Acquisition of the English Definite Article by Chinese and Malay ESL Learners

Bee Eng Wong

(bee@fbmk.upm.edu.my)
Universiti Putra Malaysia, Malaysia

Soh Theng Quek

Universiti Putra Malaysia, Malaysia

Abstract

Chinese Mandarin and the Malay language have no functional equivalents of the English article system and it has been observed anecdotally that many Chinese and Malay ESL learners have difficulty using English articles accurately, particularly the definite article *the*. Based on Hawkins' Location Theory (1978), Liu and Gleason (2002) suggest that the non-generic uses of *the* fall into four major categories: situation, cultural, structural, and textual. This study aims to determine whether the non-generic uses of *the* present different levels of difficulty for Chinese and Malay ESL learners, and whether or not these different uses are acquired at the same time. One hundred (50 Chinese and 50 Malay) upper secondary students of three levels of English proficiency (advanced, intermediate and low) participated in this study. The participants were instructed to complete a 91-item fill-in-the-article-*the* test by inserting *the* in the items wherever deemed necessary. The measures employed for data analysis were SOC (Supplied in Obligatory Contexts), TLU (Target-Like Use), and UOC (Used in Obligatory Contexts). The participants' performance revealed that: (i) the four non-generic uses of the English article *the* pose different levels of difficulty; (ii) the acquisition order of the four non-generic uses of *the* follows a natural order independent of the ESL learners' first languages; and (iii) the participants' accuracy rate on article usage also depends on their proficiency level.

1 Introduction

Many studies on the acquisition of a second language carried out in the 1970s and 1980s have shown consistent acquisition orders for grammatical morphemes for both children and adults (VanPatten & Williams, 2007, p. 29). These morpheme studies were inspired by work conducted in first language (L1) acquisition by Roger Brown (1973). In a longitudinal study, Brown found a consistent order of emergence of 14 grammatical morphemes in English among three children of different backgrounds. Studies on the acquisition of grammatical morphemes among second language (L2) learners were conducted by Dulay and Burt (1973, 1974). In the first study (1973), the subjects were one hundred and fifty-one Spanish-speaking children of three different levels of exposure to English. Using the Bilingual Syntax Measure, an elicitation technique, to collect data for the study, they found that "the acquisition sequences obtained from the groups of children were strikingly similar although each of the group on the whole was at a different level of English proficiency" (Dulay, Burt & Krashen, 1982, p. 204). In another study, Dulay and Burt (1974), using a similar approach, but with different subjects (L1 Chinese and Spanish children,) found that the subjects acquired eleven grammatical morphemes in a similar order. Based on these findings, they

concluded that second language acquisition (SLA) is staged and systematic. SLA is staged in that learners work their way through a number of predictable stages or developmental sequences, from off-target L2 forms to increasingly more native-like forms (Towell & Hawkins, 1994, p. 10; Mitchell & Myles, 2004, pp. 15–16; VanPatten & Williams, 2007, p. 10). It is also systematic in that learners speaking different L1s develop L2 linguistic knowledge in a way that is not directly attributable either to their L1, or to the L2 input and their development follow a common route, although the rate of attainment may be very different (Towell & Hawkins, 1994, p. 11).

In addition to L2 development being staged and systematic, Towell and Hawkins also highlighted other phenomena that are prevalent among L2 learners. One of them is the phenomenon of variability, where L2 learners "allow more than one structural variant for a given construction where the target language has only one form" (p. 13). In addition, learners also transfer properties of their L1 grammar into the L2 grammar (p. 7). This phenomenon is also known as cross-linguistic influence from the L1 (Mitchell & Myles, 2004, p. 19–20). L1 influence can occur at all levels – phonological, syntactic and discoursal (Towell & Hawkins, 1994, p. 7–9). Yet another phenomenon that is observed among L2 learners, although to a lesser extent, is that of 'overgeneralisation.' In L1 acquisition literature, this is used to describe instances where children form generalisations based on input they receive; that is, they apply linguistic rules they have hypothesised too broadly (Crain & Lillo-Martin, 1999, p. 29). Examples include the formation of past tense and plural forms such as *goed* and *foots* and the overuse of particular grammatical elements such as the definite article or indefinite article.

In this study, we investigated the acquisition of the English definite article *the* by L1 Chinese and Malay speakers of L2 English. In particular, we set out to determine the order of acquisition of the definite article in four different contexts: situation, structural, textual and cultural. Thus the study is interested in investigating the staged development of this article for learners of three different proficiency levels. Where appropriate, explanations are provided for the data in light of the various phenomena that are observed among L2 learners in general and which are highlighted above.

2   Previous studies

It has been documented in L2 studies that learners of English generally have difficulty acquiring articles especially among learners from [-ART(ICLE)] first languages. Mandarin Chinese and the Malay language are two such languages. They do not have a functional equivalent of the English article system. Correspondingly, observational evidence has revealed that L1 Chinese and Malay ESL learners have difficulties with the article system in English, which consists of indefinite article *a(n)*, the definite article *the*, and the *zero* article, *ø*.

Articles in English are used with noun phrases. According to Huebner (1983), English noun phrases are classified by two discourse features of referentiality, that is whether the noun phrase is specified with the feature of specific reference [+/- SR] and hearer's knowledge [+/- HK]. There are four main types of noun phrases based on his model, which itself is based on Bickerton's work (1981, cit. in Huebner, 1983). Type 1 is [−SR, +HR], also known as 'generics,' where the indefinite, the definite, or, if the noun is plural, the zero article is used. Type 2 [+SR, +HR] refers to referential definites, Type 3 [+SR, − HR] to referential indefinites, and Type 4 [−SR, −HR] to non-referentials. In addition to these four types, idiomatic expressions and conventional uses are classified as Type 5 (see for e.g. Goto Butler, 2002; Thomas, 1989). Each of the four semantic types of NPs has different discourse and referential constraints and thus calls for the selection of a specific article or articles from among *a*, *the* and *zero* to mark these constraints (Liu & Gleason, 2002). In fact, a number of studies on the acquisition of the article system in English have been carried out, and some of them have adopted and adapted the framework proposed by Huebner (1983).

In 1976, Hakuta investigated the acquisition order of grammatical morphemes of a 5-year-old L1 Japanese girl acquiring English as a second language (ESL) in a natural environment. Among the morphemes studied were the definite and indefinite articles. Hakuta found that both articles were acquired as a system, with performance on *the* being initially better than on *a*. The overuse of

a and *the* involved specific and non-specific distinctions and the violations of *a* for singular noun phrases. In a study by Huebner (1979, 1983) on an adult L1 Laotian subject, it was found that *the* emerged early and the learner overgeneralised it. The indefinite article was acquired late in L2.

Parrish, in a longitudinal study (1987), found that his adult Japanese learners acquired the zero article first, followed by *the* and then *a*. However, the learners demonstrated a gradual increase in the use of *the*. However, they were less accurate with *a*. The zero article was overgeneralised in the findings.

In a study by Thomas (1989) with learners who speak both [+ART] (English) and [-ART] languages (Japanese, Chinese, Korean and Finnish), it was found that these ESL learners did not show early and accurate control of *a* in contexts where there is no hearer knowledge ([-SR, -HK] contexts) and the in the reverse context ([+SR, +HK], i.e. speaker referent and hearer knowledge). Errors made by L2 learners were either overgeneralisations of the zero article, or failure to use any article. Overproduction of zero was much higher for the [-ART] group than for the [+ART] group. L2 learners overgeneralised *the* in contexts where there is positive speaker referent and lack of hearer knowledge but there was no sign of overgeneralisation of *the*.

According to Master (1987, cit. in Master, 1997), articles are acquired differently, depending on whether or not this property is present in the L2 learner's first language. His informants were L1 Japanese speakers ([-ART]) and L1 Spanish speakers ([+ART]). He found that the zero article was the first to be acquired. The definite article, *the*, also emerged early, flooding all environments for the [-ART] speakers, a phenomenon known as '*the*-flooding' (Huebner, 1983; Master, 1997). The indefinite article *a(n)* was acquired later than *the*.

Thus, it seems that *the* is acquired early and *a* later in L2 acquisition (Huebner, 1983; Master, 1997; Parrish, 1987; Thomas, 1989). In addition, *the* may be overgeneralised. They also found that *the* dominates [+SR, +HK] and [+SR, -HK] contexts. However, Thomas (1989), found the zero article overgeneralised among learners of different proficiency levels. For speakers from [-ART] backgrounds, the zero article dominates in all environments in the early stages of SLA (Parrish, 1987; Master, 1997).

