by which we mean those instances of *the* that were structurally unacceptable. Examples include *went the hiking* and *in the his book*. In short, there are four major types of overuse: cultural, general reference, structural, and ungrammatical. (For additional examples of our classification of the overuse of *the*, see Appendix B.)

Missed Obligatory Use of *The*

After we tabulated the results of the participants' performance, we calculated the mean of the missed obligatory uses of *the* in each of the four types of use for each proficiency level group. We then conducted a MANOVA using English proficiency as the independent variable on the three groups' means in each of the four types of use. The results reported in Table 1 show a significant difference on both the multivariate and univariate tests. The results supported our hypothesis that the four nongeneric uses of *the* are not equally difficult for ESL students.

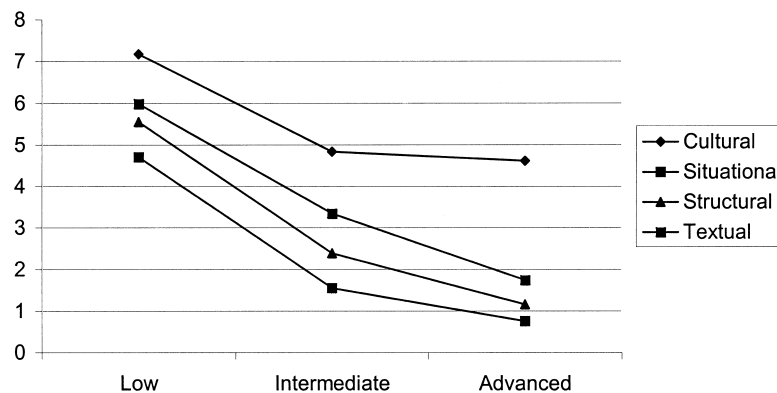
We then applied a post hoc Tukey test to see where the differences lie among the three groups in each of the categories. The results are reported in Table 2 in subscript letters together with the groups' means and standard deviations. The group means with the same subscript letter indicate no significant difference between them, and means with different subscript letters are significantly different. Figure 1 is also provided to help illustrate the differences between the groups.

As can be seen clearly in Table 2 and Figure 1, the number of missed obligatory uses of *the* in all four types of use decreases as the participants' English

Table 2. Means, standard deviations, and Tukey test results of comparison of means between groups

Group	<i>n</i>	Use type			
		Cultural	Situation	Structural	Textual
Low	41				
<i>M</i>		7.17 _a	4.71 _a	5.55 _a	5.98 _a
<i>SD</i>		2.82	4.00	3.32	2.32
Intermediate	49				
<i>M</i>		4.84 _b	1.55 _b	2.39 _b	3.35 _b
<i>SD</i>		3.05	1.95	2.04	1.51
Advanced	38				
<i>M</i>		4.61 _b	0.76 _b	1.16 _c	1.74 _c
<i>SD</i>		2.66	1.05	1.06	2.40

Note. Means with a common subscript are not significantly different by the Tukey test with $p < .05$.

**Figure 1.** Comparison of means of missed *the* by category.

proficiency level increases. More importantly, the Tukey test showed that the intermediate group's means in all four usage categories were significantly lower than those of the beginning group, which suggests an across-the-board significant decrease in missed articles from the low level to the intermediate level.

The comparison between the intermediate and advanced groups is somewhat more complex. Although the means of the advanced group were all lower than those of the intermediate group, the difference was significant in only two categories: structural and textual. The results suggest that, although ESL students' command of the structural and textual use of *the* continues to make significant improvement after their English proficiency passes the intermediate level, their grasp of the cultural usage and situation usage appears to

Table 3. Results of pairwise *t*-test

Pair	<i>df</i>	<i>t</i>
Cultural ^a vs. Situation ^b	127	11.90*
Cultural vs. Structural ^c	127	10.17*
Cultural vs. Textual ^d	127	6.30*
Situation vs. Structural	127	-3.33*
Situation vs. Textual	127	-5.59*
Structural vs. Textual	127	-3.35*

^aTotal *M* for Cultural = 5.52 (*SD* = 3.06). ^bTotal *M* for Situation = 2.33 (*SD* = 3.11). ^cTotal *M* for Structural = 3.02 (*SD* = 2.95). ^dTotal *M* for Textual = 3.71 (*SD* = 2.40).

