<table>
<thead>
<tr>
<th>Course code &amp; course title</th>
<th>Recommended readings</th>
<th>Author(s)</th>
<th>PDF Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG6210 Psycholinguistics and Second Language Acquisition</td>
<td><strong>Is the Acquisition Order of Grammatical Morphemes Impervious to L1 Knowledge? Evidence From the Acquisition of Plural -s, Articles, and Possessive ’s</strong></td>
<td>LUK Pei Sui Zoe and Yasuhiro SHIRAI</td>
<td>P.3-37</td>
</tr>
<tr>
<td>ENG6211 Sociolinguistics</td>
<td><strong>The linguistic landscape of Hong Kong after the change of sovereignty</strong></td>
<td>LAI Mee Ling</td>
<td>P.37-59</td>
</tr>
<tr>
<td></td>
<td><strong>Tracking language attitudes in postcolonial Hong Kong: An interplay of localization, mainlandization, and internationalization</strong></td>
<td>LAI Mee Ling</td>
<td>P.60-88</td>
</tr>
<tr>
<td>ENG6212 English Phonetics and Phonology</td>
<td><strong>Acoustic Analyses and Intelligibility Assessments of Timing Patterns Among Chinese English Learners with Different Dialect Backgrounds</strong></td>
<td>CHEN Hsueh Chu</td>
<td>P.89-115</td>
</tr>
<tr>
<td></td>
<td><strong>Chinese learners’ acquisition of English word stress and factors affecting stress assignment</strong></td>
<td>CHEN Hsueh Chu</td>
<td>P.116-127</td>
</tr>
<tr>
<td>ENG6216 Lexis, Morphology and Semantics</td>
<td><strong>Matching vocabulary learning process with learning outcome in L2 academic writing: An exploratory case study</strong></td>
<td>MA Qing</td>
<td>P.128-137</td>
</tr>
<tr>
<td></td>
<td><strong>Teaching Young Learners English Vocabulary with Reading-Based Exercises in a Real Classroom Situation</strong></td>
<td>MA Qing and SIN Ching Ha</td>
<td>P.138-151</td>
</tr>
<tr>
<td>Course code &amp; course title</td>
<td>Recommended readings</td>
<td>Author(s)</td>
<td>PDF Page</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>ENG6237 English in the Global context</td>
<td>Researching and teaching China and Hong Kong English</td>
<td>David C.S. Li</td>
<td>P.152-158</td>
</tr>
<tr>
<td></td>
<td>When does an unconventional form become an innovation?</td>
<td>David C.S. Li</td>
<td>P.159-175</td>
</tr>
<tr>
<td>ENG 6358 Curriculum and Assessment</td>
<td>Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism</td>
<td>David C.S. Li</td>
<td>P.176-188</td>
</tr>
<tr>
<td></td>
<td>Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling</td>
<td>XIE Qin and Stephen ANDREWS</td>
<td>P.189-211</td>
</tr>
</tbody>
</table>
REVIEW ARTICLE

Is the Acquisition Order of Grammatical Morphemes Impervious to L1 Knowledge? Evidence From the Acquisition of Plural -s, Articles, and Possessive ’s

Zoe Pei-sui Luk
University of Pittsburgh

Yasuhiro Shirai
University of Pittsburgh

In SLA, it has been often assumed that the effect of the first language (L1) is not very strong in the acquisition of grammatical morphemes (e.g., Ellis, 1994; Mitchell & Myles, 2004). However, such an assumption has not been systematically examined in the literature. This article reviews the morpheme studies conducted with native speakers of Japanese, Korean, Chinese, and Spanish to test the effect of the L1 in the acquisition of grammatical morphemes. The review reveals that although Spanish L1 learners’ acquisition order generally conforms to the “so-called” natural order (Krashen, 1977), native speakers of Japanese, Korean, and Chinese mostly acquire plural –s and articles later than, and possessive ’s earlier than, is predicted by the natural order. This indicates that learners can acquire a grammatical morpheme later or earlier than predicted by the natural order, depending on the presence or absence of the equivalent category in their L1. This suggests that L1 transfer is much stronger than is portrayed in many SLA textbooks and that the role of L1 in morpheme acquisition must be reconsidered.

Keywords morpheme studies; natural order; universal; L1 transfer

This research is supported by the Japan Iron and Steel Federation and Mitsubishi Endowments at the University of Pittsburgh and has been presented at the 27th Second Language Research Forum (University of Hawaii at Manoa, October, 17–19, 2008). We thank Kevin Gregg, Robert Schoenfeld, and three anonymous reviewers for their helpful comments on earlier versions of this paper.

We dedicate this article to the memory of Roger W. Andersen (1940–2008), whose contributions immensely helped establish SLA as a scientific field of inquiry. Although it was not possible to publish the article while he was still with us, it was fortunate that he knew we were working on this
Background

Brown (1973) investigated three children acquiring English and found that despite their different family backgrounds, the acquisition order of 14 grammatical morphemes in English was strikingly similar. Inspired by this ground-breaking study, many researchers conducted similar studies and confirmed Brown’s acquisition order of English grammatical morphemes in first language (L1) acquisition (e.g., deVilliers & deVilliers, 1973).

Dulay and Burt (1973, 1974) extended Brown’s (1973) research on morpheme acquisition to the area of second language acquisition (SLA). Their studies, as well as many subsequent studies (e.g., Bailey, Madden, & Krashen, 1974; Larsen-Freeman, 1975; Pica, 1983) suggested that there exists a universal acquisition order in the SLA of grammatical morphemes. Since then, there appears to be general agreement in the field of SLA that second language (L2) learners of English follow the pattern postulated by Krashen (1977; see also Krashen, 1981, Ch. 4), regardless of learners’ L1 background (Figure 1).

Many SLA textbooks, although noting the many methodological and conceptual weaknesses of morpheme studies, still take the position that the effect of L1 on the putative natural order is minimal. R. Ellis (1994), in summarizing morpheme studies by Krashen and associates, stated:

The picture that emerges from these studies is of a standard “acquisition order” that is not rigidly invariant but is remarkably similar irrespective of the subjects’ language backgrounds, of their age, and of whether the medium is speech or writing. (p. 94, emphasis added)

manuscript inspired by his 1983 paper “Transfer to Somewhere.” We thank him for encouraging us and providing us with some relevant information regarding Hakuta’s and Cancino’s work. His words are still relevant today as they were then:

It is comforting that the investigation of first language transfer has once again become respectable. Whatever one’s position on the relative importance of transfer in second language acquisition (SLA), transfer must be dealt with . . . . although we tried to make it go away, transfer kept coming back. And when it came back, we weren’t at all sure that it was really there. So now we again are trying to constrain transfer sufficiently to predict when, how, and to what extent it will influence the learner’s acquisition and use of a second language.”


Correspondence concerning this article should be addressed to Zoe Luk, Department of Linguistics, 2806 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA 15260. Internet: pel22@pitt.edu
More recently, Mitchell and Myles (2004) have stated:

However, the basic argument that both child and adult learners of English as a second language developed accuracy in a number of grammatical morphemes in a set order, no matter what the context of learning (classroom, naturalistic, mixed) survived the critique. . . . The existence of such an order suggested that second language learners are guided by internal principles that are largely independent of their first language; this was a serious blow for any proponents of Contrastive Analysis. (p. 43, emphasis added)

Likewise, Saville-Troike (2006) assumes the universality of morpheme acquisition order. Commenting on Dulay and Burt’s (1973) research, she stated:

Although not identical, the order of morpheme acquisition reported was similar in L1 and L2. Further, the order was virtually the same in English L2 whether children were L1 speakers of Spanish or Chinese. The existence of such a “natural order” strengthened claims for internally driven acquisition processes, which Dulay and Burt (1973) labeled creative construction . . . . The concept of natural order remains very important for understanding SLA, however, both from linguistic and from cognitive approaches . . . . These findings form part of the basis for continuing speculation that innate mechanisms for language acquisition may not be limited to early childhood (p. 43).
Thus, in many current SLA textbooks, the universality of morpheme acquisition order is emphasized, and in some cases, it is treated as a fundamental assumption on which theorizing in SLA is based. See also Ortega (2009, p. 124).

This emphasis on the universal order impervious to L1 effects is not unanimous, however. Gass and Selinker (2001), for example, are more skeptical of the claim of a natural order, stating that “there was some evidence even within these studies of the role of the NL [=Native Language]” (p. 113). They are, however, clearly in the minority, and SLA textbooks mostly “teach” introductory students of SLA that the order of acquisition is universal and that L1 influence is minimal.

Other scholars have pointed out that there is strong L1 influence on morpheme acquisition. Andersen (1983), citing Cancino (1976), pointed to the difference in acquisition order by an L1 Spanish learner and an L1 Japanese learner, attributing it to L1 transfer (more on this later). Sasaki (1987; see also 1986) compared the acquisition orders of five Japanese children acquiring English in naturalistic settings from three longitudinal studies (Hakuta, 1976; three learners from Koike, 1981; and her own study) and showed that all acquire possessive ’s earlier than in the proposed “natural order.” Sasaki (1987) further noted that only one of the five Japanese learners (Nobi from Koike, 1981) showed a significant correlation with the natural order and criticized the natural order hypothesis (Krashen, 1977), which treated Hakuta’s Japanese learner (Uguisu), a counterexample to the natural order, as an “idiosyncratic variation” (Burt & Dulay, 1980, p. 290). Terauchi (1994) also reviewed nine morpheme studies involving Japanese L1 learners and argued that, except for Makino (1979a), most of them did not show a significant correlation with the natural order.

Lightbown (1983) also reported a case of French learners of English showing a different accuracy order, which she attributed to the influence of the L1. She found that French L1 learners’ accuracy in their use of plural –s is lower than in other morpheme studies, which she attributed to the fact that final –s is not pronounced in French. Based on such observations, Shirai (1992) argued that “even though a natural order is claimed based on correlational or implicational studies, we still need to resort to an L1 transfer explanation even in the area of morpheme acquisition” (p. 100). More recently, N. C. Ellis (2006) also argued, based on Pak’s (1987) and Shin and Milroy’s (1999) morpheme studies on L1 Korean learners, that “L2 acquisition is clearly affected by the transfer of learners’ knowledge of their first language” (p. 187).
Yet the consensus among SLA textbooks, at least, is that there is a universal “natural” acquisition order, largely impervious to L1 influence. This may be because even though the evidence for such a strong L1 effect in morpheme acquisition has been pointed out in the literature, it has not been systematically tested. Even Goldschneider and DeKeyser (2001), an important study that conducted a comprehensive meta-analysis of morpheme studies, did not look at the effect of L1, due to the fact that many morpheme studies have lumped different L1 groups together. Therefore, it is imperative to assess the effect of L1 by a systematic review of previous morpheme studies.

This article presents such a review to explicate whether the proposed natural morpheme acquisition order is actually impervious to L1 effects. We focus on the morpheme accuracy orders of L2 learners of English whose L1s are Japanese, Korean, Chinese, and Spanish, for whom different acquisition orders have been reported in the literature (see the following section). The review shows that the acquisition of three morphemes—namely plural -s, articles, and possessive ’s—by native speakers of the three Asian languages deviates in large degree from the proposed natural order. We conclude that L1 transfer is not minimal but in fact is strong enough to cast doubt on the hypothesized universal acquisition order of grammatical morphemes.

**L1 Effects on Morpheme Acquisition Order?**

Weinreich (1953/1963) proposed that transfer is most likely to occur if the L1 and L2 structures are congruent and the resulting forms are free, invariant, and not complex. Andersen (1983) argued that these conditions can reasonably explain the difference between Spanish L1 and Japanese L1 learners by citing Cancino (1976), which compared the different morpheme acquisition orders of an L1 Spanish learner (Marta in Cazden, Cancino, Rosansky, & Schumann, 1975) and an L1 Japanese learner (Uguisu in Hakuta, 1976).

The top half of Table 1 (adapted from Table 4 of Andersen, 1983) shows how Spanish L1 learners transfer their L1 to English as a second language based on the following four conditions: (a) frequency in the target language, (b) whether they are free or bound, (c) whether they are congruent, and (d) whether there is phonetic similarity between L1 and L2. For example, articles are frequent in English. They exist in both Spanish and English as free morphemes, they are congruent in terms of usage, but phonetically the Spanish articles and the English articles are not similar. Because there are three favorable conditions
Table 1  Characteristics of morphemes that promote transfer from Spanish and Japanese

From Spanish to English (for Marta)

<table>
<thead>
<tr>
<th>Form</th>
<th>+/− transfer</th>
<th>Frequent in English?</th>
<th>Free/bound (L1 + L2)</th>
<th>Congruent? (L1 + L2)</th>
<th>Phonetic similarity? (L1 + L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article</td>
<td>+</td>
<td>Yes</td>
<td>Free</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Copula</td>
<td>+</td>
<td>Yes</td>
<td>Free</td>
<td>Yes</td>
<td>“Is”-yes</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>+</td>
<td>Yes</td>
<td>Free</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Plural</td>
<td>+</td>
<td>Yes</td>
<td>Bound</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>In*</td>
<td>+</td>
<td>Yes</td>
<td>Free</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>On*</td>
<td>−</td>
<td>Much less than in</td>
<td>Free</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Possessive</td>
<td>−</td>
<td>No</td>
<td>’s bound transferred</td>
<td>’s no N of/de N</td>
<td>No (but de like the)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N free</td>
<td>N yes</td>
<td></td>
</tr>
</tbody>
</table>

Go to for aux + going to − ? Transferred form free Qualified “yes” No

From Japanese to English (for Uguisu)

<table>
<thead>
<tr>
<th>Form</th>
<th>+/− transfer</th>
<th>Equivalent in Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article</td>
<td>−</td>
<td>No articles</td>
</tr>
<tr>
<td>Copula</td>
<td>−</td>
<td>Several different copulas</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>−</td>
<td>No plural</td>
</tr>
<tr>
<td>Plural</td>
<td>−</td>
<td>No plural</td>
</tr>
<tr>
<td>In*</td>
<td>+?</td>
<td>Similar to English</td>
</tr>
<tr>
<td>On*</td>
<td>+?</td>
<td>Similar to English</td>
</tr>
</tbody>
</table>


out of four, positive transfer is expected. In other words, Spanish learners of English should be able to acquire articles relatively easily.

The bottom half of Table 1, also from Andersen (1983), shows how L1 Japanese learners transfer their L1 structures to L2 English. In the case of Japanese, there are no articles and no plural markers, and, therefore, it is predicted that there will be no positive transfer; that is, Japanese learners of
English will acquire them relatively late. In contrast, Japanese possessives are similar to English (Ken no pen = Ken’s pen); thus, there is expected to be positive transfer. Note that in Table 1, the three features (Article, Plural, Possessive) have the opposite values (i.e., + vs. –), thus predicting different results for L1 Spanish and L1 Japanese learners. Cancino (1976) did find that articles and plurals were acquired early by Marta but delayed for Uguisu, whereas possessives were early for Uguisu and delayed for Marta (Andersen, 1983, pp. 183–186).

Andersen’s (1983) discussion of L1 transfer examined morpheme acquisition in only one learner from each L1. In the following sections, we review studies that have collectively examined a relatively large number of learners of L2 English in relation to morpheme acquisition. We do this in order to test the hypothesis that the ease (or difficulty) in acquiring a certain grammatical morpheme is highly related to whether the morpheme in the target language is semantically and structurally similar to the corresponding morpheme in the learner’s L1 and that the absence of a grammatical morpheme in the learner’s L1 will pose great difficulty for the learner in acquiring that morpheme in the target language. In particular, we focus on articles, plural, and possessive for which Andersen (1983) predicted different behaviors from L1 Spanish and L1 Japanese learners.

**Selection of the Studies Included in the Review**

The procedure for selecting the studies included in this review is as follows: We first surveyed the 25 studies examined for possible inclusion in the meta-analysis of Goldschneider and DeKeyser (2001). We used their list because they conducted an exhaustive computer database search on morpheme studies dealing with the acquisition of English that were conducted between 1973 and 1996, which identified 25 studies.³ We screened these 25 studies and retained all the studies⁴ that met the following criteria:

1. Reported acquisition or accuracy orders for separate L1 groups of Spanish, Japanese, Chinese, or Korean
2. Reported acquisition or accuracy orders for at least eight relevant morphemes in Krashen’s natural order, which involves nine morphemes.

This resulted in seven studies from Spanish L1 groups and two studies from Japanese groups, one from Korean groups, and two from Chinese groups. Because we intended to test the different predictions regarding Spanish and Japanese L1 learners made by Andersen (1983) and we also wanted to test
the same prediction for Korean and Chinese learners whose L1s have similar characteristics regarding the three morphemes in question, we did an additional search for Japanese, Korean, and Chinese L1 groups and added the studies that also met the above two criteria. We then excluded one study (Higuchi, 1986) that looked at a very young child (1- and 2-years-old when studied), as this case arguably involves the simultaneous acquisition of two languages, not English as a second language (ESL) (Goldschneider & DeKeyser, 2001, p. 2). We also excluded case studies other than longitudinal studies (e.g., Schmidt, 1983). This yielded seven L1 Japanese studies (Hakuta, 1976; Izumi & Isahara, 2004; Koike, 1983; Makino, 1979a; Nuibe, 1986; Shirahata, 1988; Sasaki, 1987), two L1 Chinese studies (Dulay & Burt, 1974; Mace-Matluck, 1979), two L1 Korean studies (Pak, 1987; Shin & Milroy, 1999), and seven L1 Spanish studies (Andersen, 1978; Bailey et al., 1974; Dulay & Burt, 1973, 1974; Mace-Matluck; Pica, 1983; Rosansky, 1976). Tono (2000), which was called to our attention after this article went to press, is also consistent with our hypothesis. He analyzed a corpus of Japanese EFL students’ writing and showed that possessives are acquired earlier, and plurals and articles are acquired later, than in the natural order.

The details of the studies included in the present review are listed in the Appendices. Appendix A lists the studies conducted with L1 Japanese learners of English, Appendix B lists the studies with L1 Korean learners, Appendix C lists L1 Chinese learners, and Appendix D lists L1 Spanish learners. Although some of these studies involve multiple L1 groups, they at least report results from Japanese, Spanish, Korean, and Chinese as distinct L1 groups.

Acquisition Orders of Learners From Various L1 Backgrounds

As discussed earlier, the proposed natural order (hereafter NO) seems to be violated in some studies. In particular, plural –s, articles, and possessive ‘s have been argued to show acquisition orders different from the NO, due to language transfer (Andersen, 1983). Thus, the following review will focus on these three structures, and they will be highlighted in each table in shaded rows. We first review the studies on Japanese, Chinese, and Korean learners, for which we make the same predictions. We then review studies on Spanish L1 learners, for which contrasting predictions are made (Andersen).

In the review of the morpheme studies, we did not try to come up with generalizations that are due to some methodological factors. As has been abundantly pointed out in the literature (e.g., Ellis, 1994; Hatch, 1983), there are numerous methodological problems involved in the morpheme studies, and...
different studies used different methods, which makes generalization quite impossible. Thus, we intentionally did not go into methodological details of each study, other than the basic information included in the Appendices. Rather, what we show in this review is that even with such limitations due to different methodologies and contexts of data collection, we still obtain a rather clear picture of L1 influence, which invalidates the so-called natural order.

**Acquisition Order by the Japanese Learners of English**

Table 2 shows the results of the studies on Japanese learners of English, in comparison with Krashen’s NO. Although some of the studies investigated more than 9 grammatical morphemes (e.g., Hakuta, 1976, studied as many as 17), the ranking in Tables 2 to 5 is determined by considering only the 9 grammatical morphemes in the NO. The rankings in Tables 2 to 5 are based on accuracy orders for large-scale cross-sectional studies (except for Andersen, 1978; Nuibe, 1986; Shirahata, 1988, which used implicational scaling to determine the difficulty order with the criterion of 80% correct), whereas for longitudinal case studies, the rankings are based on the order of acquisition (i.e., based on the order in which each morpheme attained the criterion for acquisition).5

The first three studies in Table 2 are longitudinal case studies of Japanese children (aged 5 to 10 years) acquiring English in the United States using spontaneous oral interaction data, whereas the rest of the studies are cross-sectional studies from postpuberty learners learning English in Japan, using oral interview data (Izumi & Isahara, 2004; Shirahata, 1988) or paper-and-pencil tests (Makino, 1979a; Nuibe, 1986). Despite the different contexts of learning and methods of data collection and analysis, it is clear from Table 2 that L1 Japanese learners show a clear deviation from the NO for the three morphemes in question: Articles and plural –s are generally lower than the NO, whereas possessive is much higher, supporting the L1 transfer hypothesis. The exceptions to this trend, however, are Makino (1979a) and two boys from Koike (1983).6

Although Makino (1979a) admitted that his results show a slight influence from the L1 of his participants (e.g., third-person singular –s was acquired late because of the lack of corresponding morpheme in Japanese), he concluded that his results correlate highly with other morpheme studies (e.g., Dulay & Burt, 1974) and thus support the hypothesis of an invariant order of English morpheme acquisition. However, his study seems to be the only one that reflects little L1 effect. This study, in its various forms (Makino, 1979b, 1980) is cited by many, including Krashen (1985, p. 20) and Ellis (1994, pp. 627–628), as constituting important evidence for the NO and therefore deserves careful scrutiny.
All studies in Table 2, except for Makino (1979a) and Jun and Nobi in Koike (1983), show that articles are acquired later by Japanese learners of English than is predicted by the natural order hypothesis (NOH). As Japanese does not have any system of articles, it is not surprising that Japanese native speakers have difficulty mastering articles. As DeKeyser (2005, p. 5) puts it, articles are “notoriously hard to acquire for learners whose L1s do not have them,” because of their “novelty” and “abstractness.”

Why, then, does Makino’s (1979a) study show quite high accuracy for articles (rank 2, 2, 3, even higher than the NO rank of 4)? There seem to be two reasons why, in his study, articles are ranked second, whereas in other studies they are ranked much lower. First, whereas most of the other studies in Table 2 analyzed oral data from the participants, Makino used a paper-and-pencil grammar test, for which participants were able to monitor their own performance. Second, Makino’s grammar test was particularly restricted, in that there are only four instances where articles were obligatory. Three of them are _a_ as in “has a pen” and “writing a letter.” The remaining one is superlative _the_ as in “is the (youngest).” This seems to raise two problems that threaten the validity of Makino’s test items as a measurement for the mastery of articles.

First, the difficulty of articles for learners partly lies in the distinction between definite vs. indefinite meaning (Hakuta & Cancino, 1977), and Makino’s (1979a) tests cannot test the mastery of this distinction. In other words, these learners do not have difficulty in _supplying_ articles, but they have difficulty in deciding _which_ article to supply. Hakuta’s (1976) longitudinal study showed that his learner, Uguisu, started to use articles quite early, but she had difficulty contrasting between definite and indefinite for a long time. The reason why the Chinese learners in Dulay and Burt’s (1974) study seemed to acquire articles early may be because Dulay and Burt did not distinguish between _a_ and _the_ in scoring, as Hakuta and Cancino noted. Dulay and Burt adopted the same scoring method as Brown (1973). Brown explained that because there were too many contexts under which an article was required, it was difficult to determine whether the reference was specific; therefore, _a_ and _the_ were put under a single category. For the three instances in Makino’s study, they are all indefinite contexts. This only gives us half of the picture of how learners used definite/indefinite articles. Learners may even stick to one form because they are not sure which one to use. This then gives us the false impression that they have mastered articles.