In their studies, Parrish (1987), Master (1997) and Ekiert (2004) also found that the zero article dominates in all environments for articles in the early stages of SLA of English. Thus, Parrish suggests an order of acquisition where the zero article precedes the definite article which then precedes the indefinite article. Master argues that learners acquire the zero article initially although he cautions that this explanation is problematic since one cannot tell the difference between the zero article and the omission of the article. Liu and Gleason (2002, p. 5), in re-examining Master's data claimed that, in Master's study, learners underused the definite and indefinite articles because acquisition of these was rather late. In Young's study (1996) on L1 speakers of Czech and Slovak, it was claimed that definiteness was not encoded by the definite article and this persisted till the more advanced stages. On the other hand, the learners encoded indefiniteness with the indefinite *a* at all levels of proficiency with greater accuracy as the process of acquisition advanced.

Thus it seems that the acquisition of articles in English is variable although it is evident that there seems to be overgeneralisation of the article morpheme, both definite or indefinite, among learners.

In the next section, the use of the definite English article is highlighted in comparison to the the notions of definiteness and indefiniteness in Chinese and Malay to show the cross linguistic differences of the three languages. A comparative analysis of the selected features of the three languages is necessary as the L1 of the respondents in this study are L1 Chinese and L1 Malay speakers and the language being studied is English.

3   Linguistic assumptions

To better understand the underlying processes in acquiring the English article system, a discussion of the grammar properties of the NP in Mandarin Chinese, Malay and English, specifically with regard to the notions of definiteness and indefiniteness, is provided here.

3.1   Definiteness and indefiniteness in Mandarin Chinese

According to Robertson (2000), "Mandarin Chinese does not have an equivalent article system that is present in English." In Mandarin Chinese, definiteness is marked on the NP through the word order or through the use of demonstratives. In Mandarin, a topic-prominent language, the sentence is constructed around the two-part division of topic and comment, whereas English has the subject and predicate. The topic which usually consists of information that is known to the speaker and assumed by the speaker to be known to the hearer is occupied by definite NPs. Indefinite NPs cannot occupy the pre-verbal topic position since they are necessarily not given information. Since the unmarked position for the subject is pre-verbal, post-verbal subjects will usually be indefinite (Robertson, 2000). Therefore, the word *rén* in 1a, which is the subject and occupies the topic position, is interpreted as definite, whereas in 1b it is interpreted as indefinite since it occurs after the verb (Li & Thompson, 1981, p. 120, cit. in Robertson, 2000, p. 141).

1)   (a)   rén   lái le
           'The person(s) has/have come.'

     lái-le rén   le
     'Some person(s) has/have come.'

As for object NPs, the same rules apply. The unmarked position for the object NP is after the verb (see 2a). The object may also, however, be marked with the particle *ba*, in which case it comes before the verb (see 2b). If the speaker wishes to emphasize the definiteness of the object NP, it may be placed in the preverbal topic position; see 2c and 2d:

2)   (a)   wǒ zài mǎi   shū   le
           'I   am buying a book.

     (b)   wǒ bǎ shū   mǎi le
           'I bought the book.'

     (c)   shū   wǒ mǎi   le
           'The book, I bought it.' (topic, contrastive)

     (d)   wǒ shū   mǎi le
           'I bought the book.' (contrastive)

(From Li & Thompson, 1981, cit. in Robertson, 2000, p. 141)

In Mandarin Chinese, there is widespread use of determiners which function in part to mark definiteness and indefiniteness. In particular, the demonstratives *zhèi* 'this' and *nèi* 'that' are used to mark definiteness and *yi* 'one' is used to mark indefiniteness. If an NP in topic position is un-marked, it can be interpreted as definite or generic (see 3a), whereas an NP pre-modified with a demonstrative must be interpreted as definite (see 3b). An NP pre-modified with *yi* 'one' is indefi-nite and therefore cannot occur in topic position (see 3c).

3)   (a)   gǒu   wǒ yǐjing kàn-guo le
                  Dog   I   already see-ASP ASP
                  'Dogs/The dog I   have already seen.'

            (b)   nèi-zhī   gǒu wǒ yǐjing   kàn-guo le
                  Det CLS[1] dog I already see-ASP ASP
                  'That   dog   I have already seen.'

(c)   * yi-zhī   gǒu wǒ yǐjing kàn-guo le
         One-CLS dog I already see-ASP ASP
       * 'A dog   I have already seen.'

<p align="center">(Adapted from Li & Thompson, 1981, p. 86, cit. in Robertson, 2000, p. 142)</p>

The distal demonstrative *nèi* 'that' and numeral *yi* 'one' (4a & 4b), in cases in which they are unstressed, are beginning to take on some of the functions of the definite and indefinite articles *the* and *a* in English respectively (Li & Thompson, 1981, p. 132; Huang, 1999, cit. in Robertson, 2000):

4)   (a)   nǐ rènshi   bu rènshi nèi-ge   rén?
           you know   not know   DET-CLS person
           'Do you know   the/that person?'

     (b)   tā   mǎi-le   yi-ge   màozi
           3PS   buy-ASP one-CLS hat
           'She/He   bought   a/one   hat.'

<p align="center">(Adapted from Li & Thompson, 1981, cit. in Robertson, 2000, p. 144)</p>

3.2   *Definiteness and indefiniteness in Malay*

Marsden (1812, cit. in Chan, 1996) regarded *itu* as equivalent to 'that, those, the' and the Malay pronouns are either demonstrative or definitive. However, he (Marsden) did not elaborate on what he meant by those two terms. Similarly, in Winstedt's Malay Grammar (cit. in Chan, 1996), *itu* was seen as being equivalent to 'the, that, those' with the sense of 'the distant, remote in space and time, the aforesaid.' Although Windstedt declared that there are no 'articles' in Malay, he admitted that there were a few exceptions in *itu* and *yang*. These are actually demonstrative pronouns which appear to have the force of the article and when it is required to indicate particular objects, use is made of these demonstrative pronouns.

In a translated work of Gonda (1939) by Khalid M. Hussain and N. Siahaan (1979, cit. in Chan, 1996), the strong statement by Gonda about the non-existence of an article in Malay is found. According to him, the morpheme *itu* below is not an article.

5)   Perempuan itu   lemah badannya.
           Women   GM[2] weak body-3PS[3]
           'That woman is physically weak.'

The subject *perempuan* in 5 refers to women in general. This is the generic perspective and *itu* is seen as a marker of generic nouns in copular equations (cf. Hopper, 1972, p. 129, cit. in Chan, 1996).

Another interpretation of the above is also possible.

6)   Perempuan itu   lemah badannya.
           Woman   Det (that) weak body-3PS
           'That woman is physically weak.'

In the example above, *itu* is a determiner with a demonstrative meaning and the noun phrase is singular.

Lewis (1963, cit. in Chan, 1996) is of the view that, at times, the demonstrative adjectives *itu, ini* are equivalent to the definite article, that is in the sense that the thing to which the noun refers has been under discussion and is familiar. As pronouns, these markers precede the noun. According to Safiah Karim, Farid M. Onn, Hashim Hj. Musa and Abdul Hamid (1997, pp. 349–361) there

are two types of determiners in the Malay language, pre-determiners and post-determiners. A pre-determiner occurs in front of nouns, and it can consist of numerals followed by classifiers, as shown in 7a and 7b.

> 7) (a) seorang askar
> one-CLS soldier
> 'a/one soldier'
>
> (b) sepuluh buah bangunan
> ten   CLS building
> 'ten buildings'

Although not equivalent, the use of *seorang* here could be described as functioning almost like the indefinite article *a*.

Post determiners, on the other hand, are elements that occur after nouns, and consist of *itu* 'that' and *ini* 'this', as shown in 8a and 8b.

> 8) (a) orang itu
> person Det
> 'that person'
>
> (b) rumah ini
> house Det
> 'this house'

The following example shows the sequence of the pre-determiner with *itu* and *ini* at the very end of the post-modifier structure, as in 9 (from Nik Safiah Karim et al., 1997).

> 9)   baju   tebal orang itu
> clothing thick person Det
> 'The person's thick clothing'

The use of *itu* 'that' and *ini* 'this' in these sentences demonstrates a notion of definiteness in the Malay NPs, although they are not equivalent to the English article *the*.

Cumming (1992, cit. in Chan, 1996) says that in Malay, "there are no morphemes which are specialised for marking definiteness, specificity, or identifiability; but as in many languages, other resources may be co-opted for these purposes." Thus the numeral *satu/se* 'one' is used to indicate indefiniteness and *ini* 'this' and *itu* 'that,' the deictic pronouns/demonstratives, may be used to mark definiteness.

3.3 Definiteness and indefiniteness in English

English has two articles, *the* and *a*, which are used in [+definite] and [-definite] contexts, respectively. While *a* is used only in singular [-definite] contexts, *the* is used in singular and plural [+definite] contexts. According to Heim (1991, cit. in Ionin & Wexler, 2003), *the* is specified [+definite], and *a* is underspecified for definiteness. This means that *the* can be used only when the conditions on definiteness have been satisfied, while *a* has no such requirement. As a result, whenever the conditions on definiteness have been met, *the* is used. If these conditions have not been met, *the* cannot be used, and *a* (or *zero*, in the case of plurals and mass nouns) is used instead. An illustration of the use of *the* and *a* is given below (see 10).