**p* < .05.

have ceased improving significantly. A closer look at the latter two categories indicates, however, that the circumstances between cultural use and situation use were very different. First, whereas the mean of missed obligatory uses of *the* in cultural use decreased by only 0.23 from the intermediate level's 4.84 to the advanced level's 4.61, the mean in situation use fell by 0.79, about four times that of cultural use, from the intermediate group's 1.55 to the advanced group's 0.76. Second and more importantly, the advanced students still missed an average of 4.6 obligatory uses of *the* in cultural use but missed only an average of 0.76 obligatory uses in situation use. When the frequency of an error falls so low, the significance of the error also becomes minuscule. After all, language users, including native speakers, all make careless errors once in a while. Hence, although it is probably safe to say that the cultural use of *the* is still a difficult problem for advanced ESL students to wrestle with, it is perhaps not the case with its situation use.

Given that the MANOVA showed a significant difference among the four category means, we conducted a pairwise *t*-test of the three groups' total means of missed obligatory uses of *the* in each of the four categories to determine if there is a significant difference between each possible pair. The results reported in Table 3 show significant differences between all pairs. The significant differences in turn suggest a hierarchy of difficulty among the four types of usage, with cultural use being the most difficult followed in order by textual use, structural use, and situation use. The finding that cultural use is the most difficult supports the finding in Master's (1995) study, in which the largest number of errors with the article *the* were in cultural use. The discovery of situation use being the easiest suggests that, as far as this particular issue is concerned, the ESL acquisition of *the* follows the same process as in L1, for the function of *the*, according to Lyons's (1977) discussion regarding native speakers' grasp of it, "is *first learned* in actual situations-of-utterance with reference to entities present in the situation context" (pp. 656–657; emphasis added). The finding that textual use is more difficult than structural use is, on the other hand, somewhat surprising. Theoretically, textual use, like situation

use, is an easier concept to understand than cultural use and structural use. We have already discussed the extreme complexity of the cultural use of *the*. The structural use is also rather complex because not all NPs with an explanatory modifier require use of the definite article. Recall, for instance, in the sentence *Children growing up with both parents . . .* no definite article is allowed before the noun *children* because the NP is a general reference. With textual use, however, the rule is rather simple. After something has been mentioned once, it becomes an object or person known to the interlocutors involved, so when it is referred to again, the definite article is required.

The finding that the textual use of *the* is more difficult than its structural use is not only puzzling but also indirectly contradicts Parish's (1987) findings. In her study, the definite article appeared to be acquired earlier than the indefinite article. The use of the definite article in her data was primarily textual. Several possible reasons may help explain the puzzle and contradiction. First, whereas the data in Parish's study came from narration (subjects telling stories or describing events), our data consisted of participants' limited response to what they were reading and was therefore essentially a judgment of grammaticality. Different tasks have been shown to cause variation in L2 accuracy. As Tarone and Parish (1988) indicated, narration—a communicative language production task—requires a speaker to rely heavily on the accurate textual use of the definite article to communicate clearly and coherently. Grammatical judgment, on the other hand, is not a communicative task and the lack of communicative function may result in lower language accuracy. In that sense, our finding, compared to Parish (1987), would support Tarone and Parish's (1988) conclusions. Another reason for the textual use being more difficult than the structural use may be that, in determining if *the* is needed for textual use, the subject has to remember whether the noun has been mentioned previously; in structural use, the information that necessitates the use of *the* is in a modifier right before or after the noun. Finally, about half of the items in our instrument consisting of a textual use of *the* were of the associative anaphoric type—that is, the noun in question was associated with, rather than the same as, the previously mentioned noun. This lack of direct previous reference might have caused some of the subjects to miss the associated reference, a hypothesis indirectly supported by some results from Poesio and Vieira's (1998) study. The subjects in their two experiments were asked to classify and locate, among other uses of the definite article, direct anaphoric and associative anaphoric uses. In both experiments, the participants performed much worse in the identification of the associative category than in the direct one in terms of agreement in their decisions.