The second problem is that the only item that required the definite article _the_ was different from the other three items in the sense that it was not followed by a noun. Instead, it is used with a superlative form of an adjective. This is
### Table 2: Comparison of the natural order and the acquisition orders for Japanese learners of English

| Study         | Morpheme order | Method 1 | Method 2 | 2nd  | 3rd  | 2nd  | 3rd  | 2nd  | 3rd  | 2nd  | 3rd  | 2nd  | 3rd  | Method 1
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural order</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Koike (1983)</td>
<td>Progressive</td>
<td>1</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Sasaki (1979)</td>
<td>Plural</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Makino (1976)</td>
<td>Copula</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Hakuta (1976)</td>
<td>Auxiliary</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Nobi (1987)</td>
<td>Article</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sachiko (1987)</td>
<td>Past-irregular</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Jun (1988)</td>
<td>Past-regular</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Nobi (1987)</td>
<td>Third person</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Isahara (1988)</td>
<td>Possessive</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Ishihara (2004)</td>
<td></td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Hakuta (1976)</td>
<td>Article</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Nobi (1987)</td>
<td>Past-irregular</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sachiko (1987)</td>
<td>Past-regular</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Jun (1988)</td>
<td>Third person</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Ishihara (2004)</td>
<td></td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Method 1 counts all answers either “right” or “wrong.” Method 2 gives partial credits to some answers. 2nd and 3rd refer to second and third graders of junior high school, respectively. I, II, and III refers to first, second, and third graders in high school.

This refers to the definite article; the indefinite article was ranked last (10).
problematic because “the (adj)-est” can be considered a simple rule that can be learned relatively easily; indeed, this rule is considered one of the important rules to be taught to students in junior high school in Japan. Therefore, it does not tell us much about whether the participants had mastered the, to say nothing of whether they can actually use it correctly in spontaneous production.

The relatively late acquisition of plural -s can also be attributed to the lack of obligatory plural marking in Japanese, as predicted by Andersen (1983) and Cancino (1976). While the NO ranks plural -s among the first morphemes to be acquired, most studies in Table 2 show low rankings of plural -s, (between rank 9 and 5), except for Makino’s (1976a) study (2 or 3), and the three children from Koike (1983) (1 or 3). The lack of plural marking and noun countability also makes it difficult for Japanese learners of English to distinguish between mass and count nouns, which has been argued to be a main source of difficulty for them, since a non-definite singular count noun requires the article a, and a non-definite mass noun does not (Hiki, 1991; Butler, 2002).

Now let us turn to the acquisition of possessive ’s. Although possessive ’s is among the last to be acquired in Krashen’s NO (ranking 7), all studies that we included in Table 2 indicate that possessive ’s is acquired much earlier by Japanese learners of English (ranking 1, 2, or 3, except for Makino, 1979a, and Nuibe’s, 1986, studies involving paper-and-pencil tests, for which ranking was mostly 4 or 5, still earlier than the NO). As noted earlier, this can be attributed to the fact that the structure of a possessive construction in English is very similar to that in their native language. In Japanese, the possessive is expressed in the form “A no B,” where A is the possessor, B is the object being possessed, and genitive no functions like ’s in English, as in Ken-no pen “Ken’s pen” (see Andersen, 1983). With this similarity, learners can easily produce possessive ’s, because it only involves mapping it to the Japanese counterpart. We will see later that possessive ’s is extremely difficult for Spanish learners of English, precisely because of the lack of an equivalent construction in their native language.

Acquisition Order for Korean Learners of English
Korean is similar to Japanese in many ways. There is no article in Korean. There is a morpheme –tul that functions like a plural marker, but its application is optional. Moreover, it can only be used with quantifiers such as manun “many,” “some,” “a few,” and so forth and cannot be used with numerals such as “two” (Jin, 2003). Korean also has a genitive marker –uy, which also attaches to the possessor as a suffix, and the resulting NP will have the same structure as
Table 3 Results of studies with Korean learners of English compared with the natural order

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Plural</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Copula</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Article</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Past-irregular</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Past-regular</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Third person</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Possessive</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

English (e.g., “John’s car” in Korean will be John-uy cha). Thus, the prediction for Korean learners of English should be the same as for Japanese learners.

Both Pak (1987) and Shin and Milroy (1999) investigated children acquiring English in the United States; Pak compared the children with adults. Pak, using the Bilingual Syntax Measure, found a significant correlation between children and adults, whereas Shin and Milroy, using oral data from classroom activities, found a significant correlation with Pak’s study, suggesting a similar acquisition order among Korean learners of English.

Regarding our predictions, articles are acquired relatively late by Korean learners of English, as we can see in Table 3. Although articles are in the second group (rank 4) to be acquired according to the NO, they are ranked sixth and seventh in Pak’s (1987) and in Shin and Milroy’s (1999) study, respectively. Plural, which ranked first in the NO, ranked eighth and ninth in the studies with L1 Korean learners. In contrast, possessive ‘ş’, one of the most difficult morphemes in the NO, ranked relatively high (4 and 2) for L1 Korean learners. These results clearly show that the absence or presence of a grammatical morpheme in the learner’s L1 will have a strong effect on the acquisition of that grammatical morpheme in English by Korean learners, as is the case for Japanese learners.

Acquisition Order for Chinese Learners of English
Chinese does not have plural marking or an article system, but it has a structure of denoting possession that is similar to that of English (Ken de bi = Ken’s pen). Therefore, we predict the same pattern of acquisition of morphemes by
L1 Chinese learners as Japanese and Korean learners.⁹ As is clear from Table 4, there are not many studies involving Chinese learners of English as a distinct L1 group, although there are studies with Chinese learners whose results were combined with those of learners with other L1s (e.g., Bailey et al., 1974).

The studies by Dulay and Burt (1974) and Mace-Matluck (1979) are both cross-sectional studies with a large number of children (55 and 442, respectively) learning English in the United States, the former using the Bilingual Syntax Measure and the latter using the SEA-CAL Oral Proficiency Test. Table 4 shows that the pattern of acquisition for L1 Chinese learners is similar to that of Japanese and Korean learners of English. First, Chinese L1 learners acquire plural later (rank 5 and 6) than the NO (rank 1), suggesting that the absence of a plural morpheme in Chinese causes difficulty for Chinese learners.

Dulay and Burt (1974), however, who conducted a study with 60 Spanish-speaking and 55 Chinese-speaking children, showed that the acquisition orders for the two groups are basically the same. In particular, their Chinese participants acquired articles quite early (see Table 4). They thus concluded that L1 has little effect on the acquisition order.

Here, too, Hakuta and Cancino’s (1977) explanation—the lack of distinguishing definite from indefinite articles in scoring—may apply. As noted earlier, Dulay and Burt (1974), following Brown (1973), did not distinguish definite and indefinite articles in their scoring. If many indefinite articles used by the participants in Dulay and Burt’s study were intended to convey definite meanings by the participants, or vice versa, the errors will be undetectable,

---

**Table 4** Results of studies with Chinese learners of English

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Plural</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Copula</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Auxiliary</td>
<td>4</td>
<td>4</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Article</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Past-irregular</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Past-regular</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Third person</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Possessive</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
and this produces the illusion that the Chinese participants could use articles in English correctly most of the time. Hakuta and Cancino argued that the most difficult problem for Japanese and Chinese learners is the distinction between definite and indefinite articles. In other words, these learners may not have much difficulty in supplying articles but may have difficulty in deciding which article—definite or indefinite—to supply.

Regarding the possessive marker, Mace-Matluck’s (1979) data from L1 Cantonese learners support the L1 transfer hypothesis in that the possessive is acquired earlier (rank 3) than in the NO (rank 7). In contrast, Dulay and Burt’s (1974) results from Chinese learners (rank 8) conform to the NO. We have no explanation for this discrepancy.

The review thus far makes it clear that in most cases, Japanese, Korean, and Chinese learners deviate from Krashen’s NO and acquire the possessive morpheme much earlier, and plurals and articles much later, than predicted by the NO, but as predicted by Andersen’s (1983) L1 transfer account (i.e., the lack of articles/plural marking, and the existence of similar possessive marking in the L1s). Some of the clear deviations (Dulay & Burt, 1974; Makino, 1979a) can be accounted for as methodological artifacts, although some might have to be attributed to individual differences (e.g., two boys from Koike, 1983). In the next subsection we turn to Spanish L1 learners.

**Acquisition Order for Spanish L1 Learners of English**

We now turn to the acquisition of English morphemes by Spanish L1 learners. These results are summarized in Table 5. The studies are mostly from learners acquiring English in the United States, except for Pica (1983), who also included English as a foreign language (EFL) learners in Mexico, and Andersen (1978), who studied EFL learners in Puerto Rico. They are all cross-sectional studies (except for Rosansky, 1976, which also had a longitudinal component that was not included in our review because it did not meet our criteria). Dulay and Burt (1973, 1974) and Mace-Matluk (1979) studied child learners, whereas Bailey et al. (1974), Andersen, and Pica (1983) studied adults. (Rosansky’s study included children, adolescents, and adults, two of each). Although Andersen used written composition data, most studies used oral data: the Bilingual Syntax Measure (Bailey et al., 1974; Dulay & Burt, 1973, 1974), the SEA-CAL Oral Proficiency Test (Mace-Matluk, 1974), and natural conversation (Pica, 1983; Rosansky, 1976). These studies are generally considered to be consistent with the NO, except for that by Rosansky, which raised some doubts on the universality of the NO by showing that the acquisition order is dependent on the method of data collection.
### Table 5 Results of studies with Spanish L1 learners of English

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Plural</td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Copula</td>
<td></td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Auxiliary</td>
<td></td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Article</td>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Past-irreg</td>
<td></td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Past-regular</td>
<td></td>
<td>7</td>
<td>n/a</td>
<td>n/a</td>
<td>5</td>
<td>n/a</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Third person</td>
<td></td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Possessive</td>
<td></td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

*Note.* Sac: participants from Sacramento; San: participants from San Ysidro; EH: participants from East Harlem; Instr.: instructed learners; Nat.: naturalistic learners; Mixed: learners received input from both naturalistic and classroom settings.

*<sup>a</sup>The data are from their Group Score Method (Table 4, p. 51).

<sup>b</sup>The data are based on the group mean from the cross-sectional data (Table 6, p. 417). Rosansky’s longitudinal analysis (from one adolescent, Jorge) only ranked six morphemes and did not meet our criteria for inclusion in this study.
By showing that the Chinese and Spanish learners exhibit similar accuracy orders for various grammatical morphemes, Dulay and Burt (1974, p. 52) concluded that their studies “provide a strong indication that universal cognitive mechanisms are the basis for the child’s organization of a target language” and that it is not the “L1 system that guides the acquisition process.” Bailey et al. (1974, p. 242), while admitting that there may be some interference from learners’ L1, also concluded that “a major source of errors is intra- rather than inter-lingual” and that the notion of “universal language processing strategies” remains the best explanation of the results.

As we see from Table 5, the results obtained from various Spanish L1 studies do seem to be consistent with the NO, but an L1 transfer explanation also seems possible. Table 5 shows that L1 Spanish learners of English acquire possessive ‘s quite late in these studies. One explanation is that possessive ‘s, as predicted by the NOH, is difficult for L2 learners and that the L1 Spanish learners went through the stages as predicted. However, if we look at the possessive construction in Spanish, we may also be able to say there is L1 transfer (Andersen, 1983). In Spanish, the possessed is ordered before the possessor, as in “the house of my mother.” Because of this difference, it is not surprising that Spanish learners have difficulty in acquiring possessive ‘s in English. Moreover, as we saw earlier, L1 Japanese and Korean learners have much less difficulty in acquiring possessive ‘s, and in fact it is one of the earliest morphemes that they acquire, contra the NO, which ranks the possessive morpheme last. Therefore, it is warranted to interpret the low ranking of possessive ‘s among Spanish learners of English as being due to the influence of their L1.

Furthermore, we can actually find utterances of Spanish native speakers that support this interpretation. Cancino (1976) cited such utterances (e.g., the books de Mommy) often produced by Spanish L1 speakers. These utterances indicate that Spanish learners of English do rely on their L1 to produce utterances containing possessive constructions and that such negative transfer delays the acquisition of the English possessive.

Articles and plurals are generally consistent with the NO, but some deviations are apparent. First, Dulay and Burt (1974) and Bailey et al. (1974) have very high ranks (1) for articles. Second, plurals were markedly low in many studies, although they are ranked 1 in the NO. In particular, the studies by Dulay and Burt (rank 4), Andersen (1978) (rank 5), and Pica (1983) (rank 5 and 4) deviate from the NO. It is not clear why this is the case. This also indicates that even though these studies are often considered to support the NO (e.g., Ellis, 1994), their agreement with the putative universal order is only approximate, although this is often masked by rank order correlations.
Additional Considerations

The above review suggests that in the acquisition of three morphemes (plural –s, articles, and possessive ’s), the effect of the L1 is very strong and that the NO is often violated due to L1 influence. Here, we discuss studies that are not included in the preceding focused review but are relevant to the issue we are addressing in this article.

The work by Fathman (1975) is a study that is very relevant, although it was not included in the tables because it only reported the results for four morphemes in the NO. Fathman compared 60 Spanish L1 ESL learners with 60 Korean L1 ESL learners on the Second Language Oral Production English (SLOPE) test, which measured 20 grammatical structures in English, and found that their accuracy levels on various morphemes were quite similar. Fathman concluded that “it appears that the sequence of learning of these structures was constant despite the fact that the individuals tested spoke native languages which were quite dissimilar” (p. 36). However, she also reported that the only statistically significant difference between the two groups was found with articles, for which the Spanish L1 group performed significantly better (p. < .01) than the Korean group, which is consistent with our prediction.10 She also found a “slight but statistically insignificant difference” for possessive ’s: the Korean L1 group performing much better than the Spanish L1 group, which, again, is consistent with our hypothesis. These are the two items for which there were the largest differences between the two L1 groups in accuracy scores (p. 37, Fig. 1).

In a nutshell, although the L1 effect has been discounted in the literature on morpheme acquisition, it appears that there are notable L1 effects, explaining why Spanish L1 learners show conformity with the NO, whereas Japanese, Chinese, and Korean learners deviate from it. Why, then, has there been a general acceptance of the idea of a universal order impervious to L1 effects? There may be several reasons.

First, it is possible that the so-called universal order is largely dependent on L1 Spanish learners, who were the most strongly represented population throughout the morpheme order studies. Goldshneider and DeKeyser’s (2001) meta-analysis of morpheme acquisition studies, which assumed and tried to identify an explanation for the common order of acquisition, reported the number of participants analyzed by L1 in their Appendix E. Spanish (N = 354) is by far the largest group.11

Further, many studies just lump the learners of different L1s together without reporting scores for each L1 group (e.g., Bailey et al., 1974; Krashen, Sferlazza, Feldman, & Fathman, 1976). Additionally, even when researchers
have tried to assess the effects of different L1s, they have sometimes lumped together various L1 groups. Krashen et al. (1976), for example, divided learners into four L1 groups (Spanish, Greek, other Indo-European, and non-Indo-European), which masks the differences among the languages categorized in the latter two groups, and, therefore, the effects of different L1s, even if very strong, may cancel out each other and will not show up in the results.

Finally, in the 1970s, when the morpheme studies on which the NO was based were carried out, researchers were more interested in universal aspects of SLA rather than language-specific influences relating to either the target language or the source language. They did this in reaction to the contrastive analysis hypothesis, which predicted very strong L1 effects (see, e.g., Kellerman, 1979, for a discussion).

In hindsight, the SLA community should have paid more attention to Hakuta’s (1976) study. Although Krashen (1981) and Burt and Dulay (1980) were puzzled by Hakuta’s Uguisu, who stood out as not conforming to the general trend, they did not pursue this. In Krashen’s Table 4, in which he compared Spearman correlation coefficients between various morpheme studies and his NO, Hakuta’s Uguisu had a correlation of .170, whereas all of the other studies using spoken data had a correlation coefficient higher than .700, except for one study (.650), and many had a correlation coefficient higher than .800. The use of rank order correlation was criticized for masking real differences (e.g., Hatch, 1983), but even then this deviation was crystal clear, and the logical next step should have been to look at other L1 Japanese learners. Several Japanese researchers conducted such studies in the 1980s (see Table 2), but their results were not introduced to the mainstream SLA community.

Although the assumption of a universal morpheme acquisition order continues in SLA textbooks, SLA research has continued to pay attention to L1 transfer in relation to the morphemes that we have focused on in this article and has shown that L1 influence plays an important role in explanation of studies examining the acquisition of these morphemes—in particular, articles.

Ringbom (1987) investigated the acquisition of articles by Swedish- and Finnish-speaking learners of English by analyzing essays written for a matriculation examination and found that the Finnish-speaking learners of low and intermediate proficiency levels used articles much less frequently than both the English native speakers and the Swedish-speaking learners of corresponding proficiency levels. He suggested that this is because the Finnish participants avoided the articles, as they do not occur in their L1 and are thus perceived as “difficult,” and concluded that this differential frequency “is the result of covert cross-linguistic influence” (p. 108).12
Apart from negative transfer, L1 influence can also give learners advantages in learning articles. Master (1987) showed that learners whose L1 has an article system (Spanish and German) learn the corresponding system more easily than learners whose L1 lacks one (Chinese, Japanese, and Russian). Odlin (1989) argued that although Spanish native speakers, like Chinese and Japanese native speakers, also omit articles in English, they tend to use articles more accurately than those whose native language does not have articles. Ringbom (1987) made a similar argument, criticizing Dulay and Burt (1974), who ignored the fact that Spanish children perform consistently better than Chinese children.

Jarvis (2002) examined the acquisition of articles at the discourse level. He showed a silent Chaplin film to Swedish-speaking and Finnish-speaking adult learners of English and asked them to write a narrative of the film. He found that Swedish speakers were likely to mark the distinction between new and not-new information with the use of articles, whereas Finnish speakers were less likely to do so. He attributed the results to the fact that it is obligatory to mark this distinction in Swedish but not in Finnish. These studies clearly show the effect of the L1 in morpheme acquisition, supporting the transfer hypothesis.

The Cognitive View of L1 Transfer

Recent advances in cognitive approaches to language acquisition research have brought back to prominence the idea of L1 influence in the processing and learning of L2 structures. The basic idea is that once an L1 is acquired, one cannot process an L2 without the filter of the L1. Ellis (2006, p. 165), for example, gave a detailed account on how the L1 can affect the learning of the L2. He argued:

- the linguistic forms that L2 learners fail to adopt and to use routinely thereafter in their second language processing are those which, however available as a result of frequency, recency, or context, fall short of intake because of one of associative learning factors of contingency, cue competition, or salience, or because of associative attentional tuning involving interference, overshadowing and blocking, or perceptual learning, all shaped by the L1.

To illustrate, overshadowing refers to the phenomenon whereby a cue, when accompanied by a known stronger predictor, will not be seen as a predictor of the outcome. For example, because Japanese does not have any plural markers, Japanese native speakers are trained to interpret plurality from other sources, such as discourse and context. When they learn the plural –s, the stronger
cue (i.e., discourse and context) overshadows the marker –s. This may prevent them from processing the plural marker as an important piece of information. This may explain why the absence of a morpheme in a learner’s L1 will create difficulty for the acquisition of that morpheme in the L2.

Similar ideas are proposed to account for the age effect (including critical period effects) in L2 learning, where adult L2 learners cannot attain nativelike proficiency in the target language. Hernandez, Li, and MacWhinney (2005) proposed a unified model of L1 and L2 acquisition that relies on a self-organizing neural network model (DevLex; e.g., Li, Zhao, & MacWhinney, 2007). In this model, late L2 learning is considered to be “parasitic” on the network of L1 language representation, which has become automatized (i.e., entrenched) over years of language processing in the L1; this means that L2 items will be interspersed with the L1 forms on which they depend, rather than clustering in a separate region of the network. Because for adult learners L1 representation is highly entrenched as a result of many years of experience with the language, it is very difficult to create a new separate system of L2 representation, which becomes even more difficult as one’s experience with L1 increases. Thus, Hernandez et al. argued that there is no need to appeal to a biologically determined “critical period” for language acquisition.

This view, which assumes L1 knowledge as a filter through which L2 knowledge is built, has similarities to the long-held view of phonological development. It is well known that children are born with the ability to distinguish any phonemes in any language (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), but this ability will decline sharply between 6 and 12 months of age after exposure to the native language (Werker, Gilbert, Humphrey, & Tees, 1981). For example, Japanese children can distinguish /r/ from /l/ initially, but after 1 year of exposure to Japanese, they cannot do so any longer because they learn to disregard the difference between the two because the difference is unimportant in processing Japanese. This is a disadvantage in learning English or other L2s for which such a phonemic distinction is important, but it is perfectly efficient for using Japanese. In other words, learning a native language is a process of learning how to ignore unimportant distinctions (see Kuhl, 2000, for further discussion).

In the area of semantics, a similar loss of sensitivity is reported. McDonough, Choi, and Mandler (2003) showed, by preferential-looking experiments, that not only infants learning Korean but also infants learning English are sensitive to the distinction between tight versus loose containment lexicalized in Korean spatial verbs (kkita vs. nohta), although this distinction is not made in English. In contrast, English-speaking adults are not sensitive to this distinction; only Korean-speaking adults are. This also suggests that by
learning our native language, we lose sensitivity to the distinctions that are not important in the language.

Thus, in various linguistic domains, learning a native language involves acquiring the ability to process it efficiently and learning to ignore—or losing the ability to make—the distinctions that are unimportant in the language. In other words, one’s native language works as a filter through which one processes incoming signals of an L2. This picture is not consistent with the view that morpheme acquisition order is universal or impervious to L1 effects. Rather, because L2 learning occurs through the filter of the L1 network, it is only natural that there are very different acquisition orders for different L1 groups, rather than a universal natural order.

Conclusion

In this article, we reviewed in detail various L2 English morpheme studies with learners of different L1s (Japanese, Chinese, Korean, and Spanish). We have shown that the acquisition order of grammatical morphemes is highly affected by the learner’s L1 such that it is possible to predict, to some extent, what is difficult and what is easy for language learners based on their L1s. Since the study by Dulay and Burt (1974), the “invariant” acquisition order of grammatical morphemes has been seen as evidence for the claim that language learning goes through universal processes impervious to L1 transfer. However, this article shows that L1 effects are strong, which means that learners are indeed heavily influenced by the previous knowledge of their native languages.

Since the publication of the study by Goldschneider and DeKeyser (2001), the interpretation of morpheme acquisition studies has entered a new stage. In particular, Goldschneider and DeKeyser reexamined the explanation for a universal order of morpheme acquisition through a sophisticated meta-analysis, which showed that the combination of five factors (perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency) can predict about .71 of the variance observed in the studies included in their meta-analysis. These roughly correspond to what Hatch (1983) called the “naturalness factors” that determine the more or less invariant order of morpheme acquisition, and we stress that these factors are at work universally in contributing to the morpheme acquisition order. We accept the universal aspects of morpheme acquisition but reject the existence of a universal order of acquisition, as posited by Krashen (1977) and presented as such in many SLA textbooks. Universal factors and L1 transfer factors interact, as has been abundantly shown in previous research in SLA (e.g., Gass, 1981; Zobl, 1982). Morpheme acquisition is not an exception to this generalization.
Our goal in this article was to show that L1 transfer is not minimal, as has been put forth by many researchers who have argued for a universal learning mechanism. L1 transfer has to be dealt with. It is clear that many SLA textbooks underestimate the effect of L1 in their discussion of morpheme studies and need to rethink the conventional wisdom.