> 10)   I saw <u>a cat</u>. I gave <u>the cat</u> some milk.

Upon the first mention of *a cat*, there is no presupposition that a unique cat exists, so the conditions on definiteness have not been met. As a result, the indefinite article *a* is used. In contrast, the mention of the same cat for the second time, the existence of a particular cat has been established. Hence, the conditions on definiteness have been met, so the definite article *the* is used.

The feature [+definite] is discourse-related. It is related to the knowledge or mind state of the speaker and/or the hearer in the discourse. If a Determiner Phrase (DP) is [+definite], then the speaker assumes that the hearer shares the speaker's presupposition of the existence of a unique individual in the set denoted by Noun Phrase (NP), in the contextually relevant domain (Hawkins, 1978; refer to 11a below). Otherwise, the DP is indefinite (Ionin, Ko & Wexler, 2004; refer to 11b below).

11) (a)   Definite: I read a book. <u>The book</u> was interesting.
11) (b)   Indefinite: I read <u>a book</u> yesterday.

Due to the extreme complexity of the English article system, the researchers chose to focus on the use of the definite article *the* because of its wide variety of usage and its higher frequency of use compared to the indefinite article *a(n)* (Master, 1993, cit. in Lu, 2001). In this respect, the study examines only the non-generic uses of the definite article *the*. The non-generic uses of *the* refers to the four major types classified by Liu & Gleason (2002). Drawing on the work of Hawkins (1978) and his rather comprehensive theory known as the Location Theory to explain the various uses of non-generic *the*, Liu & Gleason (2002) combined some of his categories and classified the non-generic uses of *the* into four main types: cultural, situation, structural and textual. Hawkins (1978, pp. 106–149) developed the Location Theory to explain the eight identified types of non-generic use of *the*. The first is anaphoric use where *the* is used when something is mentioned a second time and subsequently (e.g. *John bought a car. <u>The car</u> is a Ferrari.*). Next is visible situation where *the* is used with a noun mentioned the first time to refer to something that both the addresser and addressee can see (e.g. *Give me <u>the</u> letter*). The third type is immediate situation use where *the* is used to refer to something not visible but is known to the addresser and addressee (e.g. *Don't put your hand in the box. <u>The rodent</u> will bite it.*). The next non-generic use is larger situation use relying on specific knowledge. In this instance, *the* is used with something that one can assume people from a country or around the world should know (e.g. *<u>The moon</u> is full tonight.*). Associative anaphoric use, the sixth type, is the same as anaphoric use except that the first mention of *the* is used with a noun that is related to a previously mentioned noun rather than the same noun (e.g. *We attended a party. <u>The host</u> was very gracious.*). Next is unfamiliar use in noun phrases with explanatory modifiers (e.g. *There was a touching story on <u>the front page of the New Straits Times</u> this morning.*). The final type is unfamiliar use in noun phrases with non-explanatory modifiers, similar to the seventh (previous) type (e.g. *My sister and I share <u>the same hobbies.</u>*). Thus when an addresser uses *the*, s/he is inviting the addressee to locate the referent by using provided or assumed known cultural, situation, structural or textual information (Liu & Gleason, 2002). Liu and Gleason identify anaphoric and associative anaphoric uses as relying on textual information. In visible and immediate situation uses, the addressee relies on information that is readily available. In larger situation use relying on general knowledge, the addressee identifies the referent by using unique cultural knowledge that is shared by all the people who speak the language. In both types of unfamiliar use, the addressee locates the referent by using structural information as such information is located in a modifier such as a prepositional phrase, a relative clause, or particular adjectives.

The following are examples of each type of non-generic use of *the* based on Liu and Gleason's classification (2002). The first is cultural use, where *the* is used with a noun that is a unique and well-known referent in a speech community.

12) (a)   The Pacific Ocean is the largest Ocean in the world.
12) (b)   The cabinet meets in Putrajaya every fortnight.

                                        (from test instrument, adapted from Liu & Gleason, 2002)

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 to the members in a local community.

13) (a) While driving in their car to work, the husband asks his wife, "Could you open the window please"?
13) (b) I read a book about Kuala Lumpur. The author, however, was from Penang.
(from test instrument, adapted from Liu & Gleason, 2002)

The third is structural use, where *the* is used with a first-mention noun that has a modifier.

14) (a) The blue car across the road is very suspicious.
(from test instrument, adapted from Liu & Gleason, 2002)
14) (b) I bought the red car. It was in the show room for months.

The fourth is textual use, where *the* is used with a noun that has previously been referred to or is related to a previously mentioned noun.

15) (a) I watched several old movies last weekend. I enjoy watching the old movies.
15) (b) Fendi bought a car on Monday. On Wednesday, he crashed the car.
(from test instrument, adapted from Liu & Gleason, 2002)

3.4 Summary

Definiteness and indefiniteness in Chinese are marked by the use of word order or through the use of determiners, such as *zhèi* 'this', *nèi* 'that', and *yi* 'one'. Definiteness is represented by either *ini* 'this', *itu* 'that' or *nya* 'its' in Malay, and the numeral *satu/se* 'one' is used to signify indefiniteness. On the other hand, a definite article *the* is used in English to indicate definiteness, both in the generic and the non-generic sense.

4 The study

The English language is one of the main subjects taught to students formally beginning from the age of seven in all the national type schools in Malaysia. Despite early exposure and frequent use, English articles, including the indefinite *the*, remain a generally acknowledged marked property for the [-Article] L1 Chinese and L1 Malay speakers. Lu (2001) claims that "article acquisition appears to be late, due to the fact that article choice is complicated, context-specific, and sometimes beyond simple rules, and that articles are unstressed function words and hence perceptually non-salient and semantically light-weight." (p. 2) For the learners in this study, although the English articles are traditionally introduced in the first few lessons of the primary English syllabus, these seemingly simple morphemes, including the definite *the* are not mastered until a very late stage of second language (L2) development.

The purpose of this study is to investigate the acquisition of the definite article *the* by L1 Chinese and L1 Malay ESL learners, in terms of article accuracy in four types of non-generic uses: cultural, situation, structural, and textual (based on Liu & Gleason's classification, 2002). The following research questions are addressed in this study:

1) In what order do L1 Chinese and L1 Malay ESL learners who are speakers of [-Article] languages acquire the non-generic uses (cultural, situation, structural, and textual) of *the* based on SOC, TLU and UOC measures?
2) To what extent do the learners' English proficiency levels affect the order of acquisition of the non-generic uses (cultural, situation, structural, and textual) of *the*?

4.1 Participants

The participants of this study were 100 form five students at a secondary school in Kuala Ter-engganu, Malaysia: 50 L1 Chinese speakers and 50 L1 Malay speakers. The characteristics of the two groups of participants are given in Table 1. This school was chosen because it happened to be the only school in the town with Chinese learners who had learned Mandarin Chinese since primary one and Malay learners who were L1 Malay speakers.

Based on studies in SLA (e.g. Johnson & Newport, 1989), the critical period or cut-off point for learners to acquire a language fully is around seven years of age. The L1 Chinese speakers in this study are considered to be homogeneous since they began the acquisition process of Mandarin Chinese from the age of six or below, that is from the kindergarten level. The medium of instruction in primary school for these learners was entirely in Mandarin. In addition, all of them spoke a Chinese variety (Hokkien, Cantonese, Teochew or Hainanese) or even Mandarin Chinese as L1. Moreover, the syntax of Chinese languages is very similar due to a common written history of more than two thousand years. Added to these facts, the varieties spoken by the respondents are all southern Chinese varieties. All of the learners were first exposed to the Malay language from the age of seven onwards. This exposure was limited to a period of study of about thirty minutes per day. Their exposure to English was also about thirty minutes per day.

The L1 Malay speakers also began English lessons at the age of seven in primary school. Again their exposure was limited to thirty minutes per day.

	L1-Chinese participants	L1-Malay participants
Number	50 (28 female, 22 male)	50 (30 female, 20 male)
Age range	16 years 10 months to 17 years 10 months (mean = 17; median = 17)	16 years 10 months to 17 years 10 months (mean = 17; median = 17)
Age at first exposure to English	5 to 6 years (mean = 6; median = 6)	6 to 7 years (mean = 6 ; median = 6.5)
Proficiency	14 low, 18 intermediate, 18 advanced level	20 low, 14 intermediate, 16 advanced level

Table 1: Characteristics of the participants

The average length of tutored exposure to English study for the two groups of learners was about 11 years. Each group was thus homogenous with regard to language background, educational level and age. The scores obtained by the students in their trial school certificate examination (SPM) English test were used to measure their current English proficiency level. Those who scored above 75% were placed in the advanced group while those who scored 60% to 70% were put in the intermediate group. As for the low achievers, their scores were from 40% to 50%.