Overuse of *The*

With regard to the unnecessary uses of *the* that the students provided, we also first conducted a MANOVA. The results in Table 4 show, as in the case of missed obligatory uses of *the*, a significant difference between the three levels

Table 4. Results of MANOVA of overuse of *the* in three categories across three levels

Use type	<i>df</i>	Multivariate		
		Λ	<i>F</i>	
All types	8	0.62	8.39*	

Use type	<i>df</i>	Univariate		
		<i>SS</i>	<i>MS</i>	<i>F</i>
Cultural	2	565.24	282.62	14.41*
General reference	2	241.53	120.77	11.83*
Structural	2	93.15	46.57	11.60*
Ungrammatical	2	162.25	81.12	11.43*

* $p < .05$.**Table 5.** Means, standard deviations, and Tukey test results of comparison between group means

Group	<i>n</i>	Use type			
		Cultural	General reference	Structural	Ungrammatical
Low	41				
<i>M</i>		7.56 _a	4.02 _a	2.12 _a	2.66 _a
<i>SD</i>		4.44	2.82	1.47	4.57
Intermediate	49				
<i>M</i>		10.24 _b	6.67 _b	4.04 _b	0.51 _b
<i>SD</i>		5.00	3.65	2.28	1.02
Advanced	38				
<i>M</i>		5.13 _c	3.68 _a	2.53 _a	0.00 _b
<i>SD</i>		3.53	2.93	2.11	0.00

Note. Means with a common subscript are not significantly different by the Tukey test with $p < .05$.

of students on all four overuse categories: cultural, general reference (abbreviated “general” hereafter), structural, and ungrammatical.

We then applied a post hoc Tukey test to determine where among the three groups the differences lie. The results are reported in Table 5 in subscript letters together with the groups’ means and standard deviations. Figure 2 has also been provided to help illustrate the differences between the three groups. The results show that in the cultural, general, and structural categories, the intermediate students overused *the* more than both the low-level and the advanced students, respectively. In other words, the participants’ overuse in these three categories increased as their English proficiency improved from low level to intermediate level, and then such use began to decrease as their English proficiency improved further. In cultural use, the overuse increased

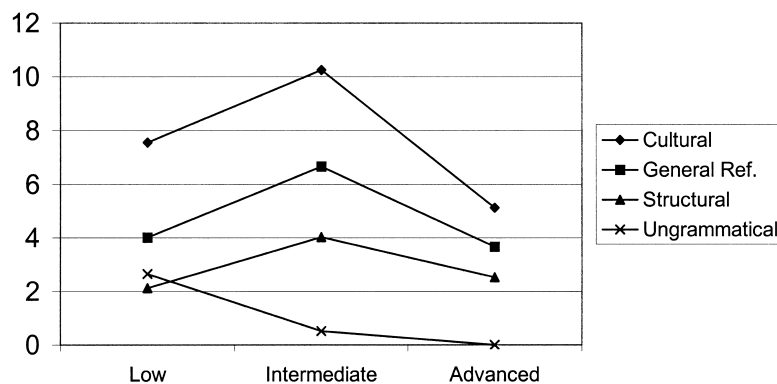


Figure 2. Comparison of means of unnecessary *the* by category.

significantly from low level to intermediate level and then decreased significantly as the participants' English proficiency reached an advanced level. More importantly, the advanced group's mean was also significantly lower than that of the low-level group. Overuse of *the* in the general and structural categories also increased and then decreased significantly, but the advanced group's mean, although significantly lower than that of the intermediate group, was not significantly different from that of the low-level group. The results seem to suggest that overuse of *the* still remains a problem for advanced ESL students.

The overuse of *the* in the ungrammatical category presents a very different picture. It decreased continuously, and understandably so, as the participants' English proficiency increased. In fact, very few intermediate subjects and no advanced subjects placed *the* in structurally unacceptable places. With their overall knowledge of grammar increasing, ESL students are expected to make fewer syntactic errors. On the other hand, the increase and then decrease in the overuse in the other three categories seems to support the existing findings regarding the acquisition process of *the* and inflectional morphemes in L2 in general. For example, Huebner (1983) and Master (1987a) both reported a significant increase of the overuse of *the* at the intermediate level. In terms of the acquisition of other morphemes, Lightbown (1983), in a longitudinal study of French-speaking children learning English at school in Quebec, found that the learners overapplied plural *-s* but that this overuse gradually decreased over time. Similarly, Pica's (1983a) study of the effects of different learning contexts on language acquisition showed that instructed learners tended to overuse grammatical morphemes more than naturalistic and mixed learners, although her study did not investigate whether such an overapplication would subsequently decrease.