Revised version accepted 29 October 2008

Notes

1 Our use of the term “natural order” is for convenience only; that is, by “natural order” we refer to the acquisition order proposed by Krashen (1977). In fact, we reject the very existence of a natural order.

2 Ellis (1994) on the next page (p. 95) discussed some longitudinal studies that go against the natural order (Hakuta, 1974; Rosansky, 1976; Schmidt, 1983) and pointed out that articles are late in acquisition for Hakuta’s and Schmidt’s L1 Japanese learners. However, he did not mention L1 as a possible cause for deviation, suggesting rather that the longitudinal nature of these studies may be the key. Incidentally, Wes (Schmidt) had plurals ranked 5th and articles 7th in his accuracy order, both lower than predicted by the natural order hypothesis, but possessive ranked 8th as predicted.

3 Goldschneider and DeKeyser (2001) eventually excluded 13 studies for various reasons (e.g., EFL data, written data, lack of scores reported, etc.) and retained 12 studies for their meta-analysis.

4 Instead of Andersen (1977), we used Andersen (1978) published in Language Learning, as both are based on the same data and the latter is more accessible.

5 All of the longitudinal studies used a criterion of 90% or 80% suppliance in obligatory contexts or targetlike use.

6 We do not have an explanation for why two boys from Koike (1983) did not show L1 effects, whereas one girl did, which we have to attribute to individual difference factors. In fact, Sasaki (1987), who compared the acquisition orders of five Japanese children, pointed out that the three girls showed very similar orders to each other and so did the two boys, suggesting that gender may be a possible contributor.

7 The other exception was the study by Nuibe (1986), which used a written translation task.

8 See Terauchi (1994, pp. 27–28) for a related discussion of Makino’s (1979a) study.

9 Some researchers suggest that Mandarin Chinese has an indefinite article. For example, Siegal (2000) claimed that the specific–nonspecific distinction in Hawaiian Creole English was a result of the influence of Mandarin Chinese, which uses the word yi “one” to denote indefinite but specific noun phrases. However, the mapping between yi in Chinese and a in English is not straightforward. The use of yi is often optional, because the definite–indefinite distinction is expressed through
context, sentence type, the position of the noun in relation to the verb in the sentence, and so forth (Yip & Rimmington, 2004). For example, shu “book” as in wo qu jie shu “I am going to borrow a book” is in the postverbal position and can be of indefinite reference without the use of yi (Yip & Rimmington). Because of this complexity, how the acquisition of the English indefinite article is influenced by L1 Chinese is not clear. Further, Zobl (1982) argued that the absence of articles in Chinese delays the acquisition of articles in terms of the developmental sequences and rates of learning by which they are acquired. He compared the Spanish child in Hernández-Chávez (1977) with the Chinese child in Huang (1971), whose age and acquisition contexts were comparable to each other. Zobl showed that whereas the Chinese child tended to use demonstratives (this), which has an equivalent in Chinese, in place of definite articles in early stages, the first spontaneous production of definite articles did not occur until much later, the Spanish L1 child was able to use definite articles much earlier without going through the stage where he used demonstratives in place of the. Thus, at least the lack of definite articles in Chinese appears to delay the acquisition of articles in English.

10 Hakuta and Cancino (1977) as well as Burt and Dulay (1980) also pointed out the discrepancy in the accuracy of articles between Korean and Spanish learners in Fathman’s study.

11 The second largest is Cantonese (N = 175), but this group is represented by only one study (Mace-Matluk, 1979). The third largest group was much smaller (Korean, N = 66), but, again, this group is represented mostly by Fathman (1975; N = 60), and these studies supports our hypothesis, as discussed earlier.

12 Ringbom (1987) also noted that Finnish learners of English might view articles as redundant because notions such as noun countability and definiteness can be determined through other cues in Finnish, which, in turn, has made the acquisition of English articles difficult for them.

13 It should be noted that this ability is not unique to humans (Kuhl & Miller, 1975).

14 Note that Goldshneider and DeKeyser’s (2001) meta-analysis excluded Hakuta (1976), Makino (1979a), and Pak (1987), which studied Japanese and Korean learners, from the original list of 25 studies, because they did not meet their criteria.

References


Appendix A

Morpheme Studies on Japanese Learners of English

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant(s)</th>
<th>Method</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hakuta (1976)</td>
<td>A 5-year-old girl living in the United States (Uguisu)</td>
<td>Recordings of spontaneous utterances (longitudinal)</td>
<td>• The learner had more difficulty learning plural and articles.</td>
<td>• The L1 has both positive and negative influence.</td>
</tr>
<tr>
<td>Koike (1983)</td>
<td>Three children in the United States: a 10-year-old boy (Nobi), a 7-year-old boy (Jun), and a 5-year-old girl (Sachiko)</td>
<td>Recordings of spontaneous utterances (longitudinal)</td>
<td>• The acquisition order of the 5-year-old girl and the 10-year-old boy correlated significantly with that of Hakuta’s (1976) girl.</td>
<td>• Many errors in the use of morphemes were caused by transfer and overgeneralization.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• The acquisition order of the 7-year-old boy showed strong correlation with Makino (1979).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Possessive ‘s was acquired relatively early across all three children.</td>
<td></td>
</tr>
<tr>
<td>Sasaki (1987)</td>
<td>One 9-year-old girl (K) living in the United States, compared with four other children from Hakuta (1976) and Koike (1981)</td>
<td>Spontaneous utterances in naturalistic settings for 10 months (longitudinal)</td>
<td>• The correlation among four out of five children and the natural order is not significant.</td>
<td>• The L1 has an effect on the acquisition order.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Plural, auxiliary, article, and possessive deviate most from the natural order.</td>
<td>• There seems to be a universal trend for some of the morphemes (e.g., –ing, regular and irregular past).</td>
</tr>
</tbody>
</table>

(Continued)
### Appendix A

#### Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant(s)</th>
<th>Method</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Makino (1979a)</td>
<td>777 high school students in Japan</td>
<td>Paper-and-pencil test (cross-sectional)</td>
<td>Accuracy order shows significant correlations with that of Dulay and Burt (1973, 1974) and Bailey et al. (1974).</td>
<td>The amount of language instruction and exposure to English did not have an effect on the ordering, thus supporting the claim of a universal strategy employed by learners.</td>
</tr>
<tr>
<td>Nuibe (1986)</td>
<td>123 junior-high school students in Japan</td>
<td>Japanese-to-English written translation (cross-sectional/longitudinal, data collected four times)</td>
<td>Articles and plurals were acquired late.</td>
<td>The so-called “natural order” is probably an artifact of the Bilingual Syntax Measure.</td>
</tr>
<tr>
<td>Shirahata (1988)</td>
<td>31 high school students in Japan</td>
<td>Recordings of English utterances (cross-sectional)</td>
<td>Possessive and plural deviate from natural order.</td>
<td>Although a natural order seems to be suggested to some degree, L1 influence also seems to be at play.</td>
</tr>
<tr>
<td>Izumi &amp; Isahara (2004)</td>
<td>167 samples from beginners to advanced (no age specified)</td>
<td>NICT Japanese Learner English spoken Corpus (cross-sectional)</td>
<td>Possessive was ranked higher than the natural order, and articles are ranked lower.</td>
<td>L1 transfer is the main cause of deviation from the natural order.</td>
</tr>
</tbody>
</table>
### Appendix B

**Morpheme Studies on Korean Learners of English**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant(s)</th>
<th>Method</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pak (1987)</td>
<td>40 adults and 40 children&lt;br&gt;40 child Korean ESL students: aged 5–12, all born in Korea&lt;br&gt;40 adult Korean ESL students: aged 25–38, graduate students&lt;br&gt;Length of residence: 6 months–7 years</td>
<td>Bilingual Syntax Measure (cross-sectional)</td>
<td>• The acquisition orders of the children and adult groups showed significant correlation.&lt;br&gt;• Possessive is learned earlier than the natural order, and plural is learned later than the natural order.&lt;br&gt;• Although the children also learned through the medium of English, their acquisition order was different from English-speaking children.</td>
<td>• The L2 learning process is somewhat different from L1.</td>
</tr>
<tr>
<td>Shin &amp; Milroy (1999)</td>
<td>12 Korean children aged about 7, living in New York City</td>
<td>Recording of three activities in classroom: storytelling, math, and play (cross-sectional)</td>
<td>• The results are significantly correlated with Pak (1987).</td>
<td>• The acquisition order is affected by the learners’ L1 knowledge.</td>
</tr>
</tbody>
</table>
### Appendix C

**Morpheme Studies on Chinese Learners of English**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant(s)</th>
<th>Method</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dulay &amp; Burt (1974)</td>
<td>55 Chinese-speaking children and 60 Spanish-speaking children</td>
<td>Bilingual Syntax Measure (cross-sectional)</td>
<td>• Results of Spanish-speaking children and those of Chinese-speaking children were similar</td>
<td>• Strong support for the existence of universal child language strategies</td>
</tr>
</tbody>
</table>
| Mace-Matluck (1979) | 442 students from kindergarten through Grade 4, with different L1s, including Cantonese (175), Tagalog (49), Spanish (168), and Ilokano (30) | SEA-CAL Oral Proficiency Tests (cross-sectional) | • Not all L2 ranks correlated significantly with Brown's L1 rank order  
• All L2 ranks correlated significantly with each other, with some major differences (e.g., Cantonese-speaking children acquired possessive relatively more easily than Spanish-speaking children, and the other way round for the acquisition of plurals). | • The result may be due to the fact that L2 learners differ from L1 learners in terms of cognitive development and to the existence of the complex grammar system of the L1.  
• Other factors such as frequency of occurrence of structures and L1 transfer contribute to differences in the orders. |
# Appendix D

## Morpheme Studies on Spanish L1 Learners of English

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant(s)</th>
<th>Method</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dulay &amp; Burt (1973)</td>
<td>151 Spanish-speaking children in three groups differing with respect to the amount of English exposure</td>
<td>Bilingual Syntax Measure (cross-sectional)</td>
<td>• Similar results were obtained across the three groups.</td>
<td>Children’s innate ability to organize structure accounts for the acquisition of L2 syntax.</td>
</tr>
<tr>
<td>Dulay &amp; Burt (1974)</td>
<td>See Appendix C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bailey et al. (1974)</td>
<td>73 Spanish and non-Spanish-speaking adults; aged 17–55 years</td>
<td>Bilingual Syntax Measure (cross-sectional)</td>
<td>• Results agreed with those of Dulay &amp; Burt (1973)</td>
<td>Children and adults use common strategies in processing linguistic data.</td>
</tr>
<tr>
<td>Rosansky (1976)</td>
<td>6 untutored Spanish (2 children, 2 adolescents, 2 adults)</td>
<td>Spontaneous conversation (mostly cross-sectional, longitudinal only for one adolescent-Jorge)</td>
<td>• The BSM-based morpheme order did not correlate with the spontaneous-speech-based order for the same individual at the same point in time.</td>
<td>The validity of cross-sectional studies of SLA based on instrument-elicited morpheme data is questionable.</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Participant(s)</th>
<th>Method</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andersen (1978)</td>
<td>89 L1 Spanish students at the University of Puerto Rico</td>
<td>Written composition (cross-sectional)</td>
<td>• The order obtained correlates significantly with Krashen’s natural order.</td>
<td>• Frequency is an explanation of natural order.</td>
</tr>
<tr>
<td>Mace-Matluck (1979)</td>
<td>See Appendix C</td>
<td></td>
<td></td>
<td>• L1 transfer is also at play.</td>
</tr>
<tr>
<td>Pica (1983)</td>
<td>18 adult native Spanish speakers belonging to Instruction Only group, Naturalistic group, and Mixed group</td>
<td>Hour-long audiotaped conversations between learner and researcher (cross-sectional)</td>
<td>• The acquisition order of all three groups significantly correlated with the natural order</td>
<td>• Learners make use of the “natural ability” to acquire their L2</td>
</tr>
</tbody>
</table>
The linguistic landscape of Hong Kong after the change of sovereignty

Mee Ling Lai

Department of English, Hong Kong Institute of Education, 10 Lo ping Road, Hong Kong

Published online: 20 Jul 2012.

To cite this article: Mee Ling Lai (2013) The linguistic landscape of Hong Kong after the change of sovereignty, International Journal of Multilingualism, 10:3, 251-272, DOI: 10.1080/14790718.2012.708036

To link to this article: http://dx.doi.org/10.1080/14790718.2012.708036
The linguistic landscape of Hong Kong after the change of sovereignty

Mee Ling Lai*

Department of English, Hong Kong Institute of Education, 10 Lo ping Road, Hong Kong

(Received 11 February 2011; final version received 7 March 2012)

This study about the linguistic landscape of Hong Kong was conducted 12 years after the People's Republic of China (PRC) reclaimed sovereignty over the city. A total of 1160 visual signs displayed in open public spaces in four selected areas of Hong Kong were analysed. These sites extend from the very heart of the city to a small town near the border with the PRC. Through analysis of the assembled photographic evidence, it is able to establish baseline data that help answer three major questions namely: to what extent is Hong Kong a multilingual city; how prominent is the influence of the PRC after the change of sovereignty; and how far is the presence of minority groups represented in the linguistic landscape of Hong Kong.

Keywords: linguistic landscape; Hong Kong; China influence

Introduction

In recent years, the study of linguistic landscapes (LL) has attracted growing interest in the fields of sociolinguistics and applied linguistics. Studies of LL have been carried out in many major cities in the world, including those in Asia, such as Bangkok (Huebner, 2006), Tokyo (Backhaus, 2006), Taipei (Curtin, 2009) and Korea (Malinowski, 2010). In Hong Kong too, Jaworski and Yeung (2010) investigated signs displayed in public spaces, yet the focus of the study was set on the semiotic meanings of 260 addresses and names of residential properties. Since few comprehensive LL surveys have been attempted in Hong Kong, the present study therefore seeks to fill the gap; and more importantly, it endeavours to establish baseline data for the documentation and evaluation of the linguistic scenery of the city. Through this, it is expected that future changes in Hong Kong's sociolinguistic evolution can be captured, especially when the city is undergoing continuous social and cultural change following the transfer of sovereignty from Britain to China. Adopting the frameworks of previous publications (i.e. Gorter, 2006; Landry & Bourhis, 1997), this study examines all kinds of visual linguistic objects displayed in the public spaces of four chosen areas in the city of Hong Kong. A total of 1160 signs were analysed, these ranging from small handwritten stickers on lamp posts to the names of shops, buildings and places, large commercial billboards, traffic signs and public notices. Through such visual signs, the current roles and status of various languages as well as the power relationships between different ethnic groupings are examined.

*Email: mllai@ied.edu.hk

© 2013 Taylor & Francis
Sociolinguistic backgrounds of Hong Kong before and after the change of sovereignty

Hong Kong is a city situated in southern China, over 95% of its population is ethnic Chinese. In 1997, the city was reintegrated into the People’s Republic of China (PRC), putting an end to 155 years of British colonial rule. Under the principle of ‘One country; two systems’, it then became a Special Administrative Region (SAR) of China administered exclusively by Hong Kong residents. It also exercises a high degree of autonomy and enjoys executive, legislative and independent judicial power (Hong Kong Basic Law, 1997, Article 2).

Given the British legacy, Hong Kong possesses a long bilingual history of English and Chinese. Although the sociolinguistic situation remains largely the same after the political handover, changes have been taking place as a result of the increased influence from the PRC. In this section, therefore, the varying sociolinguistic and demographic backgrounds of Hong Kong before and after the change of sovereignty are briefly described and the research questions consequently formulated identified.

Roles and status of English

Throughout the history of Hong Kong, English has enjoyed the status of a prestigious language strongly associated with higher education, higher social status, intelligence and wealth (Lai, 2007). Before the change of sovereignty, English was the working language for the formal institutions of government, law, education and international business. From the 1960s until 1997, the language was used as a major medium of instruction (MoI) in 90% of the secondary schools in Hong Kong. Although its role as the MoI was largely replaced by the local language (i.e. Cantonese) after 1997, English continued to be a compulsory subject in all schools for all levels. Moreover, a pass in the language remains a requirement for entrance to all local universities. English therefore still functions as a gatekeeper to higher education and better life chances up till the present.

Although the influence of Britain has subsided somewhat since the transfer of sovereignty (Lai, 2012), English continues to enjoy high status as one of the two official languages of Hong Kong (Hong Kong Basic Law, 1997, Article 9) and as a symbol of Hong Kong’s new positioning as ‘Asia’s World City’. According to the results of the last census in 2006, although the percentage of population employing English as a ‘usual language’ declined from 3.1% in 1996 to 2.8% in 2006, that using it as ‘another language’ had increased from 34.9% to 41.9% during the same period of time (Hong Kong Census and Statistics Department, 2007a). For these reasons, it was anticipated that English would form a vital and highly visible component in the LL of Hong Kong SAR, functioning not as a last vestige of past colonial history, or solely as an international language of commerce and business, but more significantly as a vital marker of ‘two systems’ that differentiates Hong Kong from other cities of the PRC.

Biliteracy and trilingualism

In depicting the sociolinguistic profile of Hong Kong, it is necessary to consider the written and spoken forms of the Chinese language separately. Before the return of sovereignty from Britain to the PRC in 1997, both English and Chinese were the
official languages. While the meaning of ‘Chinese’ was less problematic in regard to writing, there being little regional variation across the country, it was never made clear in any government documentation which spoken variety ‘Chinese’ referred to. Since the overwhelming majority of the Hong Kong population was using Cantonese as a lingua franca, the term ‘Chinese’ was customarily taken to mean Cantonese as far as the spoken form was concerned. In brief, the situation before the change of sovereignty could be described as ‘biliteracy’ (standard written Chinese and English) and ‘bilingualism’2 (Cantonese and English), where ‘Cantonese’ is a spoken vernacular syntactically different from standard written Chinese and is largely unintelligible to speakers of other Chinese varieties.

The ambiguous interpretation of ‘Chinese’ continued to exist after 1997 since it was never stated explicitly in the Basic Law whether ‘Chinese’ refers to local Cantonese or Putonghua (the national pronunciation of Chinese, also known as Mandarin). To recognise the status of the national pronunciation, the policy of ‘Biliteracy and Trilingualism’ was formally introduced in the first policy address delivered by the first Chief Executive of the new Hong Kong Special Administrative Region (HKSAR) in 1997 (Tung, 1997). Under the new policy, the postcolonial generation of Hong Kong is expected to be able to write standard Chinese and English, and speak not only Cantonese and English, but also Putonghua:

Confidence and competence in the use of Chinese and English are essential if we are to maintain our competitive edge in the world. The Education Commission Report No.6 has already laid down a framework to achieve our goal for secondary school graduates to be proficient in writing English and Chinese and able to communicate confidently in Cantonese, English and Putonghua.

Appearance of Cantonese and Putonghua on signs

In addition to enjoying an enhanced status in education and government administration after the political handover, Putonghua has become the most commonly heard non-local language in the streets of Hong Kong. In fact, it is often more important for retail salespersons to be able to speak Putonghua rather than English since the PRC has become Hong Kong’s largest source of tourists, reaching a record number of 15 million visitor arrivals in 2007 as compared to 1.23 million from North America and 0.75 million from Australia and New Zealand (Hong Kong Tourism Board, 2008). As pointed out in Wikipedia (2011), the transliteration of English proper names based on Putonghua pronunciations has become more common in Hong Kong after the political handover, this trend being pioneered in the media:

Before 1997, Cantonese pronunciation was the basis for transliterating English proper names into Chinese. After the handover, however, the media in Hong Kong began to adopt already established transliterations based on Mandarin pronunciations in order to align with the mainland… In some cases, the Mandarin-based transliteration sounds far from the original when pronounced in Cantonese.

For this reason, it is expected that the romanisation of Chinese proper names based on Putonghua pronunciations might also become more salient in the LL of Hong Kong, as a result of the burgeoning PRC influence in the city. In addition, as Hong Kong is predominantly a Cantonese-speaking community, Cantonese texts with their unique grammar, syntactic structure and written colloquial form (often coined by
adding a mouth sign next to a standard Chinese character of similar pronunciation), were also expected to be common in the LL of Hong Kong, especially where the intention is to communicate affective meanings to the local Cantonese audience.

Traditional vs. simplified Chinese script

After the communists took over China in 1949, Hong Kong was secluded from the development of the PRC. For this reason, Hong Kong preserved the use of the traditional Chinese script even though simplified Chinese has been enforced throughout the PRC since 1950. The use of traditional Chinese, together with the use of Cantonese and English, bestow a distinctive identity upon the Hong Kong territory that contrasts with mainland realities, where people speak Putonghua and write simplified characters. Although simplified Chinese has never been granted any official status in Hong Kong and is not taught in schools, it was still expected that it might be evident in the city’s LL due to continuous integration with the PRC and the influx of visitors from the mainland.

Minority groups

In July 2009, the population of Hong Kong reached seven million (Ming Pao Daily News, 2009). Although the Chinese group remains dominant, the presence of other ethnic groups, such as Filipinos, Thais and Indians, ensures a noticeable degree of ethnic diversity and suggests additional areas of multilingualism. Despite this, it has yet to be ascertained whether or not these minority groups are fairly represented in the LL of the city.

Research method

As Landry and Bourhis (1997, p. 205) state:

> linguistic landscape is concerned with the issue of language in the written form in the public sphere. It is the language of public road signs, advertising billboards, street names, place names, commercial shop signs and public signs on government buildings combined to form the linguistic landscape of a given territory, region or urban agglomeration.

And, by further extension, the data samples used for the research of LL also include language(s) displayed on vehicles, T-shirts, magazines (Curtin, 2009) and even graffiti (Pennycook, 2009). A wide range of themes have also been explored either through the quantitative or qualitative approach or a mix of both, including ethnolinguistic vitality (Landry & Bourhis, 1997; Leeman & Modan, 2010), language diversity and language shift (Huebner, 2006), level of multilingualism (Backhaus, 2006), competition between languages and identities (Lanza & Woldemariam, 2009), use of imagery (Jaworski & Yeung, 2010) and second language acquisition (Cenoz & Gorter, 2008).

As the present study aims to establish a starting point for the documentation of the linguistic changes in Hong Kong, a quantitative approach was adopted to furnish an overview of the LL of the city. Following the framework established by Landry and Bourhis (1997), photographs of signs were collected from four selected areas of
the city, through which, answers were sought for the following questions in the social context of Hong Kong 12 years after the change of sovereignty:

(1) How dominant are Chinese and English in the LL of Hong Kong?
(2) How prominent is the linguistic influence of the PRC?
(3) How much is the presence of other minority groups represented?
(4) How does the LL vary from the heart of the city to the suburban area at the border?

Research procedures

Data were collected in four areas of the city of Hong Kong from July to August 2009. The four sample areas are situated at various points along a major Mass-transit railway (MTR) line running from south to north, from the heart of the city to the border of the PRC. The four chosen areas include the main business district, a tourist hub, a local shopping area and a residential town near the border. However, as there are 19 districts in the city, data collected from the four sample areas can only be treated as indicative of the linguistic outlook of Hong Kong, but not as an absolute linguistic composition of the territory.