4.2 Instrument

The instrument consisted of 91 items, adapted from Liu and Gleason (2002). In 51 of the items, there were a total of 60 deleted obligatory uses of *the* (15 per category or context), with some items containing one and others containing more. The remaining 40 items were included as distractors or control items (10 per category). These distractors were items that had (unobligatory) contexts that do not require the definite article. The items determined the respondents' ability to identify these contexts. Liu and Gleason (2002) ensured the validity and reliability of the instrument by conducting a Kuder-Richardson 20 reliability test on the instrument and the test yielded a K-R 20 reliability of .843, a result indicating that the instrument indeed had good reliability.

As for the test format, no blanks were used for the missing obligatory uses of *the* nor were they used for the unnecessary *the* distractors. The participants were asked to read the sentences and insert *the* wherever they deemed it necessary. The rationale for not including blanks was that some of the students, especially the low level students, might fill in every blank with *the*, making the data unreliable.

4.3   Procedure

The test was administered to the learners in a classroom environment. Prior to the test, the participants were assured of anonymity and confidentiality. Then, they were given a brief explanation to facilitate the administration of the task, and were asked to respond to the test items as spontaneously as possible. The learners were asked to insert *the* wherever deemed necessary. They were also told to leave the correct sentences as they were without any insertions if they believed them to be so. The participants took approximately 30 to 40 minutes to complete the fill-in-the-article-*the* test, although there was no time limit.

4.4   Coding and data analysis

The measures employed for the data analyses were SOC (Supplied in Obligatory Contexts), TLU (Target-Like Use) and UOC (Used in Obligatory Contexts). Since the English articles have no misformations as other morphemes, like past regular or irregular verb marking, a simplified version of the SOC measure was used in this study:

$$SOC = \frac{\text{Number of correct suppliances in obligatory contexts}}{\text{Number of obligatory contexts}}$$

However, if the morpheme is over-supplied or overgeneralized, SOC will overestimate the learner's accuracy. So, suppliance in non-obligatory contexts, like morpheme overgeneralization, was taken into account in the accuracy measure. The TLU measure was designed to redress this potential inflation of SOC, and it was formulated as follows (Pica, 1983, cit. in Lu, 2001):

$$TLU = \frac{\text{Number of correct suppliances in obligatory contexts}}{(\text{Number of obligatory contexts}) + (\text{Number of suppliances in non-obligatory contexts})}$$

While the SOC and TLU are used to measure article accuracy, UOC is used to measure article use. Master (1987) devised UOC to be a complementary measure to observe the learner's overuse or underuse of the article:

$$UOC = \frac{\text{The total number of suppliances in both obligatory and non-obligatory contexts}}{\text{Number of obligatory contexts}}$$

Like TLU, suppliance in non-obligatory contexts is also taken into consideration in UOC, so the learner's overall use of a certain morpheme can be inspected. Statistically, SOC and TLU cannot exceed 100%, but UOC can. So UOC is able to indicate overuse and underuse of the morpheme.

First of all, frequencies of the use of *the* in each of the contexts (cultural, situation, structural and textual) were counted for each participant. The number of suppliances in each of the obligatory context (situation, cultural, structural and textual) is 15 while for the non-obligatory context is

10. Next, the SOC, TLU and UOC for *the* were calculated for each participant as well as for each proficiency group and language group. Then, the means of SOC and TLU for each group were compared to identify the acquisition order in terms of article accuracy. To display the developmental sequences for each non-generic use, the means of SOC, TLU and UOC were plotted on line charts. In addition, the means of UOC were used to determine overuse or underuse of the English article *the* by each group.

Two-way analysis of variance (ANOVA) procedures with repeated measures were conducted for SOC and TLU scores obtained by the L1 Chinese and L1 Malay ESL learners of different proficiency levels (low, intermediate and advanced), respectively, in order to confirm whether there were significant differences in group means. The scores for the four non-generic uses of the English article *the* (situation, cultural, textual and structural) were calculated. In addition, the Scheffé post hoc follow-up test was used to adjust the significance level for multiple mean comparisons between the groups.

5   Results and discussion

The scores of SOC, TLU and UOC obtained by the L1 Chinese and L1 Malay ESL learners were recorded and analyzed to determine the order of acquisition of the four non-generic uses of the English article *the* by each language group. This study focused only on the difficulty level and acquisition order of the non-generic uses of the definite article as determined by comparing the SOC and TLU scores. In addition, the UOC scores were tabulated to find out the contexts in which the English article *the* were most overused and underused.

5.1   *The non-generic uses of the English definite article by L1 Chinese and L1 Malay ESL learners*

The performance results for the fill-in-the-article-*the* test are shown separately for the L1 Chinese and L1 Malay ESL learners. The mean scores were recorded for each use and were presented by proficiency level and by type of non-generic uses of *the*. Tables 2 and 3 show the means of SOC for the four non-generic uses of the article *the* by the three proficiency levels of the two groups of participants. By plotting the means of accuracy from the SOC measure, the interaction between the non-generic uses of the groups and the acquisition curves for the non-generic uses of *the* for L1 Chinese and L1 Malay ESL learners are obtained (see graphs in Figures 1 and 2 respectively). The graphs also illustrate the differences between the proficiency levels.

Proficiency	N	Types of Non-generic Use			
		Situation	Cultural	Structural	Textual
Advanced	18	.9556	.5550	.9300	.8922
Intermediate	18	.9283	.4517	.9183	.8322
Low	14	.8036	.4114	.7650	.6943
Total	50				

Table 2: Means of SOC for the non-generic uses of *the* of L1 Chinese ESL learners

Proficiency	N	Types of Non-generic Use			
		Situation	Cultural	Structural	Textual
Advanced	16	.9738	.6825	.8831	.8588
Intermediate	14	.9450	.6221	.7893	.7429
Low	20	.8375	.4470	.6970	.7070
Total	50				

Table 3: Means of SOC for the non-generic uses of *the* of L1 Malay ESL learners

The lines in Figure 1 for the three groups of the L1 Chinese ESL learners are wider apart for the situation, structural and textual uses. This result indicates that greater mean differences result in a significant relationship (see Table 4). The two-way repeated-measures ANOVA for the non-generic use effect on the overall SOC means of the three groups combined was found to be significant for all the contexts. The results of the Scheffé post hoc test also indicate that the mean differences between all possible pairs of the groups were significant at a level of $p < .05$, except for the intermediate and advanced proficiency level groups. The lines for the advanced proficiency and intermediate proficiency groups nearly meet each other at situation and structural contexts (in Figure 1), which indicates that a narrow mean difference of .0273 (between advanced and intermediate proficiency groups) and .0117 (between intermediate and low proficiency groups) respectively may not be significant. The results of the t-test used to make pair-wise comparisons between the contexts (in Table 4) show that all pair-wise comparisons are significant except the situation and structural contexts. The lines for the three groups also display a fairly systematic and similar pattern (see Figure 1). The mean differences between the four non-generic uses, which ranged from .0256 to .5442 (see Table 2) across the groups, are too small to expect a contrasting difference.
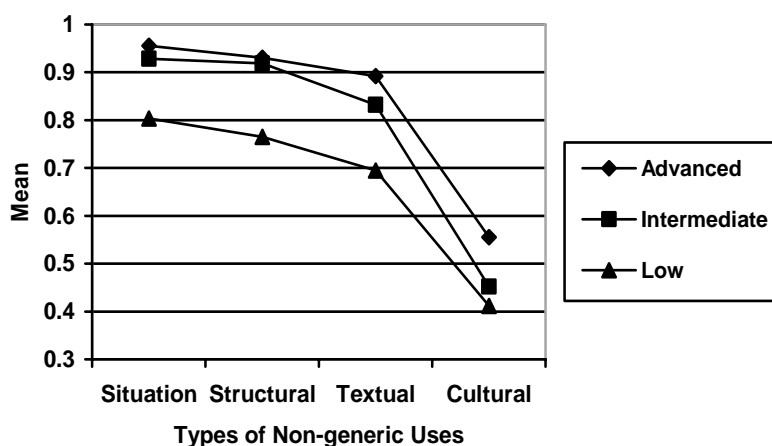