We also wanted to know if the students' native language is a factor influencing their acquisition of the use of *the*. There were 18 native languages repre-

Table 6. Means, standard deviations, and the results of *t*-tests

Use type	<i>df</i>	Indo-European (<i>n</i> = 27)		Other (<i>n</i> = 101)		<i>t</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Cultural (U.U.)	126	4.11	2.59	5.89	3.08	2.75*
Situation	126	1.19	1.47	2.63	3.36	2.18*
Structural	126	2.11	2.87	3.27	2.93	1.83
Textual	126	3.41	2.24	3.79	2.45	0.74
Cultural (O.U.)	126	7.11	3.14	8.07	5.24	0.91
General reference	126	4.48	2.83	5.06	3.61	0.77
Structural	126	2.78	2.04	3.03	2.20	0.54
Ungrammatical	126	0.67	1.86	1.15	3.09	0.77

**p* < .05.

sented, though the majority of the subjects were speakers of Chinese, Korean, and Japanese. Some languages had only a few subjects (fewer than three in a few cases). As a result, counting every language as a variable would have produced statistically unreliable results. We decided instead to divide the students into two mixed language groups: Indo-European (*n* = 27) and all others (*n* = 101). Given that English is an Indo-European language, we wanted to see if speakers of other Indo-European languages would make fewer errors than students from other language groups. To answer our question, we conducted a two-tailed *t*-test on the two groups' means in each of the underuse and overuse categories. The results in Table 6 show that, although the Indo-European language speakers performed better—that is, they committed fewer omission and overuse errors in all categories—significant differences were found only in underuse of *the* in the cultural and situation use categories. Thus, native language does not seem to be a significant factor, at least not in all the usage types. Yet because the group size of Indo-European language speakers was about one fourth of the other group, the findings may not be reliable.

SUMMARY, PEDAGOGICAL IMPLICATIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

This study has yielded two key findings. First, the four nongeneric uses of the English article *the* present different levels of difficulty for ESL students and do not appear to be acquired at the same time. Instead, ESL acquisition of the nongeneric use of *the* seems to be use dependent and follows a natural order, given that ESL students appear to acquire situation use first, cultural use last, and structural and textual uses in between. Second, in the process of the acquisition of *the*, ESL students' underuse of obligatory *the* decreases significantly as their English proficiency improves, whereas their unnecessary use of *the* appears to follow a different course. It increases significantly as the ESL stu-

dents' English proficiency increases from low to intermediate level but then decreases as their English improves from intermediate to advanced level. The results support some previous findings on L2 morpheme acquisition processes.

There are also some pedagogical implications of this study. First, because ESL acquisition appears to follow a natural order, we must take this sequence into consideration in both classroom teaching practice and instructional material writing. Although it does not mean we should not teach cultural use or structural use of *the* to beginning ESL students, it certainly makes sense not to focus on these more difficult types of use at this stage. Instead, one should start with situation use. In doing so, we are (a) following the natural sequence of the acquisition of *the* undergone by both native speakers (as suggested by Lyons, 1977) and nonnative speakers, as shown in this study, and (b) conforming to a widely believed language-teaching principle—that is, to begin with things that students can see, touch, and hear. It is a principle that the Natural Approach (Krashen & Terrell, 1983) and the TPR approach (Asher, 1982) follow closely. In teaching the situation use of *the*, the teacher can make full use of the objects readily available in class. In teaching cultural use, especially those in idiomatic expressions, the lexical approach (Nattinger & DeCarrico, 1992) may be very helpful because these expressions are best treated as frozen lexical items. In short, we need classroom teaching practice and instructional material on English articles that reflect the natural acquisition order.

Second, based on our understanding of the four types of use of *the*, we believe that a variety of learning strategies should be employed to make instruction more effective. Situation use, for example, employs the five senses, hence, the use of kinesthetic, auditory, tactile, and visual learning. With structural and textual usage, more cognitive learning may be needed because understanding and practicing these two types of use involve the ability to analyze structural and textual information to identify the known information that would require the use of *the* with the noun in question. The cultural use of *the* would certainly require both cognitive learning and a significant amount of memorization because, as pointed out earlier, such use, though rule-governed in many ways, is often conventional and the rules are often too many and too complex to be easily grasped.

Third, given that ESL students seem to undergo a process of first underuse, then overuse, and finally appropriate or close to appropriate use in acquiring the various usage types of *the*, ESL teachers should be patient and should not feel frustrated when students still make errors after many hours of instruction. We also need to understand that some students move faster in the process than others, and we cannot expect them to grasp the different uses of the definite article at the same time.