The four areas are: (1) Central – situated at the southernmost end of the MTR, the political, financial and business heart of Hong Kong, home to the Government Secretariat and the prime location for international commercial enterprises and major banks; (2) Tsimshatsui – located 2.45 km north of Central, a major tourist hub for international travellers where many of the 5-star hotels and flagship stores selling luxurious brands can be found; (3) Mongkok – situated 1.85 km north of Tsimshatsui, a shopping and entertainment area known for local specialties, small shops, food stalls, inexpensive clothes, street markets and restaurants. The area attracts visitors from other areas of Hong Kong and from abroad; (4) Sheungshui – a small town located 22.82 km to the north of Mongkok, the last stop of the MTR in the north before entering the PRC border. Services and businesses in this vicinity used to cater largely to the local working-class community, but in recent years, the area has become a popular destination for visitors from mainland China, who cross the border in search of quality daily necessities, such as milk powder and shampoo, after product safety has become a public concern in China.

In order to maintain consistency, data were collected by the same researcher, who was instructed to identify one main street and one side street in each sample area. Main streets refer to large streets on a double-decker bus route. Side streets refer to smaller streets, which allow only pedestrians, minibuses or private cars. Main streets present high visibility data reflecting their significance as major arteries of the area. Side streets are also included in the sample to increase the possibility of capturing small idiosyncratic signs that better reveal personal and group identities. In order to ensure ample data collection potential, streets in each sample area were chosen according to two overriding criteria: firstly, the length of the sample streets in each area was to be comparable; and secondly, the streets had to be the site of vibrant commercial and communal activities. It should be noted, however, that as Hong Kong is a city full of multi-storey buildings, signs above the ground level were not included except for the prominently large billboards or banners that often adorn the external walls of buildings. This restriction allowed data collection to remain manageable.
Data cleaning procedures were carried out after signs had been collected. A sign was excluded if: (1) it was a duplication of another one in the same sample street (e.g. a second branch of the same chain shop displaying the same linguistic text e.g. 7-eleven). This avoids over-representation; (2) it was visually unclear or (3) it had no linguistic text or words of any sort. In sum, this study covered a total of eight streets in four areas and a total of 1160 sample signs were used for analysis (see breakdown in Appendix 1).

In accordance to the frameworks proposed by other researchers (e.g. Ben-Rafael, Shohamy, Amara, & Trumper-Hechi, 2006), the signs in this study can be largely divided into two types: (1) official and public and (2) private and business. The official and public signs include all government notices, traffic and street signs, announcements of public interests, advertisements from charity organisations, church notices and banners for council election. In this study, only 60 signs fall into this category, which constitutes 5.2% of the total. As Hong Kong is a highly commercialised city, the vast majority of the sample signs (94.8%) are of a private or business nature. These signs can also be designated as ‘bottom-up’, meaning that they are initiated by individual social actors, aiming to project an image or to attract consumer attention.

Signs collected were then analysed according to the number and prominence of languages appearing on each sign, through that, the LL of Hong Kong was depicted.

**Problems in classifying languages**

Classifying languages was not an unproblematic task, especially in the case of Japanese signs since the written form contains a number of Kanji (Chinese characters) which overlap either with traditional or simplified Chinese. In such cases, Chinese characters were counted as Japanese Kanji if they were used: (1) with Hiragana or Katakana (i.e. Japanese phonetic symbols); (2) in Japanese names or terms; or (3) to form part of Japanese syntax. However, romanisation of Japanese names or words, for example ‘sushi’, were read as English.

Regarding words of European outlook, except some clear content items such as the French words ‘cafe’, ‘bonjour’, most of the others were famous brand names of different European origins, such as Lamborghini (Italian), Louis Vuitton (French) and Porsche (German), while some were local Hong Kong brands (e.g. Baleno, Giordano) which intended to foster an European image. As French words and brand names occurred with particular frequency, ‘French’ was accorded an independent category while other more disparate European words were grouped under the heading ‘other European languages’.

**Research results**

**Number of languages on signs**

Backhaus (2006) analysed 2321 multilingual signs collected in Tokyo. In his study, the term ‘multilingual’ was defined as signs written in any language in addition to, or instead of Japanese, thus including those written in only one language. Unlike Backhaus, this study adopts a conventional definition of the terms ‘monolingual’, ‘bilingual’ and ‘multilingual’, to refer to any signs written in one language, two languages or three (or more) languages regardless of what languages are displayed.
As shown in Table 1, more than half of the total sample signs (53.4%) are monolingual, of which, 27.7% are written in a non-Chinese language. The number of bilingual signs is slightly lower (45.1%). Multilingual signs which display three or more languages are not common, accounting for only 1.5% of the total sample. Bilingual and multilingual signs include not only texts written parallel to each other, but may also involve code-mixing at the sentential level. Had Backhaus’s definition of ‘multilingual signs’ been applied to this study, the percentage of multilingual signs would have soared to 74.3%, a far higher figure than the 19% reported in the Tokyo sample.

After identifying the number of languages on signs, we shall in the following section examine which languages contribute to the LL of Hong Kong, and where appropriate how prominent each language is in relation to other languages on the same sign.

### Languages visible in the linguistic landscape of Hong Kong

#### Monolingual signs

As shown in Table 2, the vast majority of the signs are written in either Chinese or English, which account for 569 signs out of a total of 620 monolingual signs. Other monolingual signs are written in Japanese (N = 5), French (18) and other European languages, including Italian, German and Spanish (N = 28). Most of the signs written in European languages are luxurious brand names while Japanese often appears in relation to restaurants and sushi bars.

#### Bilingual and multilingual signs

There are a total of 540 bilingual and multilingual signs, which make up 46.6% of the sample. As shown in Table 3, the large majority of the bilingual signs are written in Chinese and English (N = 505), others combine Japanese, French or Nepalese (N = 18) with either Chinese or English, which serves as an invariant component in

<table>
<thead>
<tr>
<th>Type of signs</th>
<th>Number of signs</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td>620</td>
<td>53.4</td>
</tr>
<tr>
<td>Bilingual</td>
<td>523</td>
<td>45.1</td>
</tr>
<tr>
<td>Multilingual</td>
<td>17</td>
<td>1.5</td>
</tr>
<tr>
<td>Total</td>
<td>1160</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Languages</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>298</td>
<td>25.7</td>
</tr>
<tr>
<td>English</td>
<td>271</td>
<td>23.3</td>
</tr>
<tr>
<td>Japanese</td>
<td>5</td>
<td>0.4</td>
</tr>
<tr>
<td>French</td>
<td>18</td>
<td>1.6</td>
</tr>
<tr>
<td>European languages</td>
<td>28</td>
<td>2.4</td>
</tr>
<tr>
<td>Total</td>
<td>620</td>
<td>53.4</td>
</tr>
</tbody>
</table>
the texts. Similarly, Chinese and English also appear as the constant components in the multilingual signs ($N = 17$) combining with Japanese, French or Korean.

Prominence of languages in multilingual signs

As noted by Scollon and Scollon (2003), there must be a system of preference when multiple codes are competing for space on the same sign. In such cases, the preferred code is usually put on top, on the left, or in the centre position, and the marginalised code is on the bottom, on the right, or on the margins. This study employs the same system in assessing the relative importance of languages on multilingual signs. Taking HSBC as an example (Figure 1), the logo of the bank is placed in the middle while the linguistic versions of the bank name are written in the same font size on either side of the logo. Yet, English may be identified as the preferred code since it is placed on the right.

Apart from the position of the code, font size is also an important indicator of language prominence. Taking Figure 2 as an example, although the Chinese name of the building is placed on top, the English version of the name (LUK HOI TONG

Table 3. Languages appear on bilingual and multilingual signs.

<table>
<thead>
<tr>
<th>Languages</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese + English</td>
<td>505</td>
<td>43.5</td>
</tr>
<tr>
<td>Chinese + other (i.e. Japanese, French)</td>
<td>7</td>
<td>0.6</td>
</tr>
<tr>
<td>English + other (i.e. Japanese, French or Nepalese)</td>
<td>11</td>
<td>1.0</td>
</tr>
<tr>
<td>Bilingual Chinese–English + other (i.e. Japanese, Korean or French)</td>
<td>17</td>
<td>1.5</td>
</tr>
<tr>
<td>Total</td>
<td>540</td>
<td>46.6</td>
</tr>
</tbody>
</table>

Figure 1. English appears to be the preferred code as it is placed on the right of the logo.
BUILDING) is written in capital letters and has a noticeably larger font size, thus commanding more attention from the audience. In cases like this, English is considered to be the more prominent language on the sign.

Even when font size is added to the criteria, the task of identifying language prominence is not without its complications. For example, in Figure 3, although the English name is placed at the top, and the Chinese text to the right, both can be seen as occupying the margins to satisfy the demands of graphic design. This makes use of a characteristic of the English language which runs in a horizontal vector while the Chinese text runs in a vertical vector. The two codes are of comparable font size and are both placed in the margins, it is therefore hard to judge which one is preferred or

Figure 2. The English version of the building is in bigger font size and is thus visually more prominent.

Figure 3. English and Chinese are equally prominent in different vector.
is more distinctive. In such cases, the two languages are considered to be of equal prominence on the sign.

As summarised in Table 4, Chinese is the most prominent language in the LL of the Hong Kong city (N = 368), followed by English (N = 133). This is perhaps unsurprising as Chinese is the first language of the large majority of the local population. Signs showing equal prominence of different languages are visible (N = 28), where words of the same font size are placed in comparable positions.

**Comparison of the four sample areas**

As mentioned at the beginning of this study, the four sample areas chosen were situated along a major train line. Following the line, from south to north, from the heart of the city towards the PRC border, changes in the LL also emerge. In Table 5, consistent patterns of change can be identified the further north the location. Nearer the heart of the city (i.e. Central), there are more English-only and more bilingual signs. Foreign languages are also more visible nearer the city centre as most of these signs display European brand names. By contrast, Chinese predominates in the local areas (i.e. Mongkok and Sheungshui). Simplified Chinese also appears with more noticeable frequency in areas further north, indicating stronger mainland influence closer to the border. The exception to all these is Tsimshatsui, where monolingual English signs (44.1%) and foreign brand names (7.6%) are particularly prominent, reflecting high levels of foreign tourist density.

Table 4. Prominence of languages on bilingual and multilingual signs.

<table>
<thead>
<tr>
<th>Most prominent language</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>368</td>
<td>31.7</td>
</tr>
<tr>
<td>English</td>
<td>133</td>
<td>11.5</td>
</tr>
<tr>
<td>Equal</td>
<td>28</td>
<td>2.4</td>
</tr>
<tr>
<td>Japanese</td>
<td>10</td>
<td>0.9</td>
</tr>
<tr>
<td>French</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>Total out of 1160 sample signs</td>
<td>540</td>
<td>46.6</td>
</tr>
</tbody>
</table>

Table 5. Comparison of the four districts.

<table>
<thead>
<tr>
<th></th>
<th>Central (%)</th>
<th>Tsimshatsui (%)</th>
<th>Mongkok (%)</th>
<th>Sheungshui (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chin only</td>
<td>43.1</td>
<td>60.7</td>
<td>52.8</td>
<td>57.6</td>
</tr>
<tr>
<td>Eng only</td>
<td>13.4</td>
<td>9</td>
<td>28.6</td>
<td>51.7</td>
</tr>
<tr>
<td>Other foreign languages</td>
<td>25.2</td>
<td>44.1</td>
<td>21</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>7.6</td>
<td>3.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Bi/Multilingual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chin prominent</td>
<td>56.9</td>
<td>39.3</td>
<td>47.2</td>
<td>42.4</td>
</tr>
<tr>
<td>Eng prominent</td>
<td>29.3</td>
<td>20</td>
<td>26.9</td>
<td>30</td>
</tr>
<tr>
<td>Appearance of other language</td>
<td>10</td>
<td>7.6</td>
<td>5.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Simplified Chinese</td>
<td>3.3</td>
<td>1.7</td>
<td>4.8</td>
<td>1.7</td>
</tr>
</tbody>
</table>
Summary
To sum up, the present sample presents a three-tiered picture of the LL of Hong Kong. Chinese and English are the two most dominant components, together constituting the foreground of the city. Although far weaker both in terms of quantity and prominence, Japanese, French and other European languages are also easily discernable, embellishing and enriching the cityscape. Other languages, such as Nepalese and Korean, can, however, be noticed only through careful observation. More evidence of foreign languages is found towards the heart of the city, especially in the tourist areas as these languages are usually associated with famous international brand names. As our focus shifts to the areas in further north, Chinese is found to be more dominant and simplified Chinese more frequently employed.

Discussion
Roles and status of English and Chinese in the linguistic landscape
As revealed in the data, Chinese and English are the two most dominant languages in the LL of Hong Kong. Except for the 51 monolingual signs written entirely in another foreign language, Chinese or/and English are present in all the other 1109 signs in the sample. This clearly demonstrates the prevalent bilingual culture of Hong Kong, which has remained largely the same 12 years after the political handover. Before 1997, both Chinese and English were the official languages of Hong Kong. Although it was stated in the Official Languages Ordinance that the two languages ‘possess equal status’ and ‘enjoy equality of use’ (Dickson & Cumming, 1996), English was in practice serving higher functions being the main language of the government, law courts, international businesses and the MoI for secondary and university education whereas Chinese was more often used as the language of solidarity and the means of everyday communication among the general public. After the change of sovereignty, although the official status of the two languages remained the same, the vitality of Chinese has increased considerably with the change of governance from the British to the local Hong Kong Chinese. In addition to this, the huge expansion of trade with China, the increased appeal of local culture and the implementation of the mother tongue education policy in secondary schools since 1998, have all contributed to the enhanced vitality of Chinese in Hong Kong. Indeed, this is accurately reflected in the LL of Hong Kong. As found in this study, Chinese is highly dominant in the cityscape especially where visual signs in the street cater mainly for the local ethnic Chinese people.

Despite the fact that English has become less commonly used in some domains (e.g. government and education) after the political change, it remains highly visible in all parts of the city, largely because of Hong Kong’s long-established bilingual history. As LL is about permanent signage in public spaces, one would not expect drastic or rapid changes to these signs without strong top-down interference, and this, however, has not occurred after the political handover. Throughout the history of Hong Kong, English has played an important function as one means of communication. During the period of British governance, the language served mainly the interest of the colonising group. Yet, as Hong Kong started to develop into an international financial centre from the 1970s onwards, English was used more widely as a lingua franca for international and inter-group communications.
Apart from their communicative function, English signs remain salient in the city of Hong Kong also because of their symbolic meanings; first as an icon of ‘two systems’, second being a symbol of internationalisation, westernisation, modernisation, success and attractiveness (Brock, 1991; Cenoz & Gorter, 2008; Lai, 2007). This may explain why some brands insist on using only a foreign-name for their business, which helps project a westernised and international image, while details of their products are mostly written in Chinese (e.g. Figure 4), which ensures accessibility to the local community. As pointed out by Brock (1991), the function of English on commercial signs is often ornamental rather than communicative. In this regard, Jaworski and Yeung (2010, p. 162) have a similar observation about the symbolic functions of English and other prestigious languages in Hong Kong:

It would be inappropriate to interpret the presence of bilingual signs in our data as an index of a predominantly (Chinese and English) bilingual population. It is probably more appropriate to suggest that ... English and a few other high status languages are frequently used to bestow on various goods and services symbolic capital associated with globalization, internationalization, sophistication, reliability etc.

In many discussions, Chinese is often described as a language of solidarity and English a language of power (Pennington, 1998; Pierson, 1994). However, So (1998) argues that a language of solidarity can at the same time be a language of power for its persuasiveness and its great appeal to the vast majority of the local community. If the concept of ‘power’ can be broadened in line with So’s argument, both Chinese and English should be considered powerful languages but in different senses, the former being empowered through its accessibility and appeal to the local community and the latter by its prestige as a symbol of internationalisation. Since both Chinese and English are indispensable to Hong Kong’s core identity, it is therefore to be

![Figure 4. Only the brand name is in English while all details are written in Chinese.](image)
expected that the roles of the two languages in the cityscape of Hong Kong would remain by and large the same in the foreseeable future.

**Visibility of minority groups in the linguistic landscape**

Although Hong Kong is largely a homogenous Chinese community (95%), the remaining 5% of the population displays considerable ethnic diversity. The largest minority groups are (in order): Filipino (1.6%), Indonesian (1.3%), Caucasian (0.5%), Indian and Pakistani (0.5%), Japanese (0.2%) and Thai (0.2%; Hong Kong Census and statistics Department, 2007b). However, except for Japanese, languages of other minority groups do not appear to be salient in the present study. This is perhaps unsurprising as no attempt was made during the data collection phase to magnify the linguistic representation of these groups. Unless special attention is paid to hotspots where such groups may congregate or reside, their linguistic presence cannot be easily noticed in the city. However, this does not hold true in the case of Japanese, French and other European languages. These languages stand out in all sample areas even though their respective population sizes are small in number. Such visibility is largely due to the prestigious meanings they carry as symbols of style, fame, quality and luxury. As mentioned earlier, shops and restaurants adopting these names may not necessarily reflect the origins of the owners. In fact, many such shops are owned by local Hong Kong people with no connection to the target language countries. For example, chain stores such as ‘Café de Central’ and ‘Café de Carol’, are local fast food shops, which neither serve French cuisine nor have any connection with France. Another chain store named ‘Bonjour’ is also a local cosmetics shop, which adopts a French word as a name for its association with renowned French cosmetic products. Similarly, a well-known snack shop chain called ‘Aji Ichiban’ (優的良品) has little connection with Japan except for having Japanese cakes as one of their many selling items. This act of foreign-name branding is used not to demonstrate solidarity with the target language groups, but to acquire some of the world-renowned prestige that these languages carry.

In brief, minority groups are not represented in proportion to their demographic power in the city. The foreign languages that do stand out do so primarily because of their prestige and status. This confirms again that language choice on signage serves more to convey symbolic meaning than to fulfil informational functions (Landry & Bourhis, 1997).

From the communicative point of view, Chinese and English are deemed sufficient to cover all needs of Hong Kong society under a dichotomised view of the world, with Chinese being the language for the in-group and English for the out-group (Lai, 2009). This may explain why the two official languages are so largely dominant in the cityscape of Hong Kong. Other prestige languages are in evidence but the full scope of Hong Kong’s ethnolinguistic diversity is absent.

**Visibility of Cantonese in the linguistic landscape**

Although Cantonese is only a vernacular with no formal codification, it has been the dominant language spoken in Hong Kong for everyday communication, local popular culture, basic education and local administration. After the change of sovereignty, the vitality of Cantonese has never declined even when facing the challenge of Putonghua. Although it is considered a local language which does not
help much in gaining higher opportunities in career and academic advancement, its role is irreplaceable as a language of solidarity and a marker of the Hong Kong identity (Lai, 2005). In recent years, written Cantonese has gained increasing popularity in newspapers and magazines especially for informal topics (e.g. entertainment news). Despite this, written forms of Cantonese do not appear to be highly visible in the LL of Hong Kong as revealed by the present study. Out of all 827 sample signs which contain solely or partly Chinese, only eight were found to display colloquial Cantonese words or syntax (e.g. Figures 5 and 6). In seven of them, Cantonese is used in slogans for affective functions, aiming to establish solidarity with the local customers.

Figure 5. Cantonese is used by a bank as a catchy slogan in their advertisement: ‘One day, when cash comes to your door, how would you like to use it?’.

Figure 6. ‘HEA’ is transliteration of a colloquial Cantonese word which means strolling around in a leisurely and aimless manner.
The fact that Cantonese does not appear often on permanent signage shows that it is not well-accepted as a written language, especially for formal and referential functions. Although Cantonese has been playing a dominant role in verbal communication and enjoys covert prestige, its written form is considered improper and sub-standard, and should therefore be largely restricted to informal and affective functions only. As observed by Bauer (2004, p. 40), ‘if anything has to be taken seriously, it should not or would not be written in Cantonese’.

Visibility of the PRC

As the influence of the PRC has grown tremendously in Hong Kong after the change of sovereignty, a higher proportion of linguistic evidence of mainland influence was expected, for instance, the use of simplified Chinese and the romanisation of names from Putonghua pronunciations. However, out of a total of 827 signs which contain Chinese characters, simplified Chinese was found in only 21 cases. In such instances, simplified Chinese is sometimes used to indicate the mainland origin of the owner (e.g. Figure 7). In many more other cases, it is used for pragmatic functions appealing directly to the mainland customers (e.g. Figure 8).

Near the change of sovereignty in 1997, Scollon and Scollon (2003) attempted to track the flow of simplified writing from China to Hong Kong. They then concluded that there was almost no entrance of simplified writing into Hong Kong. Twelve years have elapsed since the political handover; although the present sample reveals an observable increase of simplified Chinese in the LL of Hong Kong, its impact

![Image](https://example.com/image.png)

Figure 7. The Chinese version of ‘China Construction Bank’ is written in simplified Chinese characters to show its mainland origin.
does not yet appear to be highly significant. From the pragmatic perspective, this may be partly because simplified and traditional Chinese are to a large extent mutually intelligible, and there is no absolute need to use simplified Chinese for communicative purposes. From the symbolic perspective, traditional Chinese is valued as a significant indicator of Hong Kong’s unique identity. In addition, as pointed out by Scollon and Scollon (2003), traditional Chinese also symbolises progression and a modern Chinese lifestyle outside the PRC, to which, people on the mainland might also aspire.

Apart from simplified Chinese, linguistic representation of the PRC through Putonghua transliteration is also uncommon. Out of 132 signs which involve the romanisation of Chinese names, only five of them are based on Putonghua pronunciations while 127 use Cantonese pronunciations. All the signs displaying simplified Chinese or Putonghua pronunciations are ‘bottom-up’ in nature, reflecting the influence of mainland culture which is mainly driven by the economic and demographic power of its people rather than by political top-down rulings from the PRC government.

Language use on official signs
Throughout the history of Hong Kong, there has been no clear official policy governing language use on public signage. During the colonial period, English was the sole language used for all official matters. Only in 1974, after immense public pressure, was Chinese granted the status of an official language (Dickson & Cumming, 1996). Although it was claimed at the time that both English and Chinese possessed ‘equal status’, English was the primary working language used in
all official documents. Whenever discrepancies arose between the Chinese and English versions, English had to be taken as the official and correct version. Regarding official signage, there was never any announced policy governing the positions and font sizes of the two official languages. However, as may be observed from some of the public signs issued during the colonial period, it was apparent that English played the role of the primary official language. Taking an old street sign (Figure 9) as an example, although the font sizes of both the Chinese and English scripts are the same, English appears to be the preferred language as it was placed on top of the Chinese version.

After the change of sovereignty, both English and Chinese remained the official languages of the Hong Kong SAR, yet the role of English seemed to have changed from primary to subsidiary, as stated in Hong Kong Basic Law (1997), Article 9:

In addition to the Chinese language, English may also be used as an official language by the executive authorities, legislature and judiciary of the Hong Kong Special Administrative Region

In fact, the dominant role of Chinese seems to be salient as shown in many public signs and banners put up by the government after the change of sovereignty. Examples can be found in Figures 10 and 11 where Chinese is often the first language displayed on top of the English version and in bigger font size.