Fig. 1: Mean comparisons of SOC of L1 Chinese ESL learners

The mean differences between the situation and structural contexts, ranging from .01 to .0386 across the groups (see Table 2), are too small to expect a significant difference. Hence, it seems that SOC measures reveal an acquisition order of 'situation = structural > textual > cultural' (see Table 4), across the groups of L1 Chinese ESL learners. The order indicates that the obligatory use of *the* in the situation and structural contexts which are acquired almost at the same time, are acquired earlier than the textual and cultural contexts.
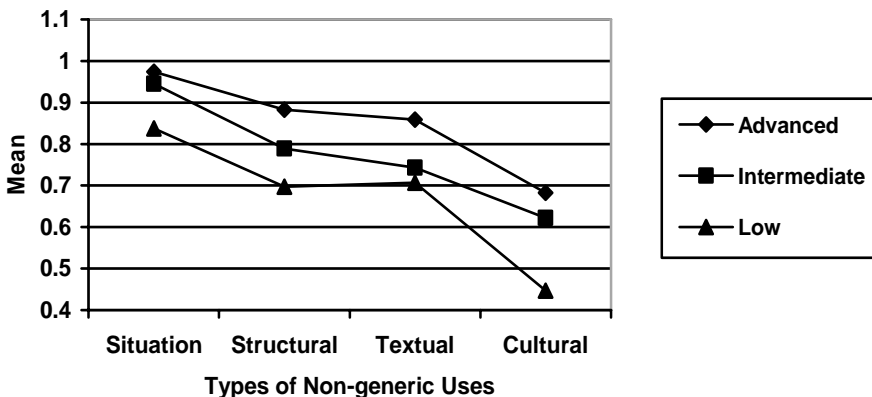
	Pair	t	df	Sig. (2-tailed)
Pair 1	Situation vs. Structural	1.589	49	.118
Pair 2	Situation vs. Cultural	20.507*	49	.000
Pair 3	Situation vs. Textual	5.622*	49	.000
Pair 4	Cultural vs. Structural	-16.503*	49	.000
Pair 5	Cultural vs. Textual	-14.819*	49	.000
Pair 6	Structural vs. Textual	4.163*	49	.000

*$p < .05$ (The mean difference is significant at the .05 level)

Table 4: Results of pair-wise t-test for comparisons of SOC of L1 Chinese ESL learners

When Table 2 is compared with Table 3, certain similarities in the acquisition of the non-generic uses of *the* by L1 Chinese and L1 Malay ESL learners become clear. The Malay learners also seem to follow the acquisition path as their Chinese counterparts (see Figure 2). The two-way repeated-measures ANOVA for the non-generic use effect on the overall SOC means of the three groups combined was also found to be significant for all the contexts. The results of Scheffé post hoc test indicate that the mean differences between all possible pairs of the groups were significant except for the intermediate and advanced proficiency level groups. In the structural and textual contexts, the low proficiency level group is not significantly different from the intermediate proficiency level group. In addition, all the possible pairs are not significantly different in the textual context. The textual context seems to be equally difficult across the proficiency groups.

Overall, the situation use shows the highest observed level of accuracy across both language groups. Even low proficiency level Chinese and Malay learners produced mean scores of .8036 and .8375 respectively (see Tables 2 and 3). The rate of accuracy reached 95.56% (M=.9556) and 97.38% (M=.9738) for advanced proficiency level Chinese and Malay learners respectively. The learners' command of situation use was indeed evident and independent of their first languages.

Fig. 2: Mean comparisons of SOC of L1 Malay ESL learners

Pair		t	df	Sig. (2-tailed)
Pair 1	Situation vs. Cultural	13.100*	49	.000
Pair 2	Situation vs. Structural	7.312*	49	.000
Pair 3	Situation vs. Textual	5.988*	49	.000
Pair 4	Cultural vs. Structural	-8.010*	49	.000
Pair 5	Cultural vs. Textual	-5.836*	49	.000
Pair 6	Structural vs. Textual	.719	49	.476

* $p < .05$

Table 5: Results of pair-wise *t*-test for comparisons of SOC of L1 Malay ESL learners

On the contrary, cultural use was the most difficult for all the learners, regardless of their proficiency levels. Low proficiency level learners were able to produce correct suppliances for less than half of the obligatory items: 41.14% (M=.4114) and 44.7% (M=.447) respectively (see Tables 2 and 3). Advanced proficiency level learners supplied correct answers for cultural use in 55.5% (M=.5550) and 68.25% (M=.6825) of the obligatory contexts. Again, language background did not appear to be a major variable for that particular usage. Nonetheless, the percentages of accuracy (mean scores) of the four types of non-generic uses increase with proficiency in a similar pattern across the two groups for these L1 Chinese and L1 Malay ESL learners (see Figures 1 and 2).

The big difference in accuracy rate between cultural and situation uses was paralleled by participants of both language groups. Structural and textual uses present a very different picture. In the case of structural use, advanced proficiency level Chinese and Malay learners scored a minimal difference of 4.69% (M=.9300 - .8831, see Tables 2 and 3) only, whereas their intermediate proficiency level counterparts registered the highest mean difference of 12.9% (M=.9183 - .7893). An interesting aspect of the data here relates to the low proficiency level Malay learners' mean scores. On structural and textual uses, both advanced and intermediate proficiency levels ESL learners scored higher in structural use compared to textual use, as did the low proficiency level Chinese learners. For low proficiency level Malay learners, the findings are reversed; they scored slightly higher in textual use than in structural use. However, the difference is rather negligible, that is of approximately 1% only (M=.7070 - .6970), which might be due to chance alone (see Table 3). In sum, the SOC measures reveal that the L1 Chinese ESL learners scored higher in the structural and textual uses, while their L1 Malay counterparts had higher mean scores in the situation and cultural uses (see Tables 2 and 3).

The results of the *t*-test for pair-wise comparisons between contexts for the L1 Malay ESL learners show that all pair-wise comparisons were significant except the structural and textual contexts. This might be due to the acquisition inconsistency demonstrated by the low proficiency level learners in these two contexts. Also, the mean differences between structural and textual contexts for the L1 Malay speakers, ranging from .01 to .0464 across the groups (see Table 3), are too small to expect a significant difference. Hence, it seems that SOC measures reveal an acquisition order of 'situation > structural = textual > cultural' across the groups for the L1 Malay ESL learners. The order indicates that the obligatory use of *the* in the situation context is acquired earlier than structural, textual, and cultural. In addition, the structural and textual contexts are acquired almost at the same time for the L1 Malay ESL learners.

However, when the results of the two groups of learners were combined, all pair-wise comparisons registered significant differences. Therefore, SOC measures reveal an acquisition order 'situation > structural > textual > cultural' across the groups for the L1 Chinese and L1 Malay ESL

learners. The order indicates that the obligatory use of *the* in the situation context is acquired earlier than structural, which is acquired earlier than textual, which is acquired earlier than cultural.

The procedure, as described above to calculate SOC, was used to identify the acquisition order for the TLU measure. Tables 6 and 7 show the means of TLU for the four types of non-generic uses of *the*. The same data are presented visually in Figures 4 and 5.

Proficiency	N	Types of Non-generic Use			
		Situation	Cultural	Structural	Textual
Advanced	18	.9044	.4961	.8311	.8194
Intermediate	18	.8083	.3789	.7194	.6822
Low	14	.6986	.3300	.5857	.5564
Total	50				

Table 6: Means of TLU for the non-generic uses of *the* of Chinese ESL learners

Proficiency	N	Types of Non-generic Use			
		Situation	Cultural	Structural	Textual
Advanced	16	.9044	.5956	.7831	.7281
Intermediate	14	.7364	.4243	.5600	.5243
Low	20	.7120	.3505	.5335	.5370
Total	50				

Table 7: Means of TLU for the non-generic uses of *the* of Malay ESL learners

The parallel lines in Figure 3 for the three groups of Chinese ESL learners show that the three groups might have only one acquisition pattern for TLU, as they do for SOC.
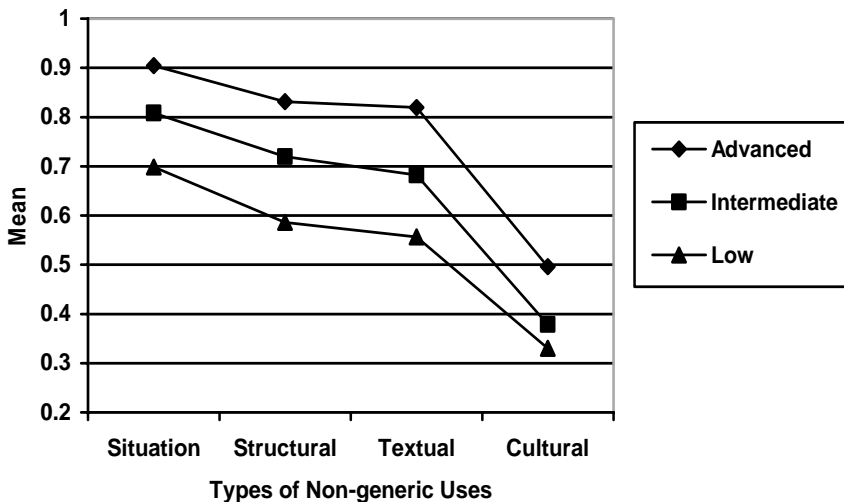
Fig. 3: Mean comparisons of TLU of L1 Chinese ESL learners

The two-way repeated-measures ANOVA for the non-generic use effect on the overall TLU means of the three groups combined demonstrates that all possible pairs were significantly different in all the contexts. The Scheffé post hoc test registers significant mean differences between all possible pairs of the groups except the low and intermediate proficiency level groups in the cultural context. The results of the *t*-test for pair-wise comparisons between contexts (see Table 8) show that all pair-wise comparisons were significant except the structural and textual contexts. This might be due to the narrow mean differences in the mean scores obtained by the L1 Chinese

ESL learners in these two contexts. The mean differences between the two contexts, ranging from .0117 to .0372 across the groups (see Table 6), are too small to expect a significant difference. Hence, it seems that TLU measures reveal an acquisition order of 'situation > structural = textual > cultural' across the groups for the L1 Chinese ESL learners.