Although we have drawn three pedagogical implications from this study, there are also three limitations. The first is sampling: Because the participants were mostly East Asians, there was insufficient representation of other language groups. The second limitation lies in the categorization of the students' levels. Instead of using a continuous measure of English proficiency, two dif-

ferent measurements were used in the present study. Although TOEFL was the quasi-measure used to separate the low-level students from the two higher level groups (the intermediate and advanced students all had a TOEFL score of 500 or above but the low-level group did not), a cloze test was employed to classify the intermediate and advanced students. The third limitation is the students' language data. This study did not involve students' own spontaneous language production. Therefore, future research calls for studies that (a) have a larger sample size with a more balanced representation of various language groups, (b) use a continuous measure of proficiency in grouping subjects, and (c) involve a greater variety of data. A comparison between subjects whose native languages contain an article system and those whose do not will also be interesting and useful, for it may help us to better appreciate the impact ESL learners' native languages have on their acquisition of English articles.

(Received 20 March 2001)

NOTES

1. Although the term "zero article" traditionally refers to any instance in which a noun requires no article, recent research (Chesterman, 1991; Celce-Murcia & Larsen-Freeman, 1999; Master, 1997a, 1997b) divides the zero article into two types: zero and null. The zero article occurs with nonspecific or generic noncount and plural nouns, such as *water* and *cats*. The null article occurs with certain singular count and proper nouns, such as *Chicago* and *lunch*.

2. We specified the indefinite article here to exclude the zero article. If the zero article is considered a full article, as it has been in some current practice, it is then the most frequently used article, followed in order by *the* and *a(n)*; more importantly, it is "the most frequently occurring free morpheme in the English language" (Master, 1997b, p. 221). If, however, the zero article is excluded, the definite article is by far the most frequently used word in the English language, according to many corpus-based studies (Francis & Kucera, 1982; Heeman & Allen, 1995; Johnsson & Hofland, 1989).

3. Master (1987b), however, argued that such generic use of *the* can occur only with singular count nouns. He contended that when plural count nouns are used with *the* in such a fashion, they essentially indicate all of the race or species in question, as seen in the example *All the Germans are athletic*.

4. The problem is further complicated by native speakers' occasional inconsistency in the use of *the*. One cause of such inconsistency is what Rastall (1995) identified as some speakers' tendency to omit *the* in certain contexts considered obligatory historically, such as in some geographical names like *Sudan*, *Ukraine*, and *Wembley Stadium*, or in phrases like *in face of* (all examples from Rastall) where, according to Rastall, *the* is deemed redundant either because the names have begun to be treated as proper nouns or the phrase is considered a fixed expression. Another source of this inconsistency relates to native speakers' pragmatic presupposition regarding whether to indicate a unique entity or to express a general class of entities or conceptual category, as can be seen in *Are you still writing a biography of Nixon?* versus *Are you still writing the biography of Nixon?*

5. As stated at the beginning of the paper, the purpose of our study is narrow and, as a result, the framework of our discussion of the definite article here is also limited. We do not attempt to provide a complete grammatical description of the use of *the* nor an exhaustive treatment of the complexity of its usage.

6. It should be pointed out, though, that not all NPs that have an explanatory modifier need the definite article (e.g., *Children born with heart problems require more parental care*). An NP like the one just cited uses the zero article instead of the definite article because, as Master (1990) explained, such a noun classifies rather than identifies information—that is, it refers to a type of people or object rather than to specific ones, as is the case in *the children who were injured yesterday*.

7. For example, concerning disease names, ailments (e.g., *cold*, *headache*) take *a(n)*; common names for common illness (e.g., *flu*, *plague*) take *the*; formal names for diseases (*polio*, *cancer*) take the null article. Regarding lakes and mountains, singular nouns (*Lake Michigan*) take the null article but plural ones (*Rocky Mountains*) require *the*. Celestial bodies in our solar system all take the null

article except *the earth*, *the sun*, and *the moon* because, whereas the latter three can be used as common nouns and require *the* to become unique referents, the rest (Mars, Jupiter, Venus, etc.) are all proper nouns.