However, there are also examples that refute the pattern. Taking Figure 12 as an example, although the sign was issued after the change of sovereignty, English obviously plays a primary role on the sign as it was placed in the top position, in larger font size and capital letters. The inconsistent design of official signs reveals a clear absence of standard policy on official signage in HKSAR.

Similarly, there is no government intervention in Hong Kong regarding the use of simplified writing and whether the romanisation of Chinese names should be based on Cantonese or Putonghua pronunciations. Such practice is quite different from that of the Taiwan government (another major region using traditional Chinese), which, in order to resist cultural invasion from the PRC, forbids simplified writing in

Figure 9. An old street sign showing the English version on top and the Chinese script at the bottom.
Figure 10. A banner put up by the Labour Department to promote employment services.

Figure 11. A message posted by the Public Transport Department to invite suggestions from the public.
official documents and websites and openly discourages its use in the private sector (Chan, 2011).

**Conclusion**

As mentioned at the beginning of this study, the aim of this study is to establish a starting point for the documentation of the LL of Hong Kong, when the city is undergoing continuous social and cultural transition in the second decade after the change of sovereignty. As evidenced in the data, Hong Kong displays a highly bilingual profile of Chinese and English. Out of 1160 signs, Chinese or/and English are found in almost all sample signs except for 51 monolingual ones which were written solely in a foreign language. Although English is still highly powerful in the LL of Hong Kong as a marker of internationalisation and the local identity, Chinese has become more dominant as it is the first written language of the large majority of the population, and its predominant role being reinforced through the change of sovereignty from Britain to China.

There is also some degree of multilingualism with other foreign languages appearing on 3.1% of the signs, and this is mainly contributed by European brand names and Japanese restaurants for the prestigious connotations that these languages carry. However, ethnic diversity does not seem to have contributed much to the multilingual profile of Hong Kong, as the minority languages are hardly visible to the mainstream society in the busy areas of the city. As regards the influence of PRC, although its presence is becoming more salient in the form of simplified Chinese and the romanisation of names from Putonghua pronunciations, its impact on the linguistic landscape of Hong Kong does not yet appear to be significant. Nonetheless, this trend is expected to grow in the coming years as Hong Kong is gradually assimilated into the PRC.
Notes
1. Due to strong demand for the language and great social pressure, English gradually resumed its role as an MoI in secondary education since 2010 when the government relaxed the mandatory mother tongue education policy to allow a more extensive use of English for teaching and learning in secondary schools.

2. An anonymous reviewer suggested that ‘oral bilingualism’ should be used instead of ‘bilingualism’. However, the reviewer’s suggestion was not adopted in this study because the policy of ‘Biliteracy and Trilingualism’ is a well-known term used in the Hong Kong context that differentiates literacy from oracy.

3. As suggested by Scollon and Scollon (2003), the difference between traditional and simplified Chinese characters can be said to resemble that between the English word ‘through’ and its short form ‘thru’.

4. Most of the simplified Chinese characters in use today are the result of the works moderated by the government of the PRC in the 1950s and 1960s. The purpose of the simplification movement is to reduce illiteracy. (Wikipedia: http://en.wikipedia.org/wiki/Simplified_Chinese#Mainland_China as retrieved on 5 January 2012)

5. Towards the end of the British governance, Cantonese started to replace English as the main language used in Legislative Council meetings. After 1997, English was also replaced by Cantonese as the MoI in secondary schools under the mandatory mother tongue education policy. The annual policy address of the Chief Executive is also delivered in Cantonese instead of English.

6. According to an email reply (dated 22 September 2011) from the Highways Department of Hong Kong, the present sign plate was issued in 2003. It was claimed in the reply that there was no background information about the graphic design of the sign plate, including why English was written in capital letters before Chinese.

References


Pennington, M.C. (1998). Perspectives on language in Hong Kong at century’s end. In M.C. Pennington (Ed.), Language in Hong Kong at Century’s End (pp. 3–40). Hong Kong: Hong Kong University Press.


### Appendix 1

**Total number of signs used in each sample area**

<table>
<thead>
<tr>
<th>Areas</th>
<th>No. of signs used in main streets</th>
<th>No. of signs used in side streets</th>
<th>Total no. of signs used for analysis</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>193</td>
<td>97</td>
<td>290</td>
<td>25</td>
</tr>
<tr>
<td>Tsimshatsui</td>
<td>198</td>
<td>92</td>
<td>290</td>
<td>25</td>
</tr>
<tr>
<td>Mongkok</td>
<td>195</td>
<td>95</td>
<td>290</td>
<td>25</td>
</tr>
<tr>
<td>Sheungshui</td>
<td>194</td>
<td>96</td>
<td>290</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>780</strong></td>
<td><strong>380</strong></td>
<td><strong>1160</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
Tracking language attitudes in postcolonial Hong Kong: An interplay of localization, mainlandization, and internationalization

Lai Mee Ling

Abstract

Four years after the sovereignty of Hong Kong was returned from Britain to China, a survey was conducted in 2001 to examine the attitudes of students toward Cantonese (the vernacular language), English (the colonizer's and international language), and Putonghua (the new ruler's language) in the early postcolonial era of Hong Kong. Eight years after, as Hong Kong moved into the second decade after the political handover, which has been characterized by an intense interplay of localization, mainlandization, and internationalization, a follow-up study was launched in 2009 to trace the changes of language attitudes over the past years. This paper reports on a comparison between the results of the two surveys, which shows 'Pragmatic trilingualism' as a future trend for the younger generation of Hong Kong. Having experienced the postcolonial changes for twelve years, informants of the 2009 study also showed significantly more positive attitudes toward Putonghua than their counterparts in 2001, as the language starts to take root in Hong Kong society along with the huge economic and demographic power of China.

Keywords: language attitudes, Hong Kong

1. Introduction

After 155 years of British colonial rule, the sovereignty of Hong Kong was returned to the People’s Republic of China (PRC) in 1997. Since then, Hong Kong is no longer a British colony but a special administrative region (SAR) of China. Under the principle of ‘One country, two systems’, Hong Kong enjoys a high degree of autonomy (Basic Law, Article 21), and the socialist system and policies are not to be practised in the city (Basic Law, Article 5). Four years after the political handover, I conducted a survey to examine the attitudes of 1,048 Secondary
Four (S4) students toward the three spoken languages promoted by the new Hong Kong SAR (HKSAR) government, namely Cantonese (the local vernacular), English (language of the colonizer and an international language), and Putonghua (the national language of China, also known as Mandarin). The respondents were from 28 different secondary schools and all of them were the first cohort of students under the mandatory mother tongue education policy and had undergone the significant political, social, and educational changes in the early postcolonial era of Hong Kong.

Years had elapsed after the last study, and Hong Kong had become an SAR of China for more than 10 years. During this period of time, further significant changes had taken place in the socioeconomic and educational domains, which might, in consequence, impact the linguistic ecology of Hong Kong. Li (2000) argues that Hong Kong people are not passive victims of linguistic imperialism (cf. Phillipson 1992) but active agents of pragmatism in their choice of languages. As Hong Kong moved into the second decade of its postcolonial era, which is characterized by an intense interplay of localization, mainlandization, and internationalization, a follow-up study was launched in 2009 to find out how students’ language attitudes are different from those of 2001. Although this study starts at a micro-level investigating how students perceive a language in specific social contexts, findings thus gathered help to 1) trace the changes of language attitudes in the postcolonial history of Hong Kong, and 2) reveal the shift of roles and status of Cantonese, Putonghua, and English in Hong Kong society after the change of sovereignty.

2. Background of the present study

2.1 The socioeconomic and educational changes in Hong Kong from 1997 to 2009

Hong Kong is a city in southern China; 95 percent of its population is ethnic Chinese. Since the political handover in 1997, Hong Kong has undergone great challenges in the social and economic domains at different times. Immediately after the political handover, the financial crisis that swept through Asia had severely impacted Hong Kong, resulting in large-scale bankruptcies, cost-cutting in all enterprises, and a soaring unemployment rate, from 2.7 percent in 1997 to 6.3 percent in 2001 when the first survey was conducted. Hardly before the slumping economy became stable, Hong Kong was once again struck by SARS, the lethal epidemic, in 2003, leading the city further into the worst economic downturn it had ever experienced in history. Just when the
city started to recover from the blows in 2006 and 2007, the world-wide financial tsunami also took effect in Hong Kong at the end of 2008, and that threw the city back to bankruptcies, cost-cutting, and unemployment. According to the government figures, the unemployment rate from April to June of 2009 was 5.4 percent when the present survey was conducted (Hong Kong Information Services Department 2009).

Apart from the socioeconomic changes, Hong Kong also experienced significant changes in the education domain after the change of sovereignty, among these, the medium of instruction policy was the one that had the greatest impact on students’ language attitudes. Two months after the handover in July 1997, as a gesture of decolonization, the new HKSAR government announced the mandatory mother tongue education policy for junior secondary education. Before that, more than 90 percent of the secondary schools in Hong Kong used English as the medium of instruction (EMI). As a result of the policy change, more than 70 percent of the secondary schools were required to adopt Chinese (i.e., standard written Chinese and spoken Cantonese) as the medium of instruction (CMI) while only about 20 percent were able to retain the EMI status. Although this policy was mandatory only for junior secondary education from Secondary 1–3, it was met with great resistance from schools and parents for fear that students would lose their competitive edge with less exposure to English during their critical years of learning. After enforcing the policy for more than ten years, the government finally decided to relax it to allow more use of English in the classroom. Starting from the academic year of 2010–2011, the label of CMI or EMI on schools should begin to fade out, and all schools should have the liberty to use either Chinese or English as the instructional language according to the needs and abilities of their students.

As regards Putonghua, it was first introduced into the formal school curriculum in 1997 as a core subject for Primary 1 to Secondary 3. Upon relaxing the mother tongue education policy, the status of Putonghua in education remains the same. Although it is only an elective subject in the senior secondary curriculum, an increasing number of schools are using it to replace Cantonese as the instructional language for the Chinese subjects (Ming Pao Daily News, 4/4/2009) as it is commonly believed that the language will become more important as the power of China increases in the world.

2.2 The sociolinguistic setting of Hong Kong from 1997 to 2009

The significant changes in the socioeconomic and educational aspects discussed above had great impacts on Hong Kong society, including the linguistic ecology. Before 1997, English (language of the colonizer) and
Cantonese (the vernacular language) formed a diglossic situation (Fishman 1967) in which the two languages were used in different domains and for different functions. English was a prestigious and a working language for the formal institutions of government, law, education, and international business while Cantonese was used by the vast majority of the Hong Kong population as their usual language in family and other informal daily-life settings. However, toward the end of British governance, Cantonese started to take over some of the high functions of English under the localization movement. After the change of sovereignty, Cantonese became even more vital as a neutral language that symbolizes ‘decolonization’ without arousing sentiments of ‘recolonization’ by the PRC (Pennington 1998).

Although Cantonese had become more dominant in use, English continued to enjoy high status as an official language of Hong Kong (Basic Law, Article 9), and a symbol of Hong Kong’s new positioning as ‘Asia’s World City’. Under the urge for internationalization, English remains indispensable after the political change as a means to maintain Hong Kong’s international status and connection with the outside world.

While Cantonese and English remained vital, the bilingual setting of Hong Kong was transformed into a trilingual one when Putonghua was formally introduced into the sociolinguistic scene of Hong Kong after 1997, through the policy of ‘Biliteracy and trilingualism’, which refers to the use of English and Chinese as the written languages; and English, Cantonese, and Putonghua as the spoken varieties. The increased importance of Putonghua, however, was not solely a result of the political change, but that of the process of ‘mainlandization’, which refers to the huge economic and cultural influences of China on Hong Kong. With special policies (e.g., CEPA and Individual Travel Scheme) that China introduced to save the fragile Hong Kong economy from the continuous financial and epidemic blows since 1997, Hong Kong’s reliance on the mainland has become increasingly great, and so is the demand for its language as a means of business and communication. In recent years, Putonghua has become the most frequently heard non-local language in Hong Kong streets, and it is often more important for retail salespersons to be able to speak Putonghua rather than English since mainland China has become the largest source of tourists for Hong Kong. In public places (e.g., buses and trains), announcements are made not only in Cantonese and English, but also Putonghua.

2.3 Informants of the present research

It was under an intense interplay among the forces of localization, internationalization, and mainlandization that the respondents of this
study (aged 15–17) were brought up in the postcolonial era of Hong Kong. This group of students was in S4 when this research was conducted in April–June 2009. While the respondents of the last survey belonged to the first cohort under the mandatory mother tongue education policy, those of the present study were members of the second last group before the policy was relaxed. As Putonghua was introduced into the formal school curriculum as a core subject in 1997, the 2001 cohort had learnt the language from Secondary 1 to 3, whereas the 2009 group had learnt it for a total of nine years from Primary 1 to Secondary 3. Unlike the 2001 cohort, who had experienced the postcolonial changes only for four years, students of the present group had undergone the significant political, socioeconomic, and educational changes of Hong Kong after a total of twelve years. For the differences dealt with above, respondents of the present study were expected to show:

1) more positive attitudes toward Cantonese as Hong Kong had become highly localized after the change of sovereignty, and the majority of the 2009 respondents had used Cantonese as the medium of instruction since Primary;

2) a higher level of acceptance toward Putonghua as the power of China had greatly increased and all of the respondents had learnt Putonghua as a core subject for a total of nine years;

3) less positive attitudes toward English despite its importance as a symbol of internationalization, since the British influence had subsided and English had lost some of its high functions in Hong Kong SAR.

3. Previous studies of language attitudes in Hong Kong

Among the studies on language attitudes of Hong Kong people, many were conducted before the change of sovereignty (e.g., Pierson et al. 1980; Pennington & Yue 1994; Littlewood & Liu 1996; Axler et al. 1998), and their main focus was on English and Chinese (Cantonese) with little attention paid to Putonghua, which was at that time not much of a social concern. Hyland (1997) reported on a questionnaire study with 900 university students shortly before the political handover. Although English, Cantonese, and Putonghua were mentioned in the questionnaire, the focus of attention was on English. The findings showed that although English was recognized for its instrumental value, it was not significant in familial contexts or as a status marker.

To extend the scope of survey from schools to the general public, Boyle and his researchers conducted quick interviews conducted at a
busy railway station in Hong Kong (Boyle 2000). Target subjects were young working adults aged 20–40. A total of 1,093 respondents answered short questions about Cantonese, English, and Putonghua. The results showed that 52 percent of the respondents believed that Cantonese would be more important than Putonghua in Hong Kong after 1997. If they had to choose between English and Putonghua, 75 percent would wish to be excellent in English. 80 percent of the respondents thought English would best benefit the economic development of Hong Kong. 75 percent said English would be the most useful language for their career prospects, 12 percent said Cantonese, and 13 percent said Putonghua. Such results revealed the supreme status and value of English on the one hand, while Cantonese and Putonghua were competing with each other on an equal footing as two varieties of Chinese on the other.

In an attempt to gauge the language attitudes of secondary students after the change of sovereignty and to put a balanced focus on Cantonese, English, and Putonghua, I conducted a questionnaire survey in 2001 with students from 28 secondary schools. All respondents were in their fourth year of secondary education, and they were the first cohort of students undertaking mother tongue education. A questionnaire was devised on a four-point Likert scale, which examined the respondents’ attitudes toward the three official spoken languages of Hong Kong SAR. A total of 1,048 questionnaires were analyzed and the results showed that the respondents felt the most affectively inclined to vernacular Cantonese and perceived English as the language of the highest instrumental value and social status. Despite the political change, Putonghua was rated lowest in both the affective and cognitive domains (Lai 2005). Although students were found showing an accommodating attitude to Putonghua, they did not perceive its roles and status as highly as those of Cantonese and English.

4. Research method of the present study

4.1 Research Instrument

In order to facilitate a valid comparison with the 2001 study, a similar research method was adopted for the present study. A questionnaire on a four-point Likert scale was devised, and the same questions were asked. Similar to the 2001 survey, the question items were categorized into six factors according to Gardner & Lambert’s (1972) sociocultural model, viz. (i) integrative orientation toward Cantonese, (ii) integrative orientation toward English, (iii) integrative orientation toward Puton-
According to Gardner (1985), ‘instrumental orientation’ refers to a positive inclination toward a language for pragmatic reasons, such as obtaining a job or higher education opportunity; and ‘integrative orientation’ refers to a favorable inclination toward a language in order to become a valued member of a given community. ‘Integrativeness’ thus implies not only an interest in a language, but also an open attitude toward another cultural group; in the extreme, it suggests emotional identification with the community of the target language (Gardner 2001).

4.2 Research procedures

In order to facilitate a direct comparison, letters were sent to the same 28 secondary schools that had participated in the 2001 research. Permission was asked to re-run the survey with one class (about 40 students) of S4 students in each school. However, positive replies were received only from 11 schools. In order to enlarge the sample size, invitations were extended to other schools in Hong Kong. In the end, questionnaires were collected from an additional 25 schools, making a total of 36 schools in the sample. However, in order to increase the representativeness of the sample, some questionnaires were excluded so as to align the ratio between female and male respondents, and that between EMI and CMI schools with those of the larger population. According to the latest census figures in 2006, the percentage of the male population of Hong Kong is 47.7 and that of the female population is 52.3 (Hong Kong Census and Statistics Department 2007). As for CMI and EMI schools, the ratio is around 3 : 1.

As a result, the analysis was carried out with a total of 1,145 questionnaires collected from 33 schools. As shown in table 1 below, the ratios on ‘Sex’ and ‘MoI’ in the 2009 sample are very close to those shown in the 2006 government census:

4.3 Comparability of the 2001 and 2009 surveys

Although the ratios on ‘Sex’ and ‘MoI’ in the 2009 sample, as shown in table 1, those of the 2001 cohort were not close to those in reality. However, this could not have skewed the overall attitude patterns because female and EMI students were found to be more positive to the nonnative languages (i.e., English and Putonghua) in the last survey (Lai 2004, 2007). As there was a lower percentage of female students
but a higher percentage of EMI students in the 2001 sample, it was believed that the effects on the overall pattern had been balanced out. Nevertheless, as the composition of the 2001 and 2009 samples is not exactly the same, results of the comparison should be read with caution, and only differences with proven significance will be addressed in the analysis.

Apart from ‘Sex’ and ‘MoI’, respondents’ ‘Place of birth’ and their ‘Home languages’ may also impact their language attitudes. As shown in table 1, the large majority of the respondents in the 2009 survey were born in Hong Kong (74.3 percent) using Cantonese as their home language (92.3 percent) while a handful of them use Putonghua (0.9 percent) and English (1.1 percent) at home. This aligns well with the government figures in the 2006 census, which shows that 90.8 percent of the Hong Kong population speaks Cantonese as their usual language while 0.9 percent use Putonghua and 2.8 percent use English (Hong Kong Census and Statistics Department 2007). Since the composition of respondents in these categories is very similar in the two surveys, it is feasible to state that any significant differences arising from the statistical analysis should not be caused by these factors.

4.4 Data analysis

A total of 1,145 questionnaires were analyzed for the 2009 study. Means and standard deviations (SD) were calculated for each item in the questionnaire. Any means greater than 2.5 indicate a positive inclination, and large SD values show a great disparity among the respondents. In addition, composite means were calculated for each of the six
factors in the integrative and instrumental domains. In order to align the negatively worded statements with the positive ones, means of the negative statements were reversed before the composite means were calculated within each factor. Cronbach’s reliability test was applied to ensure the internal reliability within each factor, any alpha value ($\alpha$) greater than 0.7 indicates high reliability while that which falls between 0.5 to 0.7 indicates moderate reliability (Fitz-Gibbon & Morris 1987). As found in the present study, the alpha values range from 0.58 to 0.78, showing a satisfactory internal reliability within each of the six factors (cf. tables 3–8 in section 5.2).

Results of the 2009 survey were compared with those of 2001 to reveal the changes of language attitudes in the past few years. Effect size values$^6$ were calculated to show whether the differences between the two surveys are large. According to Fitz-Gibbon & Morris (1987), any effect size value lower than 0.2 indicates a very small and insignificant difference; that around 0.5 refers to a moderate difference, and that higher than 0.8 indicates a large difference. Following this scale, this paper attends only to moderate to large differences with effect size values equal to or higher than 0.4.

5. Results of the comparison between the 2001 and 2009 surveys

In this section, an overall comparison between the 2001 and 2009 surveys will be reported on the six factors in the integrative and instrumental domains. After that, the means, standard deviations, and effect size values will be shown on the question items of each factor.

5.1 An overall comparison

There are two main findings regarding the change of students’ language attitudes from 2001 to 2009: 1) the overall attitude pattern toward the three spoken languages is the same; 2) attitudes toward Putonghua are significantly more positive in 2009 than 2001.

As revealed from the composite means of the six factors in table 2 below, the attitude pattern of the 2009 respondents is the same as that of the 2001 survey in both the integrative and instrumental domains. Similar to the results gathered eight years ago, students show the greatest affection toward Cantonese (mother tongue of the majority) and English second. As for instrumental values, English is rated the highest and Cantonese second. In both domains, Putonghua ranks the lowest.

While the respondents’ attitudes toward Cantonese and English are highly positive, all with composite means above ‘3’, those related to Putonghua are below. Yet the composite means on the Putonghua-re-
lated factors are well above the mid-point (i.e., 2.5) in the 2009 survey, showing a clearly positive inclination toward the language in both domains.

The largest change is found in Factor 3 ‘Integrative orientation toward Putonghua’. While the composite mean of the 2001 survey was negatively skewed (i.e., 2.47), that of the 2009 survey is clearly positive (i.e., 2.79). As shown through the effect size values, Putonghua has gained the most significant positive change in both the integrative (\(\Delta = 0.6\)) and instrumental domains (\(\Delta = 0.53\)) in the 2009 survey:

Table 2. Means of the six factors.

<table>
<thead>
<tr>
<th>Factors</th>
<th>2001 Mean</th>
<th>2001 SD</th>
<th>2009 Mean</th>
<th>2009 SD</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Integrative orientation toward Cantonese</td>
<td>3.43</td>
<td>.40</td>
<td>3.42</td>
<td>.43</td>
<td>0.02</td>
</tr>
<tr>
<td>2. Integrative orientation toward English</td>
<td>3.05</td>
<td>.45</td>
<td>3.21</td>
<td>.46</td>
<td>0.35</td>
</tr>
<tr>
<td>3. Integrative orientation toward Putonghua</td>
<td>2.47</td>
<td>.55</td>
<td>2.79</td>
<td>.51</td>
<td>0.60 #</td>
</tr>
<tr>
<td>4. Instrumental orientation toward Cantonese</td>
<td>3.19</td>
<td>.48</td>
<td>3.14</td>
<td>.57</td>
<td>0.09</td>
</tr>
<tr>
<td>5. Instrumental orientation toward English</td>
<td>3.51</td>
<td>.34</td>
<td>3.5</td>
<td>.41</td>
<td>0.03</td>
</tr>
<tr>
<td>6. Instrumental orientation toward Putonghua</td>
<td>2.66</td>
<td>.38</td>
<td>2.9</td>
<td>.52</td>
<td>0.53 #</td>
</tr>
</tbody>
</table>

\(\uparrow\): group with a higher mean; # moderate to large differences

5.2 Comparison on individual factors

After an overall comparison, this section focuses on the individual items of each factor to reveal details about the change of attitudes.