Pair		t	df	Sig. (2-tailed)
Pair 1	Situation vs. Cultural	20.806*	49	.000
Pair 2	Situation vs. Structural	5.888*	49	.000
Pair 3	Situation vs. Textual	6.503*	49	.000
Pair 4	Cultural vs. Structural	-14.681*	49	.000
Pair 5	Cultural vs. Textual	-13.945*	49	.000
Pair 6	Structural vs. Textual	1.477	49	.146

* $p < .05$

Table 8: Results of pair-wise t-test for comparisons of TLU of L1 Chinese ESL learners

Their Malay counterparts however showed significant interaction in mean scores across the groups (see Table 9). In Figure 4, the lines for intermediate and low proficiency level nearly meet each other at all non-generic uses indicating insignificant interaction but are wide apart from the advanced proficiency level learners. This suggests significant mean differences between the advanced group on the one hand and the low and intermediate groups on the other.

The two-way repeated-measures ANOVA for the non-generic use effect on the overall TLU means of the three groups combined reveal that all possible pairs were significantly different in all the contexts. The results of the Scheffé post hoc test indicate that the mean differences between all possible pairs of the groups were significant except for the low and intermediate proficiency level groups in all the contexts. That is, the low and intermediate proficiency level Malay learners obtained scores that are approximately at the same levels of accuracy. Both proficiency groups obtained averages of 72.42% (M=.7242; situation use), 54.68% (M=.5468; structural use), and 53.07% (M=.5307; textual use) for the obligatory contexts (see Table 7). The results suggest that there is similarity in the acquisition pattern between the low and intermediate Malay ESL learners. The results of the t-test for pair-wise comparisons between contexts (see Table 9) show that all pair-wise comparisons were significant except for the structural and textual contexts. The acquisition inconsistency (higher mean scores in textual than structural) demonstrated by the low proficiency level learners in these two contexts might have been attributed to their being indeterminate in their use of the definite article in these contexts. Also, the mean differences between the structural and textual contexts, ranging from .0035 to .055 across the three groups (see Table 7), are too small to expect a significant difference. Once again, it seems that TLU measures reveal an acquisition order of 'situation > structural = textual > cultural' across the groups for the L1 Malay ESL learners. The order indicates that the obligatory use of *the* in the situation context is acquired earlier than structural, textual, and cultural. In addition, the structural and textual contexts are acquired almost at the same time or they pose the same level of difficulty for the L1 Malay ESL learners.

	Pair	t	df	Sig. (2-tailed)
Pair 1	Situation vs. Cultural	13.634*	49	.000
Pair 2	Situation vs. Structural	10.596*	49	.000
Pair 3	Situation vs. Textual	9.500*	49	.000
Pair 4	Cultural vs. Structural	-8.020*	49	.000
Pair 5	Cultural vs. Textual	-5.531*	49	.000
Pair 6	Structural vs. Textual	1.330	49	.190

* $p < .05$

Table 9: Results of pair-wise t-test for comparisons of TLU of L1 Malay ESL learners

Generally, the mean scores of TLU for situation use are still the highest of the four types of non-generic uses of *the*, followed closely by structural use and textual use, across the three proficiency groups for the two language groups (see Tables 7 and 8). Nonetheless, cultural use seems to be the most challenging to master for the ESL learners of both language groups. In sum, by looking at the graphs (Figures 3 and 4 respectively), once again a similar pattern of accuracy emerges for both L1 Chinese and L1 Malay ESL learners.
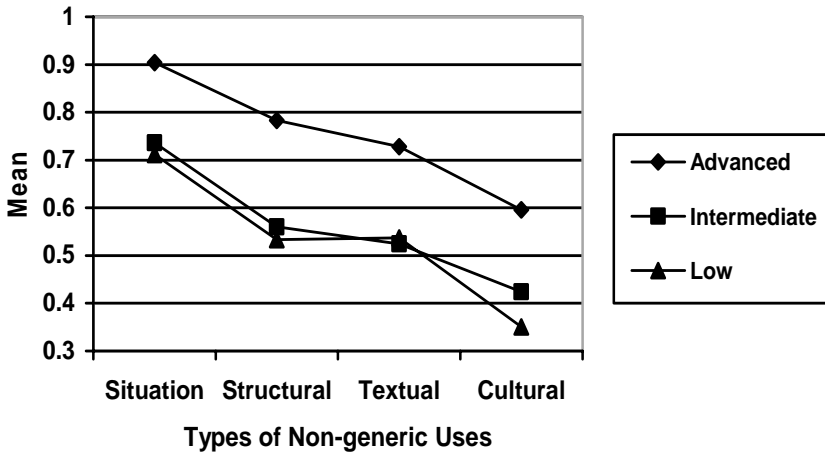
Fig. 4: Mean comparisons of TLU of L1 Malay ESL learners

So far, two aspects of the data have been considered: hierarchical orderings of non-generic uses, and levels of accuracy for each proficiency level of the Chinese and Malay ESL learners respectively. The similarities between the hierarchical orderings of acquisition across two samples justified the earlier findings. The TLU mean scores of accuracy for both language groups were plotted onto the line chart and the acquisition curves for the four types of non-generic use of the English article *the* can be seen in Figure 5. The general tendencies for each type hold true for the combined results. Therefore, it can be confirmed that the TLU measure reveals an acquisition order of 'situation > structural > textual > cultural' across the proficiency groups, for the L1 Chinese and L1 Malay ESL learners. This means that situation use is used more accurately followed closely by structural and textual uses, leaving cultural use to be the last in line, in obligatory contexts.
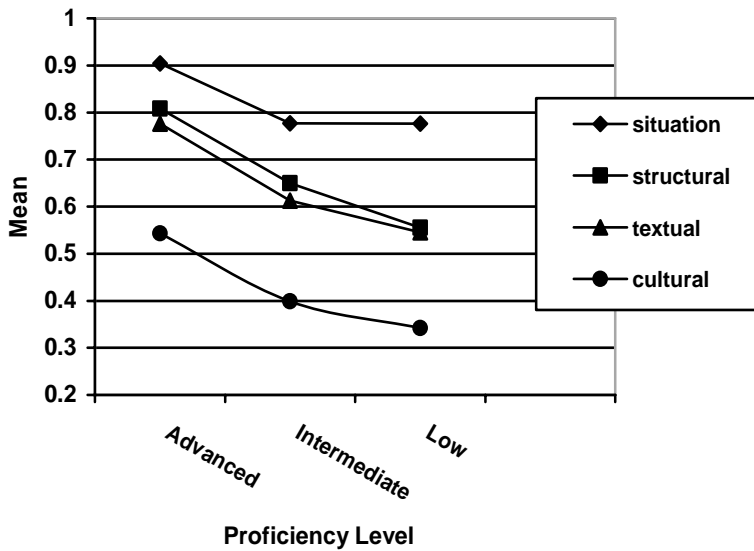
**Fig. 5: Mean comparisons of TLU of non-generic uses of *the* by proficiency levels
of L1 Chinese and L1 Malay ESL learners**

5.2   *Overuse and underuse of the English definite article*

Table 10 and Figure 6 show the means of UOC for the four types of non-generic uses of the definite article *the* for both groups of ESL learners. Reporting the participants' SOC and TLU scores would not give a thorough picture of learners' linguistic behaviour on the fill-in-the-article-*the* test. By tabulating the mean scores of UOC, a closer look at the overuse of the definite article *the* by the participants could provide an interesting insight into their developing interlanguage systems. The UOC measure shows that the definite article *the* was overused in structural and situation contexts for the advanced, intermediate and low proficiency level groups at the rate ranging from 100.71% (M=1.0017) to 121.62% (M=1.2162). The intermediate and advanced groups also overused the article in the textual context. However, UOC for cultural use indicates that it is underused at the rate of 74.65%, 82.87%, and 70.41% (M=.7465, .8287, and .7041) for the advanced, intermediate and low proficiency level groups respectively.