REFERENCES

- Asher, J. (1982). *Learning another language through action: The complete teacher's guide book* (2nd ed.). Los Gatos, CA: Sky Oaks Production.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL teacher's course* (2nd ed). Boston: Heinle & Heinle.
- Chesterman, A. (1991). *On definiteness: A study with special references to English and Finnish*. New York: Cambridge University Press.
- Christophersen, P. (1939). *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.
- Covitt, R. (1976). *Some problematic grammar areas for ESL teachers*. Unpublished Master's thesis, University of California, Los Angeles.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. Oxford: Oxford University Press.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Grannis, O. (1972). The definite article conspiracy in English. *Language Learning*, 22, 275–289.
- Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning*, 26, 321–351.
- Hawkins, J. (1978). *Definiteness and indefiniteness*. San Diego, CA: Academic Press.
- Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Heeman, P. A., & Allen, J. E. (1995). *The TRAINS-93 dialogues*. (TRAINS Tech. Note No. 94-2). Rochester, NY: University of Rochester, Department of Computer Science.
- Huebner, T. (1979). Order-of-acquisition vs. dynamic paradigm: A comparison of methods in interlanguage research. *TESOL Quarterly*, 13, 21–28.
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma.
- Jespersen, O. (1949). *A Modern English grammar on historical principles: Vol. 7*. Copenhagen: Munksgaard.
- Johnsson, S., & Hofland, K. (1989). *Frequency analysis of English vocabulary and grammar, based on the LOB corpus: Vol. 1. Tag frequencies and word frequencies*. Oxford: Clarendon Press.
- Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon.
- Lightbown, P. M. (1983). Exploring relationships between developmental and instructional sequences. In H. W. Selinger & M. H. Long (Eds.), *Classroom-oriented research on second language acquisition* (pp. 217–243). Rowley, MA: Newbury House.
- Lyons, J. (1977). *Semantics: Vols. 1–2*. New York: Cambridge University Press.
- Master, P. (1987a). *A cross-linguistic interlanguage analysis of the acquisition of the English article system*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Master, P. (1987b). Generic THE in scientific American. *English for Specific Purposes*, 6, 165–186.
- Master, P. (1988a). Teaching the English article system (Part 1). *English Teaching Forum*, 26, 2–7.
- Master, P. (1988b). Teaching the English article system (Part 2). *English Teaching Forum*, 26, 18–25.
- Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly*, 24, 461–498.
- Master, P. (1994). Effects of instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229–252). New York: Cambridge University Press.
- Master, P. (1995). Consciousness raising and article pedagogy. In D. Belcher & G. Braine (Eds.), *Academic writing in second language: Essays on research and pedagogy* (pp. 183–204). Norwood, NJ: Ablex.
- Master, P. (1997a, March). *Acquiring the two zero articles in English*. Paper presented at the annual meeting of the American Association of Applied Linguists, Orlando, FL.
- Master, P. (1997b). The English article system: Acquisition, function, and pedagogy. *System*, 25, 215–232.
- McEldowney, P. L. (1977). A teaching grammar of the English article system. *International Review of Applied Linguistics*, 15, 95–112.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

- Parish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, 37, 361–383.
- Pica, T. (1983a). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning*, 33, 465–497.
- Pica, T. (1983b). The article in American English: What the texts don't tell us. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 222–233). Rowley, MA: Newbury House.
- Poesio, M., & Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, 24, 183–216.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Rastall, P. (1995). Definite article or no definite article? *English Today*, 11, 37–39.
- Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, 35, 373–404.
- Tarone, E., & Parish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, 38, 21–44.
- Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics*, 10, 335–355.
- Whitman, R. L. (1974). Teaching the article in English. *TESOL Quarterly*, 8, 253–262.

APPENDIX A

TEST INSTRUMENT

I. Please tell us:

Your native language _____
 Number of years you have studied English _____

II. In some of the following sentences, the definite article “the” is missing. Please read the following sentences carefully and insert the article “the” wherever you believe necessary.

1. Fred bought a car on Monday. On Wednesday, he crashed car.
2. I look after a little girl and a little boy on Saturdays. Little girl is smart but boy isn't.
3. I read a book about New York. Author, however, was from Arizona.
4. Jane bought a ring and a necklace for her mother's birthday. Her mother loved ring but hated necklace.
5. Rocket ships are launched from Cape Canaveral in Florida.
6. We rented a boat last summer at a lake. Unfortunately, boat hit another boat and sank.
7. My mother has a white dog and a black dog. White dog is taller than black one.
8. The mother says to her children, “Come on, it's time to go to Grandma's house.”
9. I watched several old movies last weekend. I enjoy watching old movies.
10. I have read a few science fiction books this semester. Science fiction books are really interesting.
11. Congress meets on Capital Hill.
12. At the zoo I saw several tigers. I think that tigers are beautiful animals.