5.2.1 Factor 1: Integrative orientation toward Cantonese

As shown in table 3 below, the composite mean for the ‘integrative orientation toward Cantonese’ in the 2009 survey (3.42) is very close to that of 2001 (3.43). Although the means in nearly all items have declined in the 2009 survey, the differences are insignificant, as is evident from the small effect size values, and the overall decrease in this factor is made up by a significant increase in item 3.12b ‘As a Hongkonger, I should be able to speak fluent Cantonese’ (\(\Delta = 0.77\)). This may reflect the role of Cantonese as a dominant language and an important symbol of the Hong Kong identity:
Table 3. Integrative orientation toward Cantonese.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 (α = 0.67)</th>
<th>2009 (α = 0.68)</th>
<th>Eff. size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>2.3</td>
<td>I like Cantonese because it is my mother tongue.</td>
<td>3.31</td>
<td>.72</td>
<td>3.15</td>
</tr>
<tr>
<td>2.5</td>
<td>Cantonese is the language which best represents Hong Kong.</td>
<td>3.45</td>
<td>.64</td>
<td>3.36</td>
</tr>
<tr>
<td>2.10</td>
<td>Cantonese should be replaced by Putonghua since it is only a dialect with little value.</td>
<td>1.57</td>
<td>.69</td>
<td>1.59</td>
</tr>
<tr>
<td>3.3b</td>
<td>I like Cantonese.</td>
<td>3.63</td>
<td>.53</td>
<td>3.45</td>
</tr>
<tr>
<td>3.9b</td>
<td>I like Cantonese speakers.</td>
<td>3.59</td>
<td>.54</td>
<td>3.44</td>
</tr>
<tr>
<td>3.12b</td>
<td>As a Hongkonger, I should be able to speak fluent Cantonese.</td>
<td>3.18</td>
<td>.71</td>
<td>3.71</td>
</tr>
</tbody>
</table>

Composite mean of Factor 1 3.43 .40 3.42 .43 .02

# moderate to large differences

5.2.2 Factor 2: Integrative orientation toward English

Although the difference between the composite means in this parameter is not noticeably significant (Δ = 0.35), students’ integrative orientation toward English has generally become more positive in the 2009 survey. Such a positive change is mainly contributed by items 3.12a and 3.13a, in which students in the present study take English a lot more strongly as a symbol of Hong Kong identity (Δ = 0.82), and also that of a higher group who are usually well-educated, intelligent, and well-off (Δ = 0.4):

Table 4. Integrative orientation toward English.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 (α = 0.67)</th>
<th>2009 (α = 0.58)</th>
<th>Eff. size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>2.6</td>
<td>I would like to speak fluent English because it makes me feel modern and westernized.</td>
<td>3.11</td>
<td>.81</td>
<td>3.06</td>
</tr>
<tr>
<td>2.8</td>
<td>A person who speaks fluent English is usually arrogant, snobbish and show-off.</td>
<td>1.88</td>
<td>.70</td>
<td>1.76</td>
</tr>
<tr>
<td>3.3a</td>
<td>I like English.</td>
<td>3.18</td>
<td>.70</td>
<td>3.07</td>
</tr>
<tr>
<td>3.9a</td>
<td>I like English speakers.</td>
<td>3.31</td>
<td>.63</td>
<td>3.33</td>
</tr>
<tr>
<td>3.12a</td>
<td>As a Hongkonger, I should be able to speak fluent English.</td>
<td>2.90</td>
<td>.75</td>
<td>3.50</td>
</tr>
<tr>
<td>3.13a</td>
<td>A person who speaks fluent English is usually educated, intelligent and well-off.</td>
<td>2.71</td>
<td>.90</td>
<td>3.08</td>
</tr>
</tbody>
</table>

Composite mean of Factor 2 3.05  .45 3.21  .46 .35

# moderate to large differences
5.2.3 Factor 3: Integrative orientation toward Putonghua

As shown in table 5 below, the difference between the 2001 and 2009 surveys is moderately significant, as indicated by an effect size value of 0.6. This is also the largest difference found in the six factors, showing a rather significant positive change in students’ integrative orientation toward Putonghua. Apart from this, a consistent increase is also found in all items in this factor, showing a higher level of acceptance toward Putonghua from the affective perspective. The greatest change in attitude appears in item 3.12c. Unlike the 2001 respondents, who clearly disagreed that Hongkongers should speak fluent Putonghua (mean = 2.26), the 2009 group agrees rather strongly with the statement (mean = 3.13), suggesting that Putonghua has become part of Hong Kong identity after the introduction of the policy of ‘Biliteracy and trilingualism’:

### Table 5. Integrative orientation toward Putonghua.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 (α = 0.75)</th>
<th>2009 (α = 0.67)</th>
<th>Eff. Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>2.4</td>
<td>Putonghua should be more widely used in Hong Kong so that Hong Kong will quickly integrate with the PRC</td>
<td>2.32</td>
<td>.87</td>
<td>2.39</td>
</tr>
<tr>
<td>2.19</td>
<td>I’m afraid that if I speak fluent Putonghua, others will think I am a new immigrant from the mainland.</td>
<td>2.19</td>
<td>.83</td>
<td>2.00</td>
</tr>
<tr>
<td>3.3c</td>
<td>I like Putonghua.</td>
<td>2.40</td>
<td>.89</td>
<td>2.79</td>
</tr>
<tr>
<td>3.9c</td>
<td>I like Putonghua speakers.</td>
<td>2.69</td>
<td>.84</td>
<td>2.93</td>
</tr>
<tr>
<td>3.12c</td>
<td>As a Hongkonger, I should be able to speak fluent Putonghua.</td>
<td>2.26</td>
<td>.83</td>
<td>3.13</td>
</tr>
<tr>
<td>3.13c</td>
<td>A person who speaks fluent Putonghua is usually educated, intelligent and well-off.</td>
<td>2.14</td>
<td>.70</td>
<td>2.48</td>
</tr>
<tr>
<td>Composite mean of Factor 3</td>
<td>2.47</td>
<td>.55</td>
<td>2.79</td>
<td>.51</td>
</tr>
</tbody>
</table>

# moderate to large differences

5.2.4 Factor 4: Instrumental orientation toward Cantonese

As shown in table 6 below, the effect size values in all items are very small, showing no notable changes in students’ instrumental orientation toward Cantonese. However, slight declines are found in 3 out of the 4 items in this factor. Although such a result is rather surprising, it seems understandable as the government had just announced the relaxation of the mandatory mother tongue education policy at the time of this research, suggesting that Cantonese is not the right language to increase the competitive edge of Hong Kong students:
Table 6. Instrumental orientation toward Cantonese.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 (α = 0.69)</th>
<th>2009 (α = 0.73)</th>
<th>Eff. size</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4b</td>
<td>Cantonese will help me much in getting better opportunities for further studies</td>
<td>2.99 .71</td>
<td>2.89 .77</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5b</td>
<td>Cantonese will help me much in better career opportunities in the 21st Century</td>
<td>3.04 .71</td>
<td>3.08 .77</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6b</td>
<td>Cantonese is highly regarded in Hong Kong society.</td>
<td>3.08 .67</td>
<td>3.03 .76</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.7b</td>
<td>I wish to master a high proficiency of Cantonese.</td>
<td>3.67 .59</td>
<td>3.57 .76</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Composite mean of Factor 4</td>
<td>3.19 .48</td>
<td>3.14 .57</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2.5 Factor 5: Instrumental orientation toward English

Similar to Cantonese, the difference between the 2001 and 2009 surveys in students’ instrumental orientation toward English is also very small. Students’ evaluation of English remains highly positive in all items. Throughout the years after the change of sovereignty, English has been perceived as an important language in Hong Kong (item 2.9), a highly useful means for better academic and career opportunities (items 3.4a & 3.5a), and a crucial factor that helps to increase the competitive edge of Hong Kong in the world (item 2.12):

Table 7. Instrumental orientation toward English.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 (α = 0.6)</th>
<th>2009 (α = 0.68)</th>
<th>Eff. size</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9</td>
<td>English is less important in Hong Kong after the change of sovereignty.</td>
<td>1.99 .84</td>
<td>1.92 .82</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>The use of English is one of the crucial factors which has contributed to the success of Hong Kong's prosperity and development today.</td>
<td>3.03 .74</td>
<td>3.06 .73</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.13</td>
<td>To increase the competitive edge of Hong Kong, the English standard of Hong Kong people must be enhanced.</td>
<td>3.35 .71</td>
<td>3.49 .63</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.4a</td>
<td>English will help me much in getting better opportunities for further studies.</td>
<td>3.88 .40</td>
<td>3.83 .92</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5a</td>
<td>English will help me much in getting better career opportunities in the 21st Century.</td>
<td>3.88 .39</td>
<td>3.83 .61</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6a</td>
<td>English is highly regarded in Hong Kong society.</td>
<td>3.85 .40</td>
<td>3.77 .64</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.7a</td>
<td>I wish to master a high proficiency of English.</td>
<td>3.58 .64</td>
<td>3.45 .79</td>
<td>.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Composite mean of Factor 5</td>
<td>3.51 .34</td>
<td>3.50 .41</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2.6 Factor 6: Instrumental orientation toward Putonghua

As shown in table 8, the overall difference between the 2001 and 2009 survey in this parameter is moderately large as indicated by an effect size value of 0.53. This shows that the overall instrumental value of Putonghua has been rated significantly higher by students of 2009. Although Putonghua still ranks the lowest after English and Cantonese, a consistent increase in positive response is evident in all items in this factor. The most notable change appears in item 3.4c (effect size = 0.43), which shows that students tend to believe more strongly that Putonghua can help them gain better academic opportunities in the future:

Table 8: Instrumental orientation toward Putonghua.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 Mean ± SD</th>
<th>2009 Mean ± SD</th>
<th>Eff. size</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.14</td>
<td>If Putonghua is widely used in Hong Kong, Hong Kong will become more prosperous.</td>
<td>2.51 ± .82</td>
<td>2.71 ± .79</td>
<td>.25</td>
</tr>
<tr>
<td>2.15</td>
<td>Putonghua is an important language in Hong Kong.</td>
<td>2.45 ± .79</td>
<td>2.73 ± .74</td>
<td>.37</td>
</tr>
<tr>
<td>2.16</td>
<td>The importance and status of Putonghua will soon be higher than that of English in Hong Kong</td>
<td>2.03 ± .71</td>
<td>2.24 ± .77</td>
<td>.28</td>
</tr>
<tr>
<td>3.4c</td>
<td>Putonghua will help me much in getting better opportunities for further studies.</td>
<td>2.74 ± .90</td>
<td>3.11 ± .85</td>
<td>.43 #</td>
</tr>
<tr>
<td>3.5c</td>
<td>Putonghua will help me much in getting better career opportunities in the 21st Century.</td>
<td>3.16 ± .91</td>
<td>3.43 ± .79</td>
<td>.32</td>
</tr>
<tr>
<td>3.6c</td>
<td>Putonghua is highly regarded in Hong Kong society.</td>
<td>2.69 ± .84</td>
<td>2.89 ± .79</td>
<td>.25</td>
</tr>
<tr>
<td>3.7c</td>
<td>I wish to master a high proficiency of Putonghua.</td>
<td>2.93 ± .86</td>
<td>3.19 ± .79</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>Composite mean of Factor 6</td>
<td>2.66 ± .58</td>
<td>2.90 ± .52</td>
<td>.53 #</td>
</tr>
</tbody>
</table>

# Moderate to large differences

5.2.7 Comparing Cantonese and Putonghua from the instrumental perspective

From the research results, the superior position of English in the instrumental perspective seems unarguable, whereas there is room for an alternative view regarding that of Cantonese and Putonghua. As shown in table 9 below, if only the means of the four statements related to Cantonese in Factor 4 (i.e., items 3.4 to 3.7) are compared with those of Putonghua in Factor 6, one can see that the composite mean for Putonghua (i.e., 3.16) is in fact higher than that of Cantonese in 2009.
Table 9. Comparison between Factor 4 and 6 on four key statements.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>This language will help me much in getting better opportunities for further studies.</td>
<td>2.99</td>
<td>2.74</td>
<td>2.89</td>
<td>3.11</td>
</tr>
<tr>
<td>3.5</td>
<td>This language will help me much in getting better career opportunities in the 21st Century.</td>
<td>3.04</td>
<td>3.16</td>
<td>3.08</td>
<td>3.43</td>
</tr>
<tr>
<td>3.6</td>
<td>This language is highly regarded in Hong Kong society.</td>
<td>3.08</td>
<td>2.69</td>
<td>3.03</td>
<td>2.89</td>
</tr>
<tr>
<td>3.7</td>
<td>I wish to master a high proficiency of this language.</td>
<td>3.67</td>
<td>2.93</td>
<td>3.57</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>Composite mean on 4 key statements</td>
<td>3.19</td>
<td>2.88</td>
<td>3.14</td>
<td>3.16</td>
</tr>
</tbody>
</table>

CAN = Cantonese; PTH = Putonghua

(i.e., 3.14) while it was obviously lower in 2001 (PTH: 2.88 vs CAN: 3.19). This shows that Putonghua has made up a lot of ground in the instrumental domain in the past eight years and is comparable to the position of Cantonese although still significantly lower than that of English.

6. Summary of findings

6.1 Putonghua ranked last in the two surveys

Through the comparison between the 2001 and 2009 surveys, it was found that respondents of the latter display highly similar language attitude patterns as those of 2001. In the integrative domain, respondents favor Cantonese (mother tongue of the majority) most, English second. In the instrumental domain, the positions of Cantonese and English are reversed, with English being rated most highly for its instrumental values and social status, then Cantonese second. In both surveys, Putonghua ranks last in both the integrative and instrumental domains. Despite this, the attitudes of the 2009 respondents are significantly more positive toward Putonghua than the 2001 group, showing that the language is much better accepted in Hong Kong society from the affective perspective and is evaluated a lot more highly for its instrumental values and social status than before.

6.2 Most significant differences appeared for Putonghua

The two surveys are compared using a total of 36 question items, and notable differences are found in seven of them with effect size values
above 0.4. As shown in table 10 below, four out of the seven items displaying significant attitude changes are related to Putonghua, while two are about English and only one about Cantonese. The greatest differences were found in items 3.12a, b & c, indicating that students in 2009 think much more strongly that a Hongkonger should be able to speak Cantonese, English, and Putonghua. This illustrates the success of the government’s policy of ‘Biliteracy and trilingualism’. Although language attitudes may not necessarily correlate with language behaviors (Baker 1992; Eagly & Chaiken 1993), this shows good potential for Hong Kong in the development of a higher level of trilingualism in the future.

Apart from the above, the 2009 students also tend to take English more strongly as a symbol of education, intelligence, and wealth than before. Throughout the history of Hong Kong, English has played the role of a prestigious language (Pierson 1998). Despite the fact that Hong Kong has entered the second decade after its sovereignty was returned to China, the superior status of English remains unchanged, and it has become even more prominent as a symbol of education, intelligence, and wealth. On the other hand, the rating on a corresponding statement about Putonghua (item 3.13c) is still negatively skewed although it has already become significantly higher in 2009 (i.e., mean = 2.48) than 2001 (i.e., mean = 2.14). As indicated in the results, the 2009 respondents tend to like Putonghua more (mean of item 3.3c = 2.79), and they also tend to believe much more strongly that Putonghua can help them acquire better opportunities for further studies (mean of item 3.4c = 3.11):

Table 10. Significant differences between the 2001 and 2009 surveys.

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>2001 (α = 0.69)</th>
<th>2009 (α = 0.73)</th>
<th>Eff. size (Δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean   SD</td>
<td>Mean   SD</td>
<td></td>
</tr>
<tr>
<td>3.12a</td>
<td>As a Hongkonger, I should be able to speak fluent English.</td>
<td>2.90   .75</td>
<td>3.50   .71</td>
<td>.82</td>
</tr>
<tr>
<td>3.12b</td>
<td>As a Hongkonger, I should be able to speak fluent Cantonese.</td>
<td>3.18   .71</td>
<td>3.71   .68</td>
<td>.77</td>
</tr>
<tr>
<td>3.12c</td>
<td>As a Hongkonger, I should be able to speak fluent Putonghua.</td>
<td>2.26   .83</td>
<td>3.13   .78</td>
<td>1.09</td>
</tr>
<tr>
<td>3.13a</td>
<td>A person who speaks fluent English is usually educated, intelligent and well-off.</td>
<td>2.71   .90</td>
<td>3.08   .95</td>
<td>.40</td>
</tr>
<tr>
<td>3.13c</td>
<td>A person who speaks fluent Putonghua is usually educated, intelligent and well-off.</td>
<td>2.14   .70</td>
<td>2.48   .83</td>
<td>.44</td>
</tr>
<tr>
<td>3.3c</td>
<td>I like Putonghua.</td>
<td>2.40   .89</td>
<td>2.79   .86</td>
<td>.45</td>
</tr>
<tr>
<td>3.4c</td>
<td>Putonghua will help me much in getting better opportunities for further studies.</td>
<td>2.74   .90</td>
<td>3.11   .85</td>
<td>.43</td>
</tr>
</tbody>
</table>
6.3 Prediction about Putonghua supported

As mentioned in section 2.3, there are three predictions regarding students’ language attitudes in 2009 in comparison with the 2001 survey: 1) students will show stronger emotional attachment to Cantonese; 2) students will display a more positive attitude toward Putonghua; and 3) the attitude of students toward English will be less positive. As shown from the results of this study, only the prediction about Putonghua is supported while those about Cantonese and English are not. Despite eight more years’ of localization, students’ attitudes toward Cantonese remain more or less the same. On the other hand, although the British influence has subsided and the use of English in society has decreased, students in 2009 not only perceive it as a superior language for its supreme instrumental values, but also feel affectively more attached to it.

Although both surveys targeted only S4 students and their language attitudes and therefore cannot be overgeneralized, there is nevertheless a clear indication that their response, to a large extent, better reflects the collective psychology of society, since students are often less diplomatic than adults in expressing opinions. In the following section, I shall try to relate the students’ response to the social context of Hong Kong and explore what has contributed to their language attitudes as they are in 2009.

7. Discussions

7.1 Vitality of Cantonese – the empowering and undermining factors

Pierson (1998) predicted that Hong Kong would reach a triglossic situation of two high languages and one low language in a single community. Putonghua will become the language of politics and administration; English the language of technology, commerce, and finance; and Cantonese the language of the family and intimacy. Twelve years after the political handover, Pierson’s predictions seem to have been proved only partly correct. After the change of sovereignty, Cantonese continues to thrive in its unique status as a marker of Hong Kong identity and also a politically neutral language, which helps to erase the British colonial marks while at the same time staving off the fear of political imposition from Communist China. As stated in the Basic Law, Hong Kong is to be governed by local Hong Kong permanent residents and the lifestyle of people is to remain unchanged for 50 years (Basic Law, Article 5). Under the movement of localization, Cantonese has become the main language for politics and administration instead of Putonghua, especially when Hong Kong is moving toward a greater degree of self-
governance and democracy. In Hong Kong, Cantonese is the language that a candidate must know to show solidarity and win votes in the District Board or Legislative Council elections. With increasing pressure for a higher degree of transparency and accountability, heads of nearly all government departments and large enterprises would need good competence in the local language in order to gain support from the general public. Up till now, Cantonese has been not only a language used at home for day-by-day purposes, but also in the legislative council, the government, and the legal system. Having been educated under the mandatory mother tongue education policy, young people in this generation are more used to using Cantonese as the language at school and at work.

Apart from the formal institutions, Cantonese has also dominated the local mass media and the entertainment business in the past few decades (Sun 2004). With the advancement of technology, the vernacular is used even more often than before when communication between individuals can be carried out efficiently through mobile phones, MSN, and web cameras in the mother tongue. Information provided by Wikipedia (2009) about languages of Hong Kong may well summarize the broadened role of Cantonese in the postcolonial era:

The majority of the population in Hong Kong speaks Cantonese, a Chinese spoken variant originating from Guangdong province. It is a main variety used in education, broadcasting, government administration, legislature and judiciary as well as in daily social communication.

It is the enhanced role of Cantonese and its widened use in the Hong Kong community that has helped to formulate a positive attitude of the students toward the vernacular language in 2009. However, students’ attitudes have not become much more positive toward Cantonese than their counterparts in 2001. One of the possible reasons is that even though students are aware of the usefulness of Cantonese in society, it is still a regional dialect after all that has little national or international status. Government propaganda on TV often emphasizes the importance of English and Putonghua, but never Cantonese. Although the language is essential within the Hong Kong community, it does not empower students for upward and outward mobility, nor will anyone praise a student for being able to speak fluently in their mother tongue. Hong Kong people did not use to think highly of their mother tongue (Bauer 2000), mainly because it is only a basic skill that everyone takes for granted. Cantonese facilitates convenience and solidarity yet it cannot help to make one stand out from the crowd and increase one’s
competitive edge. Such an ideology was clearly manifested when the government announced its relaxation of the mandatory mother tongue education policy in 2009, suggesting that it was a mistake to use Cantonese as the medium of instruction for junior secondary education. The lack of overt prestige is perhaps a main reason to explain why students in 2009 do not appear to be more positive toward the local language, especially when they are the second last cohort of students under the mandatory mother tongue education policy, and are yet victims of the ‘mistake’!

7.2 Status of English unchallenged

When describing the rising position of Chinese, Sun (2004: 135–136) proclaims the death of English as the sole language of importance in Hong Kong:

When writing letters decades ago, the students were prone to use English, their sole educated language; but today more and more people are using PC, writing emails and all that, in Chinese … Formerly, in the job market of Hong Kong, the higher and more lucrative the positions were, the better the command of English was required. Today, though English competence is still frequently a sine qua non, Chinese is often more than just an advantage, because of the enormous market in China has changed everything. Enterprising business executives are learning Putonghua, and far-sighted young lawyers taking remedial lessons in Chinese law. The demise of English as the sole language of importance has arrived.

In fact, English did meet with great challenges from Cantonese and Putonghua in the postcolonial era of Hong Kong as a consequence of the movements of localization and mainlandization. As early as the late 1980s when Hong Kong began to prepare for the political transition, Cantonese had started to replace English in certain high functions. After 1997, English even lost its role as the medium of instruction for junior secondary education under the decolonization movement. About a decade ago, when PCs and Internet were mainly operated in English, the language was considered indispensable for modern technology and communication. However, as these technologies have become more sophisticated, most computer functions can now be performed in one’s mother tongue. According to Graddol (2006: 44), the percentage of browsers using English on the internet had dropped from 51.3 percent in 2000 to 32 percent in 2005. In Hong Kong, while many people from the older generation can only read English instructions
and type English words on computers, nearly everyone in the younger generation can send SMS messages and surf on the internet through Chinese (their L1). Although English is still an important language for international business and finance, there is no doubt that the English arena has become smaller in Hong Kong. Such a decrease in the use of English is also noted in Wikipedia (2009):

For most of the population who are ethnically Chinese, English is only a foreign language acquired from school education … On average, it is rare for a Hongkonger of Chinese ethnicity to achieve a fluent command of English, and Chinese is required for most daily communication purposes outside districts frequented by tourists … Code-switchings have become rarer alongside the decrease of use of English since the transfer of sovereignty in 1997.