UOC	Proficiency Level	N	Mean	Std. Deviation	Std. Error
Situation	Low	34	1.0071	.24661	.04229
	Intermediate	32	1.1619	.22343	.03950
	Advanced	34	1.0374	.08877	.01522
	Total	100	1.0669	.20730	.02073
Cultural	Low	34	.7041	.32173	.05518
	Intermediate	32	.8287	.38224	.06757
	Advanced	34	.7465	.25365	.04350
	Total	100	.7584	.32305	.03230
Structural	Low	34	1.0250	.29511	.05061
	Intermediate	32	1.2162	.21321	.03769
	Advanced	34	1.0468	.18460	.03166
	Total	100	1.0936	.24871	.02487
Textual	Low	34	.9871	.29456	.05052
	Intermediate	32	1.1256	.20499	.03624
	Advanced	34	1.0235	.17234	.02956
	Total	100	1.0438	.23535	.02354

Table 10: Means of UOC for the non-generic uses of *the* of L1 Chinese and L1 Malay ESL learners

Obviously, the total overuse rates of the definite article *the* in the structural use by the ESL learners are the highest, which can be clearly noted in Figure 6, resulting in an inflated SOC rate. Because the SOC measure does not take overuse into consideration, the means of SOC for the four non-generic uses of *the* (see Tables 2 and 3) are much higher than the means of TLU for the four non-generic uses of *the* (see Tables 6 and 7). This shows that SOC measures tend to overestimate accuracy scores and need to be revised by a more accurate measure, TLU, as suggested by Stauble (1981) and Pica (1983) (both cit. in Lu, 2001). Therefore, TLU is probably a more reliable accuracy measure since it takes overuse of *the* into consideration. The results suggest an acquisition order of 'situation > structural > textual > cultural' for the four non-generic uses of the definite article *the* in this study which is identified by TLU, and is probably more reliable than the acquisition order identified by SOC.

In the case of overuse of the English article *the*, however, the picture is different from the acquisition order across the groups. There was a substantial increase in the unnecessary use of *the* by the intermediate proficiency level learners compared to those of the low proficiency level. The levels of overuse of this definite article then dropped for advanced proficiency level learners.

The graphs in Figure 6 also show that overuse of the definite article *the* increases from low proficiency level to intermediate proficiency level but decreases at the advanced proficiency level. The data indicate that learners became indeterminate at the intermediate proficiency level and then stabilises at near-native level, the advanced proficiency level.
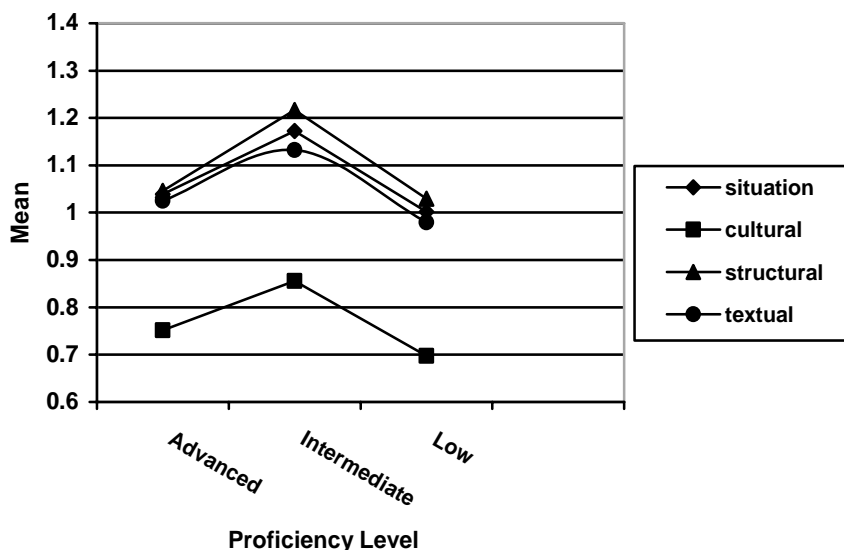
Fig. 6: Mean Comparisons of UOC of L1 Chinese and L1 Malay ESL Learners

6   Discussion

Consistent with Liu and Gleason's finding (2002), the ESL learners in this study had difficulty with the definite article *the* in the following (difficulty) order: cultural > textual > structural > situation. The finding that cultural use is the most difficult supports the findings in Liu and Gleason's (2002) study and Master's (1995, cit. in Liu & Gleason, 2002) study, in which the largest number of errors with the article *the* was in cultural use. The rules in cultural use are often too complex and numerous, and more importantly, they are often shaped by conventional practice. Hence, ESL learners from different L1s will probably face the most difficulty in this context due to dissimilarity in the grammar rules between their L1s and English. For example, the use of the definite article before the names of rivers and diseases in English is not applicable in Mandarin Chinese and Malay. This difference also explains the underuse of *the* in the cultural contexts.

On the contrary, the finding of situation use being the easiest suggests that the ESL learners' acquisition of the definite article *the* follows the same process as in L1 for the function of definiteness. According to Lyon (1977, cit. in Liu & Gleason, 2002), the definite article is learnerd first by native speakers in the acquisition process with reference to entities that are present in the situation context.

However, the finding that textual use is more difficult than structural use is somewhat unexpected. Theoretically, textual use which is similar to situation use is an easier concept to understand compared to cultural use and structural use. Structural use is rather complex because not all NPs with an explanatory modifier require use of the definite article. For example, in *people from around the world are meeting here today*, no definite article *the* is allowed before the noun *people* because the NP is a general reference (Item 30, test instrument). In contrast, the rule with textual use is rather simple. After something has been mentioned once, it becomes an object or person known to the interlocutors involved. Therefore, when it is referred to again, the definite article *the* is required. For example, *Fendi bought a car on Monday. On Wednesday, he crashed the car* (Item 1, test instrument).

Based on the results of the SOC and TLU measures used for article accuracy rates, the acquisition orders of the non-generic uses of the definite article *the* for the L1 Chinese and L1 Malay ESL

learners across the three groups (low, intermediate, and advanced proficiency levels) were identified as being the following:

1. (a) SOC reveals the acquisition order: situation > structural > textual > cultural.
    (b) TLU reveals the acquisition order: situation > structural > textual > cultural.

In terms of article accuracy, SOC and TLU help to identify the acquisition orders of the four non-generic uses of the definite article *the*: situation, cultural, structural and textual. This acquisition order is similar to the article acquisition order found in the study by Liu and Gleason (2002). On the other hand, the UOC measure helps interpret the acquisition processes underlying the orders in terms of the mean differences that occurred between the two acquisition orders. Thus the UOC measure provides an explanation for the mean differences between the two acquisition orders, namely, the relationship between the situation and structural contexts.

Due to overgeneralization of the definite article *the* in the structural context, the SOC measure for the use of the definite article was inflated, and thus the SOC measure revealed an acquisition order situation: structural > textual > cultural for the Chinese ESL learners. However, the TLU measure identified the acquisition order 'situation > structural = textual > cultural' as well as the frequency for overuse of *the*. This measure was deemed to be more reliable than the SOC measure (Stauble, 1981; Pica, 1983; cit. in Lu, 2001). The UOC measure aids in understanding the differences found in the four contexts (situation, cultural, structural, and textual) across the groups.

It is found that the Chinese and Malay ESL learners tended to overuse the definite article *the* in the situation, structural, and textual contexts but underused this definite article in the cultural context. This accounts for the low TLU mean scores obtained by the L1 Chinese and L1 Malay learners across the three proficiency groups in the cultural context. The data also indicate that only the low proficiency level ESL learners tended to underuse the English article *the* in the textual context. The low proficiency level learners overused *the* in the situation and structural contexts only (M=1.0015 and 1.0297). However, the intermediate proficiency level group demonstrated a substantial increase in the overuse of the definite article *the* in three non-generic uses, namely, situation, structural and textual. Similarly, Ekiert's (2004) study of the effects of different learning contexts (EFL and ESL) on article acquisition showed that intermediate-ability participants had the highest rate of the overuse of the definite article *the*. The relatively high rates of UOC for the intermediate proficiency level learners, was followed by a decrease in the data obtained for the advanced proficiency level learners in this study. Although the advanced proficiency group's mean was lower than that of the intermediate proficiency group's, the results of this study seem to suggest that overuse of the definite article *the* remains a problem for the advanced proficiency group. This finding is in line with Young's claim (1996) that the definite article was not encoded at the early stages of acquisition and the problem persisted even at the advanced stages among his subjects. This is also consistent with the occurrence of the "*the*-flooding" phenomenon chronicled in earlier studies (Huebner, 1983; Master, 1987, cit. in Master, 1997). Both Huebner and Master reported a significant increase in the overuse of *the* at the intermediate level.

In this study, the Chinese and Malay ESL learners of similar levels of proficiency exhibited parallel acquisition orders of the non-generic uses of the definite article *the*. One could conclude that the learners' L1s did not have an impact on the overall acquisition order of the non-generic uses of the definite article *the*. However, we need to examine the other data before such a claim can be made.

It is not surprising that the cultural use of *the* in particular is most problematic for the L1 Chinese and L1 Malay learners. As can be seen from the examples offered earlier, the use of some noun phrases are culturally bound. For example, *the sun* would be uttered as *matahari* in Malay and *tai yang* in Mandarin Chinese (or any variety for that matter) without any determiner or article to indicate definiteness. Some proper nouns such as *The Republic of China* also does not have the definite determiner or article if it is uttered in Malay and Chinese (any variety).