13. While driving in their car to work, the husband asks his wife, "Could you open window please?"
14. Our office got some new computers last week. Someday, I really think that computers will replace people everywhere.
15. Before the examination begins, the teacher says to the students, "Write your answers in blanks."
16. I saw a man in a car across the street. At first I wasn't sure, but then I realized that man driving car was a friend of mine.
17. Handle of that cup was broken.
18. When I grow up, I want to be a doctor. Medicine is a widely respected profession.
19. Austin is capital of Texas.
20. Do you know pilot who flies this airplane?
21. Man I met in New York later became my husband.
22. Blue car across the road is very suspicious.
23. Did you hear house we saw last week was burned down last night?
24. I know man who runs this university.
25. Can you turn on light on top of that table?
26. In his office, the boss says to her secretary, "Turn on computer."
27. Children growing up with both parents are healthier than those growing up with only one parent.
28. Mary is not tall but she plays basketball very well. Usually short women aren't so good at playing basketball.
29. I've heard of parents who don't give their children enough to eat.
30. People from around the world are meeting here today.
31. We went to a basketball game on Saturday. Players at game were all very tall.
32. Shade on this lamp is really ugly.
33. Things of beauty always bring great joy.
34. We went hiking in Lake District last autumn.
35. She is only American woman to have run for vice-president.
36. I generally don't read newspaper articles from low-class papers.
37. Sally Ride was first woman in space.
38. Professor who teaches the physics class explains things very well.
39. A woman, with her hands full, says to a man standing in front of the office, "Open door for me, would you?"
40. Water in this glass is dirty.
41. A man says to his wife at the breakfast table, "Can you pass me newspaper?"
42. While driving in their car to work, the father says to his son, "Please turn on radio."
43. Tom and his friend are playing basketball. Tom says loudly to his friends, "Pass me ball."
44. Shoes in department stores tend to be expensive.
45. We went to a wedding. Bride was beautiful and groom was handsome.
46. The manager asks her secretary, "Could you please check schedule for me?"
47. I like to read books about philosophy.
48. Pacific Ocean is the largest in world.
49. Sun is shining. It's a beautiful day.
50. Moon is full tonight.
51. We got a new television for our house. I enjoy watching some programs, but in general I think that we shouldn't watch television so much.

52. Do you think we can move car that's blocking my driveway?
53. At dinner, the mother reminds her children, "Keep your elbows off table."
54. Who is leader of your club?
55. President of the United States lives in White House.
56. My mother likes to have salads at dinner because salads are very healthy.
57. Ladies of the night is a euphemism (an indirect word) for prostitutes.
58. Bill caught Malaria (a disease) while traveling in Africa.
59. In a bright sunny room, the woman asks the man "Could you close curtains, it's too bright in here."
60. I like to watch movies that are black and white.
61. The teacher says to his pupils, "Read Chapter Twenty in your book."
62. There are very poor people who are living in this community.
63. The wife hears a noise, then tells her husband "Doorbell is ringing. Answer door."
64. Mississippi river runs through Louisiana.
65. There has been a great deal of effort to clean up Chesapeake Bay.
66. A woman says to her friend "Why don't you come over for dinner tonight?"
67. Jim made a salad to go with dinner. Lettuce and tomatoes are always delicious in salad.
68. Lake Michigan is a large lake in North America.
69. The teacher says to her students, "Don't forget that your papers are due next week."
70. I start back to work on Monday.
71. The man says to his friend "I'm off on vacation tomorrow."
72. The teacher says to her students, "The meeting will not be held until next week."
73. The man says to his date, "I'll see you at eight o'clock."
74. A lot of people died of plague (a disease) in the 17th century.
75. Mojave Desert is in California.
76. New York Times is a very well known paper.
77. The mother asks the father, "Is baby sleeping?"
78. A plane crashed in Florida Everglades.
79. I'm sick. I've come down with flu.
80. The boss says to his employees, "I'm not happy with your work. Things are really going to have to change around here."
81. At dinner, the guest says to the host, "Could you please pass salt?"
82. England is part of United Kingdom.
83. Jill had polio (a disabling disease) when she was a little girl.
84. The daughter says to the mother, "I'll come visit you in June."
85. Yellowstone Park is in Wyoming.
86. John's wife died of cancer in 1996.
87. The wife says to her husband, who is hanging a picture in the room, "Picture isn't straight."
88. Salt Lake City is in Utah.
89. The game show host says to the contestant, "What's behind door number one?"
90. Mount Etna in Sicily is still an active volcano.
91. In their living room at bedtime, the mother says to the children, "Turn off television."