The change in the linguistic ecology led me to believe that students in 2009 would feel less positive toward English since they were brought up in a Chinese-boom environment. However, the results of this study do not offer evidence of support for this assumption. On the contrary, students in the present study showed a stronger integrative orientation toward English, and the language is still rated most highly for its instrumental importance and social status, which remains unchallenged by Cantonese and Putonghua. The result of this study appears to be surprising especially when the majority of the present respondents (i.e., 74.5 percent as compared to 53.8 percent of the 2001 survey) were CMI students who had undertaken mother tongue education for 9 years. What, then, had lent English this power?

During the British governance, Hong Kong had developed into an international financial center. It is this international advantage that has enabled Hong Kong to stand out as a valuable and unique place for China. It is also this international advantage that upholds Hong Kong pride and enables Hong Kong to remain competitive when facing challenges from neighboring cities. It is widely believed that if Hong Kong has to remain prosperous and affluent, the city must be further internationalized. The less westernized and internationalized Hong Kong becomes, the more advantages it will lose to the other cities in Asia. The most direct tool to maintain internationalization and westernization is therefore ‘English’. This is perhaps the reason why the public felt so disturbed when the standard of students in English was found to have declined. Even after the mother tongue education policy has been in place for 11 years, the public is unwilling to change the role of English as the gatekeeper to higher education; for fear that Hong Kong graduates and pillars of future society will become unable to participate in
international affairs if less English is used in universities. When Professor Tsang Wing-kwong, a scholar from the Chinese University of Hong Kong, released his research findings in 2008, proclaiming that the English standard of CMI students was lower than that of their EMI counterparts and their success rate in entering universities was only half of that of the EMI students, instead of questioning whether higher education should also turn CMI to accommodate the majority, the public felt greatly alarmed and clamored to revive English-medium education for secondary levels (South China Morning Post, 15/3/2008).

Upon entering the second decade after the change of sovereignty, English is not only a language for better employment prospects, a gatekeeper to higher education, and a crucial factor that has contributed to the prosperity of Hong Kong today, but it is also a gateway to the world. English is psychologically indispensable for many Hong Kong people as it was once an important means that enabled many middle-class families to flee the communist threat from the PRC after the June 4th massacre in 1989 and resettle in other English-speaking countries like Canada and Australia. Although many of these have returned to Hong Kong after the ‘One country; two systems’ policy has proved successful, yet having English within reach as a means of outward mobility helps to give a sense of security. As a means of upward and outward mobility, no individuals or Hong Kong society as a whole can afford to give up English, because losing it would mean restricting one’s future opportunities. It is this social psychology that reaffirms the status of English as a ladder to success and hope for the future. One may not treasure something as much when possessing as when losing it. Respondents in the present study showed stronger integrative orientation toward English than their 2001 counterparts because they were not allowed to use the language as the medium of instruction under the mother tongue education policy. As an emotional rebound, they agreed more strongly that a Hongkonger should be able to speak fluent English. For the same possible reason, they perceived English even more strongly as a symbol of education, intelligence, and wealth since English education had been, for years, the privilege of the elites, even more so after the change of sovereignty. In fact, as pointed out by Dörnyei & Csizer (2002), owing to the rapid development of English as a world language, it is losing its national cultural base since it is no longer clear who the people are in the English community. Hence, it would be more meaningful to interpret students’ positive integrative orientation toward English, as revealed in the present study, as an admiration for their fellow elite Hongkongers and a general aspiration to Western culture.
Although there may be some truth in what Sun (2004) says about the end of English as the sole important language of importance and what Wikipedia (2009) says about the decrease of English use in Hong Kong, English is still the first choice of students. As shown in a supporting question in the survey, 64.2 percent of the respondents would choose English if they could speak only one language. Twelve years after the change of sovereignty, the superior status of English remains unchallenged and the language is still highly treasured as a means for upward and outward mobility. English is indispensable for Hong Kong, at least psychologically, if not practically.

7.3 Soft landing for Putonghua in Hong Kong

Pennington (1998) suggests that the next language shift in Hong Kong will take place with Putonghua being the center of change. As found in the present study, her prediction seems to have come true. Although Putonghua ranks last after Cantonese and English, as shown in the results of the present study, the greatest change in attitude took place in the case of Putonghua, showing that the role and status of the language have been greatly enhanced over the past few years. This, however, does not occur overnight with political imposition, but softly and gradually with the huge demographic and economic powers of China. In fact, the rise of Putonghua in Hong Kong is more circumstantial than planned. After the change of sovereignty, the connection between Hong Kong and the mainland has become closer than ever with intensive exchange at commercial, social, and individual levels. Lo (2008) interprets ‘mainlandization’ as a policy change after the handover to make Hong Kong politically and economically more reliant on mainland support. Common people use ‘mainlandization’ as a general term to refer to any practices or social phenomena (from neutral to negative) which resemble those on the mainland.7

It is this wave of mainlandization that helps Putonghua to land softly in Hong Kong, yet it is also the resistance to mainlandization that undermines the positive attitude toward the language. This may explain why respondents of this study show much more positive attitudes toward Putonghua than those in 2001, while still rating the language lower than Cantonese and English.

As mentioned in section 2.1, Hong Kong fell into its worst economic downturn in history after the change of sovereignty. Blows from the Asian Financial Crisis, the SARS Epidemic, and the world-wide economic recession triggered by the Financial Tsunami in 2008 all combined to throw Hong Kong into the arms of China for economic support. It is such a thirst for business opportunities that has driven Hong
Kong businessmen and professionals to learn Putonghua in order to gain access to the enormous market on the mainland. It is also the thirst for investment and economic stimulation that has further opened the doors of Hong Kong to mainland investors and visitors, who have brought along not only money, but also their language. In September 2009, the property market of Hong Kong reached its peak with record-high prices of HK$ 70,000 (around USD 9,000) per square foot despite the fact that the economic foundation had shown few clear signs of recovery. The biggest buyers of these super-luxurious properties are often from the mainland (Ming Pao Daily News 14/10/2009).

Being desperate to bring in more tourists to support the retail market, Hong Kong welcomes the ‘Individual Travel Scheme’ as a gift from the PRC that opened up all affluent cities of China to the tourist market of Hong Kong (Hong Kong Daily News 28/12/2008). As a result of the economic downturn, both mainland people and their money were wanted in various domains, including schools that brought in students across the border to boost student enrolment (Apple Daily News 13/9/09). Similarly, universities also sought to fill the surplus international student quota by admitting students from the mainland (Ming Pao Daily News 3/6/2009). Even hospitals have to wave their hands to the mainland residents so as to ensure a stable source of income (Ming Pao Daily News 3/11/2009). All these were bottom-up initiatives that enabled a soft landing for mainland people and their language. On top of all this, there are many other regular immigration policies (e.g., One-way Permit Scheme; the Admission Scheme for Mainland Talents and Professionals) that brought in large numbers of immigrants from mainland China. In July 2009, the population of Hong Kong reached seven million, much of the growth of which was contributed by new arrivals from the mainland (Ming Pao Daily News 6/10/2009).

The irresistible economic power of the PRC has helped to increase the vitality of Putonghua in Hong Kong, especially in the instrumental domain. In the integrative domain, the huge demographic power of the mainland group has also brought Hong Kong people into closer contact with Putonghua speakers and thus fostered a more accommodating attitude among them toward the language and its people. This may explain why the 2009 respondents claimed to like Putonghua and its speakers a lot more than their counterparts in 2001.

Despite the facilitating factors above, there are factors undermining the integrative orientation of students toward Putonghua and its speakers. No matter how rich some mainlanders may be, and no matter how successful China was in hosting the Beijing Olympic Games and sending its spaceship to the moon in 2008, negative news about the bad quality of Chinese products and the fact that many of the new arrivals
from the mainland come from a lower social stratum (Ming Pao Daily News 6/10/2009) have formulated a negative image of Putonghua speakers, detaching the language from a positive association with education, intelligence, and wealth (mean of item 3.12c = 2.48). For similar reasons, respondents do not favor the idea of quick integration with the mainland by allowing Putonghua to be used more widely in Hong Kong (mean of item 2.4 = 2.39). This may explain why the integrative orientation of the 2009 respondents was still not strong toward Putonghua (composite mean = 2.79) even though a significant increase has been shown in comparison with that of the 2001 study (mean = 2.47). As found in a supporting question in the present survey, only 7.1 percent of the respondents opted for Putonghua if they could speak only one language.

7.4 Trilingualism as a norm of the younger generation

As mentioned in the previous section, if students could only speak one language, 65.2 percent of them would choose English, 26.4 percent would choose Cantonese, and 7.1 percent would choose Putonghua. This may show their ultimate language choice after balancing the love for and the need for each language. However, multilingualism has replaced monolingualism as the norm of the 21st century, and languages can coexist at the same time without being mutually exclusive. As shown in the result of item 3.7, students in the present study display a strong wish to achieve a high proficiency in Cantonese (mean = 3.57), English (mean = 3.45), and Putonghua (mean = 3.19). Similarly, they agree significantly more strongly than the 2001 group that a Hongkonger should be able to speak fluent Cantonese (mean = 3.71), English (mean = 3.5), and Putonghua (mean = 3.13). This shows that the present students perceive trilingualism as a norm for Hongkongers, at least for the younger generation.

Dörnyei (2009) proposed the framework of ‘ideal self’ (i.e., personal hopes and aspirations) and ‘ought-to self’ (i.e., sense of responsibilities) and hypothesized that if proficiency in the target language is part of one’s ideal or ought-to self, it will serve as a powerful motivator to learn the language. As revealed from the results of the present study, there is no doubt that the three languages are part of the respondents’ ‘ideal self’ and ‘ought-to self’ as a student, after the notion of ‘biliteracy and trilingualism’ had been promoted in schools for 12 years.

Apart from this, the government’s pragmatic approach to languages also helps trilingualism to develop in Hong Kong. Frequent advertisements on TV have often packaged English as a language for better business and job opportunities. Recently, such tactics are also used in
promoting Putonghua. In an advertisement featuring a young local popular singer, Putonghua is depicted as a language for wider communication and a better future. Li (2000) argues that Hong Kong people are active agents of pragmatism in their choice of languages. Such a pragmatic approach that detaches a non-native language from a specific country or a specific group of speakers seems to have won special appeal for the people of Hong Kong. This has also helped English and Putonghua to thrive and stand up against the dominance of Cantonese in Hong Kong. Under an intense interplay of localization, mainlandization, and internationalization, pragmatic trilingualism seems to be the resolution which best serves the needs of Hong Kong for its special position in the 21st century, as China’s international city. Although a positive attitude toward the three spoken languages may not necessarily result in keener learning behaviors (Baker 1992), at least it paves the way to a higher level of trilingualism in the future.

8. Conclusion

After the change of sovereignty, Hong Kong has undergone the process of localization and mainlandization while adhering to internationalization as a means to maintain its advantage as a world-class city. An intense interplay among these forces then contributed to the language attitudes of the students.

As discussed in the previous section, there are factors facilitating the growth of vitality of each language, and yet there are also factors inhibiting the development of each language in postcolonial Hong Kong. It is the localization movement after the change of sovereignty that upholds the role of Cantonese as the dominant language, and mainlandization that helps Putonghua to land softly in Hong Kong. Under these challenges, English has lost some of its high functions in Hong Kong; yet it is the urge for internationalization that maintains its superior status, and it is also the fear of being too localized and mainlandized that makes English indispensable for Hong Kong. Under an intense interplay between localization, mainlandization, and internationalization, respondents in the present study display similar language attitude patterns as their counterparts in 2001, reflecting no drastic change in the linguistic equilibrium in Hong Kong society over the past years.

Similar to the 2001 study, respondents in 2009 tend to like Cantonese most and English second, and they rate English most highly for its instrumental values and social status, then Cantonese second. In both the integrative and instrumental domains, Putonghua still ranks lowest. Yet the most significant positive changes have taken place with Putonghua, indicating a gradual language shift toward the language (Pen-
nington 1998). In fact, from the instrumental perspective, there are signs that Putonghua has caught up substantially to the position of Cantonese. With the roles and status of Cantonese and English remaining stable and those of Putonghua rising along with the increasing influence of China in the world, gaps between the three languages are becoming narrower. Given a longer period of time, attitudes toward the three languages may become more equal when Putonghua puts down deeper roots in Hong Kong society. Until then, a higher level of trilingualism can be reached.

*Hong Kong Institute of Education*

**Notes**

1. There are 11 articles in Chapter 1 (General Principles) of the Basic Law Full Text, which can be retrieved from http://www.basiclaw.gov.hk/en/basiclawtext/chapter_1.html

2. Cantonese and Putonghua are two spoken varieties of Chinese, which are syntactically and phonologically different from each other and are, to a large extent, mutually unintelligible. Cantonese is a dialect of Southern China and Putonghua (now the lingua franca of the Chinese people) used to be a variety spoken in the Northern regions of the country.

3. CEPA stands for ‘The Mainland and Hong Kong Closer Economic Partnership Arrangement’. This policy opens up the huge market in China for Hong Kong goods and services. The ‘Individual Travel Scheme’ helps to bring in millions of mainland tourists from the most affluent cities in China to Hong Kong.

4. In 2007, China accounted for 15 million visitor arrivals as compared to a total of two million from North America, Australia, and New Zealand (Hong Kong Tourism Board, 2008).

5. There was no published information about the number of CMI and EMI schools in Hong Kong. According to an email enquiry to the Hong Kong Education Bureau on Dec. 17th 2008, the ratio between CMI and EMI schools is 3 : 1.

6. Given the large sample size of the two surveys, small differences between means can easily lead to significant t-test values smaller than 0.05 in statistical comparisons. For this reason, effect size values are adopted instead to reveal the differences between the two surveys.

7. A wide range of discussions about ‘mainlandization’ among common people was retrieved on 1st Dec. 2010 from Yahoo: http://hk.search.yahoo.com/search?_ylt=A8tU3wrDjPdMfE5%4%A7%E9%99%9B%E5%8C%96&ei=UTF-8&meta=rst%3Dhk&fr=FP-tab-web-t&xargs=0&pstart=1&b=1. Issues include self-censorship of mass media, lack of democracy, use of mainland language terms, scenic spots in Hong Kong, charge for admission, etc.

**References**


Hong Kong Daily News. 28/12/2008. Mainland visitors uphold the Hong Kong tourist industry (news in Chinese).


Mee Ling Lai is Assistant professor in the English department at the Hong Kong Institute of Education. Her research interests are in the areas of sociolinguistics and teaching English as a second language. Her recent publications include ‘Social class and language attitudes in Hong Kong’ (*International Multilingual Research Journal*)
Tracking language attitudes in postcolonial Hong Kong


Address for correspondence: English department, HKIED, 10, Lo Ping Road, Tai Po, Hong Kong.
e-mail: mllai@ied.edu.hk
Acoustic Analyses and Intelligibility Assessments of Timing Patterns Among Chinese English Learners with Different Dialect Backgrounds

Hsueh Chu Chen
Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: “The final publication is available at link.springer.com”.
Acoustic Analyses and Intelligibility Assessments of Timing Patterns Among Chinese English Learners with Different Dialect Backgrounds

Hsueh Chu Chen

Abstract This paper includes two interrelated studies. The first production study investigates the timing patterns of English as spoken by Chinese learners with different dialect backgrounds. The second comprehension study explores native and non-native speakers’ assessments of the intelligibility of Chinese-accented English, and examines the effects of the listeners’ language backgrounds on their perceptions of Chinese-accented English. The results showed that the Hong Kong (HK) group performed better in unstressed syllable duration compared with the Taiwan (TW) and Beijing (BJ) groups. The results also revealed that all six listener groups achieved at least 78 % intelligibility, with the native speaker accent achieving the highest rating, followed by the HK, TW, and BJ accents. A shared first language (L1) background may have little or no impact on intelligibility. The speech properties might prevail over the shared L1 effect. All listeners perceived inappropriate word-stress shift and consonant cluster simplifications to be the most unintelligible features.

Keywords Intelligibility · Foreign accent · Production · Perception · Inter-language phonology

Introduction

A number of scholars have envisioned a bright future for Chinese English as the most common variety of English spoken in Asia, with the largest population of speakers. However, most previous studies on the phonological features of Chinese-accent English focused on the segmental features (e.g., Chang 1987; Deterding 2006); the suprasegmental parts of English with Chinese accent received much less attention, even though a number of studies have indicated that suprasegmental factors—such as stress, linking, pause, and intonation—contribute significantly to listeners’ perception of intelligibility and foreign accent (Anderson-Hsieh et al. 2006).
As we know, Cantonese and Mandarin are both dialects of the Chinese languages, but speakers of these two dialects are not mutually intelligible. Very few studies have even considered speech timing in learners of varying dialect backgrounds—such as Chinese people from Taiwan (TW), Hong Kong (HK), and Beijing (BJ)—as well as learning environments, for example, learning English as a second language (ESL) in Hong Kong or learning English as a foreign language (EFL) in Taiwan or mainland China.

The distinction between the ESL and EFL contexts is becoming increasingly fuzzy with the rapid growth of English in EFL areas. The classification criterion between the ESL and EFL contexts should be adapted to the current surge of English worldwide. According to Graddol (1998), “The main distinction between a fluent EFL speaker and an ESL speaker depends on whether English is used within the speaker’s community (e.g., country, family)” (p. 11).

Unlike Mandarin speakers in Taiwan or mainland China, Cantonese speakers in Hong Kong inhabit a former British colony and have extensive experience of its English-speaking environment. Most elite students in the country receive their secondary and higher education in English. Outside the classroom, they have a specific, practical need for English and ample opportunity to use it. It is expected that greater exposure to native English speakers in everyday life will help Cantonese learners in Hong Kong gain better timing patterns, such as speech rhythm and fluency. Therefore, these speakers are considered to be ESL speakers.

On the contrary, Taiwan and Beijing are places where English is not the dominant language. Students share the same language, Mandarin, and culture. The English teacher may be the only English speaker they are exposed to. Moreover, students have limited exposure to English-speaking culture, most often only through TV, movie, or music. Outside of the classroom, students have very few opportunities to use English. For some, learning English may not have any obvious practical benefit. Therefore, these speakers are considered to be EFL speakers.

A more dynamic classification could be made based on speakers’ English proficiency and the settings in which English is actually used instead of the rigid classification based on speakers’ national identity. Therefore, it would be of great value to determine whether and in what manner ESL learners (e.g., Hong Kong learners) with extensive experience in English-speaking districts could significantly outperform EFL learners (e.g., Taiwan and Beijing learners) with minimal language exposure in timing components of speech intelligibility.

In terms of timing patterns, languages have been classified as either stress-timed, such as English, or syllable-timed, such as Chinese (Abercrombie 1967; Pike 1945). A stress-timed language can be distinguished by the stressed syllables that recur at equal intervals of time. Ladefoged (2006) mentioned that “we now know that this is not true...in English the rhythm of a sentence depends on several interacting factors, not just the stress.” (p. 243). English has a tendency to stress-timed rhythm, but it is not strictly stress-timed. On the other hand, a syllable-timed language is so named because of the role of the syllable in the utterance. Although this distinction between syllable-timed and stressed-timed languages may seem rather basic and remain debatable, it has been widely recognized in the literature of second-language phonology (Celce-Murcia et al. 2001; Cruttenden 2001; Roach 2009). Because of the differences between the timing patterns of Chinese and English, it is critical to further explore how these differences influence intelligibility of Chinese English. A number of factors make it more complicated: Chinese is a tone language and has four basic lexical tones. Each syllable receives a tone and each tone is more or less of equal length. Among Chinese dialects, Mandarin speakers from Beijing and Taiwan somewhat differ in the use of neutral tone (Jun 2005); moreover, Cantonese has no neutral tone in its phonology.
The neutral tone in Mandarin usually comes at the end of a word or phrase, and is pronounced in a light and short manner. Because of this characteristic, it is considered analogous to an unstressed syllable. These unstressed syllables are often referred to as having the “fifth tone” or a “neutral tone” (Chao 1968). In Beijing Mandarin, an unstressed syllable reduces (Duanmu 2007) and is shorter in duration (Chao 1968). However, in Taiwan Mandarin, the neutral-tone syllables behave differently compared to Beijing Mandarin. In Taiwan Mandarin, unstressed syllables are less frequent (Duanmu 2007). The neutral tone in Taiwan Mandarin functions like a lexical tone rather than a neutral one. Moreover, Taiwan Mandarin is often described as more syllable-timed than Beijing Mandarin (Kubler 1985). Thus, it would be reasonable to hypothesize that Beijing speakers would perform better in vowel reduction in unstressed syllables than both Cantonese and Taiwanese speakers. Due to the differences of tone use among people from different dialect backgrounds, when Chinese learners learn the English language, different patterns of language transfer might occur. Speech perceived by people from different language backgrounds may demonstrate different levels of intelligibility.

This study aims to examine the timing properties of English spoken by three groups of Chinese ESL learners: those from BJ, mainland China; TW; and HK. In addition, a comparison between these learners and native speakers of American English (NS) is conducted. Seven timing variables—stressed and unstressed vowel duration, pause duration, consonant-vowel linking duration, consonant cluster duration at initial and final positions, and speech rate—were acoustically measured and compared. In the current study, the tasks were administered as part of a larger project (Chen and Wang 2013) that investigated multiple aspects of English phonological patterns in Chinese speakers with different dialect backgrounds, and compared native and non-native listeners’ assessments of foreign accent and intelligibility. This study seeks to address the following research questions:

1. To what extent do Chinese English learners with different dialect backgrounds (TW, HK, and BJ) display specific patterns on the seven acoustic timing variables (stress vowel duration, unstressed vowel duration, consonant-vowel linking duration, pause duration, consonant cluster duration at initial and final positions, and speech rate) that deviate from those of native speakers of American English (the norm)?
2. What are the differences between native and non-native listeners’ perceptual judgments of intelligibility of Chinese English learners with different dialect backgrounds?
3. What are the effects of the listener’s language background in interaction with their perceptions of Chinese-accented English?

**Literature Review**

Intelligibility studies identify the aspects of a foreign accent that negatively affect intelligibility in a specific context, which helps prioritize the sounds that must be produced as accurately as possible in order for communication to succeed. According to Derwing and Munro (2005), Jenkins (2000), and Schiavetti (1992), there is no satisfactory definition of intelligibility. Pickering (2006) noted that a common conceptualization is Smith and Nelson’s (1985) definitions on both intelligibility and comprehensibility: Intelligibility is the ability of the listener to recognize individual words or utterances, whereas comprehensibility is the listener’s ability to understand the meaning of the word or utterance in its given context. Munro and Derwing define “intelligibility” as the “extent to which a speaker’s message is actually understood by a listener” (Munro and Derwing 1995a, p. 76), as measured by written
transcriptions or oral repetitions. Further, they define “comprehensibility” as the degree of difficulty involved in processing the speaker’s message, as measured by subjective “perceived comprehensibility” ratings or listener processing times (Munro and Derwing 1995b).

According to Kachru (2008), the first study to determine if this view of intelligibility is testable was reported in Smith and Rafiqzad (1979). It required the participating subjects to listen to a tape of a passage and fill in words from it in its printed cloze text version. Being able to approximate a representation of sounds heard by writing them down was counted as a successful demonstration of intelligibility. Smith and Bisazza (1982) designed their experiment to determine whether select varieties of English are comprehensible to users who are not familiar with them by requiring subjects to match pictures with taped sentence- or passage-level texts read by speakers of different varieties. The subjects also answered questions about what they heard. The study concluded that “one’s English is more comprehensible to those people who have had active exposure to it” (p. 269). In Rogers et al.’s (2004) experiment in investigating the effects of noise and proficiency on the intelligibility of Chinese-accented English, four Chinese speakers were selected based on the intelligibility of their speech in quiet from a group of eight speakers recorded in a previous study (Rogers 1997), in which native listeners orthographically transcribed sentences produced by the Chinese speakers. In their study, sentence intelligibility was scored as number of words correctly transcribed.