The overuse of grammatical elements resulting in the phenomenon of overgeneralisation, at lower levels of proficiency, particularly at intermediate levels has been noted earlier (see for e.g. Huebner, 1979, 1983; Master, 1987, cit. in Master, 1997) and in SLA literature in general. Master found that although the zero article is the first to be acquired, the definite article was generalised to all environments at the earlier stages of acquisition. Other studies also reveal similar findings. In Parrish's study (1987), ESL Japanese learners demonstrated a gradual increase in the use of the definite article but the learners overgeneralised or overused the zero article at the intermediate stages. Thomas (1989) also found that learners with both [+ART] and [-ART] as L1s (Japanese, Chinese, Korean and Finnish) overused the zero article. However, overproduction of zero was much higher for the [-ART] group than for the [+ART] group. These L2 learners also overgeneral-ised *the* in contexts where there is positive speaker referent and lack of hearer's knowledge and there was no sign of '*the*-flooding'. This phenomenon is highlighted in Archibald (1996, p. 506) when he explains that such errors are due to the developmental nature of SLA. He also mentions that such errors should initially be small but then increase at the intermediate levels before de-creasing. This is referred to as 'U-shaped' development (VanPatten & Williams, 2007, p. 10), where learners start out "doing something correctly, then subsequently does it incorrectly, and then 're-acquires' the correct form." The finding in this study which shows that the intermediate level learners overused the article more is in line with these other studies in the literature. This finding indicates that the overgeneralisation of *the* by the learners could be developmental 'errors'.

In relation to the developmental nature of SLA, the findings of past studies show that this is true for the main grammatical properties. However, this study and that of Liu and Gleason's (2002) indicate that the phenomenon is also evident for the subcategories of particular grammatical properties, in this case, the non-generic uses of the definite article *the*. Although the learners are from two different language backgrounds, they seem to acquire the definite article and its uses in the four contexts in a similar fashion. Therefore, it would appear that in terms of natural sequences in language acquisition of the definite article *the*, a learner's L1 (native language) does not seem to have an important influence on the acquisition of this property as systematicity is observed in the two sets of learners. But before such a claim is made, the other data need to be examined first.

The data also reveal that the acquisition of the definite article in all contexts improves with pro-ficiency level, similar to Liu & Gleason's finding (2002). This fact lends support to the claim that learners' output often follow predictable stages (VanPatten & Williams, 2007, p. 10) for a given structure, in this case, the non-generic uses of the definite article *the*. It is said that 'errors' due to cross linguistic influence (transfer) decrease as learners are exposed to increasing L2 input (see e.g. Archibald, 1996, p. 506).

Thus, it can be said that although the pattern of acquisition of the definite article *the* in the four contexts seemed to be developmental in nature (see e.g. Archibald, 1996, p. 506), one cannot dis-regard the role of L1 influence or transfer in the process since the data also reveal that the learner's proficiency level does affect the order of the acquisition of the non-generic uses of the definite article *the*. Generally, the learners do become better in their use of the definite article in the differ-ent contexts with increased proficiency. Such an increase is indicative of errors due to L1 influ-ence or transfer (Archibald, 1996, p. 506). However, VanPatten and Williams (2007, p. 11) claim that the influence of L1 is "somehow selective and also varies across individual learners."

7   Conclusion

As we have mentioned earlier, the study assumes and presupposes that Mandarin Chinese is similar to the L1 varieties of the Chinese speakers and since the acquisition of Mandarin Chinese occurred before the critical period for language acquisition had ceased to operate, and that it is the main language they have been exposed to in primary school for six years. And although the Chi-nese variety spoken at home may not be structurally congruent with Mandarin, it is assumed that they have an equal if not better level of competence in Mandarin Chinese than their other Chinese variety. It is this internal linguistic competence that may affect their learning of the English *the*.

The fact they had learnt Mandarin in a formal context would make it a stronger influence on their English.

Since both Chinese and Malay are [-ART] languages and structurally more similar in terms of definiteness and indefiniteness (neither has explicit articles and both rely on determiners to indicate definiteness and indefiniteness) for this property on the one hand than English, an [+ART] on the other, it may be the case that the L1s may have influenced the acquisition of the English definite article to a certain extent. This lack of a specific morpheme to indicate definiteness could have affected the ESL learners' acquisition of the English definite article *the* in ways resulting in parallel acquisition orders. The L1 Chinese learners' underlying structure of definiteness and indefiniteness is reinforced by their exposure to Malay. Although the L1 Chinese learners have also been exposed to Malay, the amount of exposure during their primary years was the same as their exposure to English although exposure to Malay did increase at the secondary level. We will assume that if Malay does play a role in the Chinese learners' acquisition of English *the*, then it would be minimal since the learner would have passed the critical period of acquisition (see Johnson & Newport, 1989).

Thus, it can be concluded that it is likely the acquisition of grammatical properties and their sub-categories of a second language is due in part to a natural order, resulting in a particular developmental pattern and systematicity across groups of learners from different L1 backgrounds on the one hand and in part to the influence of the learners' L1s on the other. More studies are required to determine which phenomenon is more dominant among the local ESL learners. Based on the findings obtained thus far, it is fair to say that they have implications for pedagogy which requires further investigation and discussion which is not within the scope of this study.

Notes
[1] CLS - Classifier
[2] GM – Generic marker
[3] 3PS – Third person singular pronoun

References

Archibald, J. (1996). Second language acquisition. In W. O'Grady, M. Dobrovolsky & F. Kambata (Eds.), *Contemporary linguistics: An* introduction (pp. 503–539). Longman: London and New York.
Brown, R. (1973). *A first language*. Boston, MA: Harvard University Press.
Chan, K.K. (1996). *"Itu" and the notion of definiteness in Malay*. Unpublished dissertation (M.M.L.S.), Universiti Malaya, Kuala Lumpur, Malaysia.
Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Oxford: Blackwell.
Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, *24*, 245–258.
Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*, 37–53.
Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York: Oxford University Press.
Ekiert, M. (2004). Acquisition of the English article system by speakers of Polish in ESL and EFL settings. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 4*(1), 1–23.
Goto Butler, Y. (2002). Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system. *Studies in Second Language Acquisition, 24*(1)*,* 481–488.
Hakuta, K. (1976). A case study of a Japanese child second language acquisition. *Language Learning*, *24*, 37–53.
Hawkins, J.A. (1978). *Definiteness and indefiniteness*. London: Croom Helm.
Huebner, T. (1979). Order-of-acquisition vs. dynamic paradigm: A comparison of method in interlanguage research. *TESOL Quarterly*, *13*, 21–28.
Huebner, T. (1983). *A longitudinal analysis of acquisition of English*. Ann Arbor: Korama press.

Ionin, T., & Wexler, K. (2003). The certain uses of the in L2-English. In J. M. Liceras, H. Zobl & H. Good-luck (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition* Conference (pp. 150–160). Somerville, MA: Cascadilla Press.

Ionin, T., Ko, H., & Wexler. K. (2004). Article semantics in L2-acquisition: the role of specificity. *Language Acquisition*, *12*(1), 3–69.

Johnson, J.S., & Newport, E.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Liu, D., & Gleason, J.L. (2002). Acquisition of the article the by nonnative speakers of English: An analysis of four non-generic uses. *Studies in Second Language Acquisition*, *24*(1), 1–26.

Lu, C.F.C. (2001). The acquisition of English articles by Chinese learners. *Second Language Studies, 20*(1), 43–78.

Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System*, *25*, 215–232.

Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). London: Hodder Arnold.

Nik Safiah Karim, Farid M. Onn, Hashim Hj. Musa & Abdul Hamid Mahmood (1997). *Tatabahasa Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, *37*, 127–152.

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, *16*(2), 35–172.

Thomas, M. (1989). The acquisition of English articles by first and second language learners. *Applied Psycholinguistics*, *10*, 335–355.

Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon: Multilingual Matters.

VanPatten, B., & Williams, J. (Eds.). (2007). *Theories in second language acquisition: An introduction.* Mahwah, N.J.: Lawrence Erlbaum.

Young, R. (1996). Form-function relations in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 135–175). Amsterdam: John Benjamins.

Appendix 1: Sample items from test instrument (adapted from Liu & Gleason, 2002)

I. Please write down:

> Your name : ………………………………………………………….
>
> Age   : …………
>
> Class   : ………………………………………………………….
>
> 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. Fendi bought a car on Monday. On Wednesday, he crashed car.
3. I read a book about Kuala Lumpur. Author, however, was from Penang.
5. Rocket ships are launched from Cape Canaveral in Florida.
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."
19. Shah Alam is capital of Selangor.
21. Man I met in New York later became my husband.
22. Blue car across the road is very suspicious.
33. Things of beauty always bring great joy.
37. Sally Ride was first woman in space.
38. Water in this glass is dirty.
40. A man says to his wife at the breakfast table, "Can you pass me newspaper?"
47. I like to read books about philosophy.
49. Sun is shining. It's a beautiful day.
55. President of the United States lives in White House.
56. My mother likes to have salads at dinner because salads are very healthy.