APPENDIX B

ADDITIONAL EXAMPLES AND EXPLANATIONS OF OUR CLASSIFICATIONS OF THE UNEXPECTED USE OF *THE*

Ungrammatical Uses of *The*

In classifying unexpected uses of *the*, we first separated them from the rest of the ungrammatical ones. For example, in sentence 69 (S69), *the* was found in three unexpected places (though not inserted by the same student): *The teacher says to her the students, Don't forget that your the papers are due the next week.* We considered the first two ungrammatical because the two nouns *students* and *papers* each already had a specific core determiner. We deemed the last one a structurally acceptable *the* but a cultural overuse because, although *the* is not allowed here, the NP *next week* does sometimes take *the* (we will explain its conventional rule in our discussion of examples of cultural overuse below). Thus, in our analysis, the ungrammatical uses of *the* were strictly those that were structurally unacceptable, and they were a tiny minority group compared with the structurally acceptable overuse of *the*. Other examples include: *We rented a boat last summer at the a lake . . .* (S6); *I watched several the old movies . . .* (S9); *The handle of that the cup was broken* (S17); . . . *the house we saw last week was the burned down last night* (S23); and *People from around the world are the meeting here today* (S30).

Cultural, General Reference, and Structural Overuses of *The*

We have included many more examples of cultural overuse than of the other two types because it constituted the largest group of the three categories and also because of the extreme complexity of cultural overuse coupled with the rather straightforwardness of general reference and structural overuse.

Cultural overuse. Some obvious examples are the use of *the* before those disease and geographical names that take the null article instead, such as *the polio* (S83), *the cancer* (S86), *the Arizona* (S4), *the Austin* (S19), and *the Mount Etna* (S90). Some less obvious examples include *the* in *she plays the basketball* (S28) because it is a cultural practice to say *play the piano* but not *play the basketball*, and *the* in *Sally Rider was the first woman in the space* (S37) because, although space here is a unique referent alluding to the region beyond the earth's atmosphere, convention stipulates that it take a null article rather than *the* when so used. For similar reasons, *the* in *on the top of that table* (S25) was deemed a cultural overuse. As mentioned above, we also classified as cultural overuse *the* in *the next week* (S69) because the NP here belongs to familiar use meaning "next week from here and now" and it contrasts with the unfamiliar use of the phrase that means "next week from there and then" as shown in the example *He visited me the next week*. Even though the phrase refers to a specific time in either use, convention dictates that it take a null article in familiar use but *the* in unfamiliar use. One more example of cultural overuse is *the* in *I'm off on the vacation tomorrow* (S71) because *on vacation* is an idiomatic expression without *the* or *a*, meaning "being away." The phrase may take *a*, as in *go on a vacation* but it means "taking a trip" instead.

General reference overuse. One example is *the* placed before the word *computers* mentioned the second time in *Our office got some new computers last week. Some day, I think that the computers will replace people everywhere* (S14). This use of *the* is a general

reference overuse because its head noun is a general conceptual noun rather than a reference to the computers mentioned earlier. Another example is *the* in *The rocket ships are launched from Cape Canaveral in Florida* (S5). Its head noun is not a reference to specific rocket ships and hence no *the* allowed. The same can be said of *the* in . . . *The things are really going to have to change around here* (S80). Again, the word *things* is not a specific but a general referent, and hence the definite article is a general reference overuse (e.g., *the* in *I like to read books about the philosophy* [S47]).

Structural overuse. One example is *the* used before *people* in *The people from around the world are meeting here today* (S30). Although the noun *people* is indeed modified in structure by a prepositional phrase, the modifier does not make the noun a specific referent, so the definite article is a structural overuse—that is, an NP that has a modifier but does not take *the*. The same is true of *the* placed before *parents* in *I have heard of the parents who don't give their children enough to eat* (S29). Again, the noun *parents* also has a modifier (a relative clause) but the modifier does not work to turn the noun into a referent that identifies specific individuals, hence no *the* is allowed. For the same reason, *the* inserted before *ladies* in *The ladies of the night is a euphemism for prostitutes* (S57) should be a structural overuse because the noun *ladies* also has a postnominal modifier. This example, however, was one of a few cases in our data where one could also classify it as another type of overuse—a cultural overuse in this case, given that *ladies of the night* (plural of *lady of the night*) is a fixed idiomatic expression, whereas *a lady of the night* or *the ladies of the night* are not. We decided to include it in structural overuse because the item was written as a distracter of structural use (i.e., a noun with a modifier but one not allowed to take *the*).