In Matsuura et al. (1999), comprehension check questions and the partial dictation task are used to detect the intelligibility of English. Each dictation question included 10 blanks, with both content and function words missing. In addition, five multiple-choice comprehension check questions were prepared for each speech sample to assess the listeners’ understanding of the uttered message. Field’s (2005) research on intelligibility focused on the role of lexical stress; he adopted the most widely adopted approach to researching intelligibility (Derwing and Munro 1997; Munro and Derwing 1995a, b). He asked participants to evaluate samples of non-native speech for prosody, accentedness, and other features, and then to rate them objectively for intelligibility, often on the basis of how accurately they had been transcribed.

Based on the reviews above, dictation is one of the most effective ways to assess intelligibility. By asking participants to transcribe the English words spoken by Chinese speakers rather than to complete comprehension questions commonly used in accent studies (e.g., Kirkpatrick et al. 2008), the current research measured the basic level of intelligibility of the pronunciation of Chinese speakers for listeners of different language backgrounds. This attempt helps us better understand the true nature of intelligibility.

This research is a follow-up extension of Chen and Chung (2008) and Chen’s (2010, 2011) studies. In these studies, they investigated the difficulties in English speech timing patterns encountered by 30 Taiwanese learners with different English proficiency levels. By examining both acoustic measures and perceptual ratings given by 10 English native listeners, they identified critical variables that affect native listeners’ perceptions of foreign accents. This led to some major findings: Taiwanese learners display very different patterns in acoustic timing variables compared to native English speakers. The acoustic timing patterns in American English are stress-timed oriented, whereas those in the English spoken by Taiwanese learners are more syllable-timed. Speech rate is the primary predictor of the perception of foreign accent. If the variable of speech rate is excluded, then vowel reduction and linking duration are the most heavily weighted variables displayed by native speakers in discerning foreign accents. In addition to a group of Chinese learners from Taiwan, the current study examined two more Chinese first language (L1) groups from different dialect backgrounds, that is, learners from Hong Kong and Beijing, so that a comprehensive comparison could be made and the timing patterns produced by various groups of Chinese learners of English could be better captured.
When evaluating foreign-accented speech, listeners are commonly influenced by their own linguistic backgrounds, as well as their experiences and familiarities with different speech varieties. Van Wijngaarden et al. (2002) claimed that Dutch listeners found native English speakers to be more intelligible than their own non-native English accent. Bent and Bradlow (2003) stated that non-native listeners might find L2 speech more intelligible than native speech, whereas the opposite might be true for native listeners. Further, Major et al. (2002) showed that native speakers of Spanish found Spanish-accented English easy to understand, but native speakers of Chinese found Chinese-accented English difficult to understand. Since the results of previous studies are mixed, it is not known to what different degree a diverse group of listeners might share a response to L2 speech.

Further, properties of the speech itself also influenced the judgments of foreign-accented speech. Chen (2011) investigated multiple aspects of English phonological patterns spoken by Chinese low-achieving English learners with two dialect backgrounds (Mandarin and Cantonese) and explored native/non-native speakers’ perceptual judgments of the intelligibility of Chinese-accented English. The results showed that all listener groups perceived both the Cantonese and Mandarin accents to be at least 70% intelligible. Mandarin-accented English was easier to understand than Cantonese-accented English was. In Cantonese-accented speech, several word-stress shifts and double-primary stresses evidently made words unintelligible. Among five groups of listeners, the native English-speaker group and two Chinese-speaker groups gave lower ratings than the ESL and EFL groups, thereby demonstrating strict native English norms.

However, in Chen’s study, the backgrounds of Mandarin speakers were not well defined or controlled. There are numerous Mandarin speakers in Greater China for whom Mandarin is the second dialect rather than their home language. Several properties of Beijing Mandarin, such as the use of a neutral tone, differ from those of Taiwan Mandarin. The effects of negative transfer from different first languages (dialects) on second language (i.e., English) learning may vary. Moreover, English proficiency of the speakers was another factor determining the intelligibility of the listeners, which was less emphasized in Chen’s study. In addition, in being assessed on timing properties, lower-achieving English speakers and higher-achieving English speakers likely have different perceptual judgment of intelligibility.

In addition to the critical issues mentioned above, several other measurement issues have not yet been explored thoroughly. First, Chen and Chung (2008) found that Taiwanese speakers of English often dropped the final consonant or inserted an additional schwa between consonants or after the final consonant. Chan (2006) examined the strategies employed by Cantonese speakers of English when pronouncing initial consonant clusters. She found that deletion and substitution were the most common strategies in initial consonant clusters used by secondary and university students in Hong Kong. Therefore, it is worth further exploring this variable at both initial and final positions rather than only at either one.

Second, apart from the adoption of the variability index (VI), as proposed in Chen and Chung (2008), another way of describing rhythmic differences is to consider how much variation in length occurs within a sentence. Such differences can be quantified by calculating the pairwise variability index (PVI) developed by Low et al. (2000). The PVI is calculated by finding the average ratio of all the adjacent units in the utterance. The PVI for vowels has been calculated for many languages, for example, the PVI value of the English language is around 55, whereas that for Chinese languages is around 25, as shown by Ladefoged (2006, p. 246). Through these calculations, we can further explore the differences in speech timing produced by Chinese learners with different dialects and learning environments. This will enable better understanding of variations among the groups.
Jenkins (2002) noted that because non-native speakers (NNSs) outnumber NSs, teaching English as an international language is more realistic and relevant than teaching NNSs to imitate NSs’ accents. However, the nature of intelligibility from the perspective of both native and non-native listeners has not been fully clarified and neither has the perceptual significance of phonetic and phonological variables of intelligibility been fully assessed. The effects of the listener’s language backgrounds in interaction with their perceptions of Chinese-accented English remained unclear. Listener judgments of variables such as intelligibility may help us both to understand the manner in which people respond to particular accented speakers and to ascertain the difficulty they encounter in understanding a particular phonological variable. We can also prioritize the components of teaching pronunciation, thereby further improving learners’ intelligibility and facilitating their international communication.

To clarify the aspects mentioned above, the two studies—the production study by acoustic measurements and the comprehension study by intelligibility assessments—were conducted and are discussed separately in the following sections.

The Production Study

Methods

Participants

Four groups of 10 participants were invited to read aloud from an English text, including 10 EFL Chinese learners from TW, 10 EFL Chinese learners from BJ, 10 ESL Cantonese learners from HK, and 10 native speakers of NS. All the Chinese learners possessed English proficiency at the beginner–intermediate level, with TOEFL scores of around 500 points (or 60 on the internet-based test), which is equivalent to an IELTS overall band score of 5.0. Each group consisted of an equal number of male and female speakers between the ages of 18 and 28 years. Further, all the subjects were recruited from local universities or community colleges. The NS group included 10 college students from the University of California, Berkeley. The speech samples produced by this group were considered as the NS norm to which those of the three Chinese groups were compared. They were monolingual English speakers, Californians from the West Coast of the US. Each speaker produced approximately 83 syllables in 5 sentences selected from the diagnostic passage. Thus, a group of 10 speakers generated approximately 830 pieces of information on this timing variable. According to the literature, and the power calculation of the software G*Power 3, this amount of information was sufficiently large to yield a high power of above 0.95 in revealing group differences.

Materials

The materials employed in the acoustic section included a short questionnaire on personal background and a diagnostic passage for acoustic measurement. A diagnostic passage with 14 sentences selected from Teaching Pronunciation (Celce-Murcia et al. 2001) was used as the reading material. This material was used in Chen and Chung (2008); using this article has certain advantages (as stated in their article). For example, this passage contains different types of sentences, which can help counterbalance the effects of various sentence types on the timing patterns produced by Chinese learners. The following five sentences, corresponding to these sentence types, were selected for further acoustic analysis.
1. Wh- questions: Why do people usually have an accent when they speak a second language?
2. Declarative sentences: Most native speakers of English can, for example, recognise people from France by their French accents.
3. Yes–No questions: Does this mean that accents can’t be changed?
4. Tag questions: Old habits won’t change without a lot of hard work, will they?
5. Closed-choice questions: Will you manage to make progress, or will you just give up?

Data Collection Procedure

A reading aloud task rather than spontaneous speaking activity was used because consistency across samples is important for measurement and comparison. Use of spontaneous speech for measurement would be problematic because it varies so widely in content and utterance complexity that these variables would undoubtedly contribute to large variability.

The participants were informed of the research goal and recording procedure before the session began. Each participant then read the passage individually in a quiet room while being recorded on a notebook computer. Before recording began, each participant was asked to read the diagnostic passage and allowed to request for help and practice words that he/she was not familiar with. The recording took approximately 30 min for each participant.

Measurement of the Seven Timing Variables

Seven variables (syllable duration, unstressed syllable duration, pause duration, linking duration, consonant cluster duration at initial position, consonant cluster duration at final position, and speech rate) were measured.

Stressed and Unstressed Syllable Duration

The speech data were segmented into syllables and then analyzed for stressed and unstressed vowel durations using the Praat program. The total number of syllables analyzed was 3,320 (4 groups × 10 participants × 83 syllables in 5 sentences). In the measurement of vowel duration, if there is no vowel to be measured in both polysyllabic content words and monosyllabic function words, particularly in the case of syllabic consonants, the following are the general guidelines adapted from Ong et al. (2005): “a voiceless plosive is to be measured from the onset of the plosive release, and this was devised for polysyllabic content words. It was also applied to all function words containing a voiceless plosive, such as ‘to’, preceding a /s/ since there is a tendency for the /s/ to be deleted in fast speech. This adaptation of the guideline was eventually extended to other instances where the /s/ was deleted, such as ‘just’ where /st/ was measured to represent the vocalic segment and the /st/ in the coda was not included.” (p. 79)

Pause Duration

According to Duez (1982), a pause is defined as an interval of the oscillographic trace where the amplitude is impossible to differentiate from that of the background noise. For an exact definition of a pause, it is necessary to determine a threshold where a pause begins to be a pause. While some studies work with a threshold of below 50 milliseconds (ms) (pause and segment detection with an automatic procedure) (e.g., Lee and Oh 1999), in other studies only silent pauses longer than 500 ms are regarded (for a review see O’Connell and Kowal 1983).
Other examples are 150 (Tsao and Weismer 1997) and 100 ms (de Pijper and Sanderman 1994; Griffiths 1991). As different thresholds were listed in several studies, ranging from 50 ms up to several hundred ms, in this research, only those pauses that were longer than 100 ms were considered pauses. Pauses were identified by ear perception and the acoustic cue of intensity.

**Linking Duration**

To facilitate fluent and natural speech, native speakers often link adjacent words together instead of pronouncing them in isolation. Linking is the phonological process of linking two or more words and pronouncing them as one. As Hieke (1987) observed, in English, linking between two words can take place between two vowels, between two consonants, and between a consonant and a vowel (CV). Because the high frequency of CV linking occurs in daily speech, this paper confines itself to CV linking only. When the first word ends with a consonant and the next word begins with a vowel, the durations of the final consonant of the first word and the beginning vowel of the following word were measured. For example, the duration of n and a in “An apple” are measured. The following are the 11 tokens of potential linking (L) and were those that were selected for analysis.

1. Why do people usually have an accent (L1) (L2) when they speak a (L3) second language?
2. Most native speakers of (L4) English can, for example (L5), recognize people from France by their French accents (L6).
3. But old (L7) habits will not change without a (L8) lot of (L9) hard work, will they?
4. Does this mean that accents (L10) cannot be changed?
5. Will you manage to make progress, or will you just give up (L11)?

**Consonant Cluster Duration**

A consonant cluster (CC) is a group of consonants that appear together in a syllable without a vowel between them. For example, in the second tested sentence, the sp in “speakers” and fr in “from” are the initial CCs. In addition, the st in “most” and the nts in “accents” are final CCs. For each participant, a computation is made based on the measured values of the initial and final CC duration.

**Speech Rate**

Speech rate is based on the total duration of a sentence, including pauses. The most common measurements of speech rate are syllables per second (SPS) and words per minute (WPM) (Buck 2001). The total duration and SPS will be calculated. Because the measurement of syllable duration is the major focus of this study and because the selected speech samples from the four groups are short and identical, the measurement of WPM will not be employed.

**Calculation of the Seven Timing Variables**

The VI of each of the seven variables (syllable duration, unstressed syllable duration, pause duration, linking duration, consonant cluster duration at initial position, consonant cluster...
duration at final position, and speech rate) was calculated. The VI for norm-referenced comparison is defined as

\[ \text{VI} = \sqrt{\frac{\sum_{k=1}^{K} (X_k - E_k)^2}{K}}, \tag{1} \]

where \(X_k\) is the duration of the \(k\)th component, \(E_k\) is the mean duration of the \(k\)th component over native speakers of English (treated as the norm), and \(K\) is the number of components in the sentence. When a speaker produces exactly the same timing pattern as the norm, VI will be zero. The larger the VI, the greater the deviation of the speaker’s timing pattern from the norm. NNSs are expected to show larger VIs than native speakers.

Further Analysis of PVI

To examine the timing patterns from a different perspective, we measured the vowel duration for each syllable in five sentences across four groups, calculated the successive vowel durations, and used the PVI developed by Low et al. (2000) in an attempt to capture the timing patterns of the three Chinese groups. In Ladefoged (2006), English and Chinese have been categorized as a stress-timed language and a non-stress-timed language with PVI values of 55 and 25, respectively. The higher of the PVI value, the larger variations in vowel lengths. The PVI is expressed mathematically as

\[ \text{PVI} = 100 \times \left[ \frac{\sum_{k=1}^{m-1} \left| d_k - d_{k+1} \right|}{(d_k + d_{k+1})/2} \right] / (m - 1), \tag{2} \]

where \(m\) is the number of items in an utterance and \(d\) is the duration of the \(k\)th item. The index has been multiplied by 100 to avoid fractional values. According to Low (2010), two modifications are necessary:

The minimum duration for each vowel was set at 30 ms so that we can overcome issues such as a vowel being completely elided due to extreme reduction. Second, when two extremely short vowels occur successively, the utterance was not considered, as this might return large PVI values, which might not be meaningful in capturing overall rhythmic patterning. (p. 399)

These two suggestions were considered in this study.

Results

Overall Performance of Seven Timing Variables

The seven acoustic timing variables performed by the three groups of Chinese speakers were compared with those of the native speakers of English, using VI. Because VI involves the norm \((E_k)\), it is a norm-referenced VI. If a subject produces exactly the same timing pattern as the norm, his or her VI will equal zero. As the value of the VI increases, the subject’s timing pattern will deviate more from the norm. The difficulties in English speech timing patterns encountered by Chinese learners could then be identified by calculating the VI.

Table 1 summarizes the means, standard deviations (SD), minimum values (Min), and maximum values (Max) of VI in seven timing variables for the four groups. The results of one-way analysis of variance (ANOVA) of the VI of the four groups in terms of the seven variables for all three groups of Chinese learners showed significant differences compared with those
Table 1 Means, standard deviations (SD), minimum values (Min) and maximum values (Max) of VI in seven timing variables for the four groups

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD</td>
<td>TW</td>
<td>10</td>
<td>0.018</td>
<td>0.003</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>0.016</td>
<td>0.002</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>0.016</td>
<td>0.002</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.008</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>USVD</td>
<td>TW</td>
<td>10</td>
<td>0.010</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>0.009</td>
<td>0.001</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>0.011</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.005</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>PAU</td>
<td>TW</td>
<td>10</td>
<td>0.033</td>
<td>0.009</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>0.037</td>
<td>0.016</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>0.023</td>
<td>0.007</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.011</td>
<td>0.006</td>
<td>0.005</td>
</tr>
<tr>
<td>LIN</td>
<td>TW</td>
<td>10</td>
<td>0.027</td>
<td>0.011</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>0.033</td>
<td>0.015</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>0.024</td>
<td>0.008</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.009</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>ICC</td>
<td>TW</td>
<td>10</td>
<td>0.010</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>0.011</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>0.010</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.005</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>FCC</td>
<td>TW</td>
<td>10</td>
<td>0.021</td>
<td>0.005</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>0.023</td>
<td>0.005</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>0.022</td>
<td>0.005</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.012</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>SR</td>
<td>TW</td>
<td>10</td>
<td>3.817</td>
<td>1.746</td>
<td>2.100</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>10</td>
<td>3.164</td>
<td>1.371</td>
<td>1.092</td>
</tr>
<tr>
<td></td>
<td>BJ</td>
<td>10</td>
<td>2.291</td>
<td>0.499</td>
<td>1.216</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>10</td>
<td>0.322</td>
<td>0.131</td>
<td>0.113</td>
</tr>
</tbody>
</table>


for native English speakers. Following the ANOVA, for further exploration and comparing the mean of one group with that of another, Fisher’s Least Significant Difference (LSD) test was used. All the results will be reported in the following sections. In general, among the three Chinese groups, the BJ group performed significantly closer to native speakers in speech rate (*NS < BJ < TW = HK*) and pause (*NS < BJ < TW = HK*), while the HK group performed significantly better in unstressed vowel duration (*NS < HK < TW = BJ*). Apart
from this, not many significant differences could be found in other timing variables among the three Chinese groups.

**Stressed and Unstressed Vowel Durations**

This section measures and compares the VI of the vowel durations from the four groups. In these calculations, the mean timing pattern of the native speakers served as the norm, and those of the Chinese learners were then compared to the norm. Table 1 illustrates the means, standard deviations, minimum values, and maximum values of VI in stressed vowel duration (SVD) and unstressed vowel duration (USVD) for the four groups. On average, TW ($M = 0.018, M = 0.010$), HK ($M = 0.016, M = 0.009$) and BJ ($M = 0.016, M = 0.011$) had larger VI than NS ($M = 0.008, M = 0.005$). All three Chinese groups generated approximately twice the deviance in vowel duration in both stressed and unstressed syllables from the norm as NS did. The range for NS was 0.005–0.023 for SVD and 0.004–0.007 for USVD; whereas for Chinese groups, it was 0.012–0.023 for SVD and 0.007–0.014 for USVD. The minimum value for Chinese groups in USVD (0.007) was similar to the maximum value for NS (0.007), thereby suggesting that all the Chinese subjects deviated from the norm far more than any NS subject.

The VI of SVD for these four groups was statistically significant ($F_{3,36} = 15.590, p < 0.001$). Following the ANOVA, the post-hoc LSD test was used. There was a significant difference between the NS group and the other three Chinese groups ($p < 0.01$). No statistically significant differences were found among the TW, HK, and BJ groups. The VI of USVD for these four groups was statistically significant ($F_{3,36} = 36.468, p < 0.001$). Following the ANOVA, the post-hoc LSD test was used. There was a significant difference between the NS group and the other three Chinese groups ($p < 0.01$). There was a significant difference between the HK and BJ groups ($p = 0.000$). There was also a significant difference between the HK and TW groups ($p = 0.023$). The results suggest that the HK group performed significantly closer to the NS group than the TW and BJ groups in terms of unstressed syllable.

**Further Analysis of PVI**

Since HK performed slightly but significantly better in unstressed vowels than the other two Chinese groups (see Table 1) in terms of VI, which seems the closest to native speakers, it is reasonable to hypothesize that the PVI of HK would be much closer to NS. It must be noted that PVI was calculated by measuring successive vowel durations in an attempt to further capture the timing patterns of the current three Chinese groups. As mentioned before, English and Chinese have been measured with PVI values of 55 and 25, respectively. In the current study, the results show that there are the differences among BJ with PVI 32.64, TW with PVI 37.55, HK with PVI 39.76, and NS speakers with PVI 52.30, as shown in Table 2.

The one-way ANOVA showed that differences in PVI syllable duration for these four groups were statistically significant ($F_{3,36} = 29.286, p < 0.05$). Based on the post-hoc LSD test, there was a significant difference among the BJ, HK, TW, and NS groups ($p < 0.001$). Among the three Chinese groups, there is a significant difference between HK and BJ groups ($p = 0.002$), as well as between TW and BJ groups ($p = 0.031$). No statistically significant differences were found between HK and TW groups ($p = 0.318$). Figure 1 illustrates the timing patterns among four groups. If the groups are connected by a line, then they do not differ statistically significantly at the 0.05 nominal levels. Even though there is no language
Table 2  PVI values for four groups

<table>
<thead>
<tr>
<th>Participants</th>
<th>TW</th>
<th>HK</th>
<th>BJ</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVI value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>37.55</td>
<td>39.76</td>
<td>32.64</td>
<td>52.30</td>
</tr>
<tr>
<td>SD</td>
<td>5.32</td>
<td>4.83</td>
<td>5.29</td>
<td>4.02</td>
</tr>
<tr>
<td>Max</td>
<td>49.56</td>
<td>48.36</td>
<td>41.21</td>
<td>58.69</td>
</tr>
<tr>
<td>Min</td>
<td>30.15</td>
<td>33.91</td>
<td>26.26</td>
<td>46.30</td>
</tr>
</tbody>
</table>

Fig. 1  The timing patterns among four groups on the syllable-stress timing continuum

which is totally syllable-timed or totally stress-timed, the results could provide empirical proof that on the syllable-stress timing continuum, TW English and HK English were found to be significantly more stress-timed than BJ English.

The PVI has been shown to effectively capture the broad rhythmic typology between stress-based and syllable-based varieties of English, as in Low (2010). As predicted, we found that the acoustic timing patterns in American English are stress-timed oriented, whereas those in the English spoken by Chinese learners (particularly for the BJ group) are more syllable-timed. In this study, all the Chinese participants, being ESL or EFL learners, had rather limited English experience and proficiency and remained at the developmental stage of phonological acquisition; that is, they performed the syllable-timing nature of rhythm patterns in varying degrees: BJ performed more syllable-timed orientation than the HK and TW groups did. This disparity can be attributed to the native language transfer (i.e., the syllable-timed nature of the Chinese language), which strongly influenced the syllable duration produced by Chinese learners of English.

Pausing Duration

As mentioned above, the VI of pause duration was measured to show participants’ timing ability in terms of pause. Only those pauses greater than 100 ms, the cut-off for a pause (Griffths 1991), were considered pauses. In Table 1, the BJ group (M = 0.023) performed significantly closer to native speakers (M = 0.011), while the HK group (M = 0.037) and TW group (M = 0.033) did not show any significant differences in pause duration. The ANOVA showed that the differences between these four groups were statistically significant (F3,36 = 12.737, p < 0.001). The post-hoc LSD test showed a significant difference between NS and the other three groups (p < 0.001). Among the three Chinese groups, there is a significant difference between the HK and BJ groups (p = 0.005). There was no significant difference between the HK and TW groups (p = 0.349). Since there is a close relationship between the pause duration and the overall speed (discussed in the speech rate section), it is reasonable to hypothesize that the less number of inappropriate pauses produced by the BJ group is due to their fast speaking speed.

The major pauses (MP) produced by the four groups of speakers are presented in “Appendix 1”. It is noted that MP in “Appendix 1” is defined as the pause produced by three or more
Table 3  Mean of each CV linking duration for the four groups

<table>
<thead>
<tr>
<th>Group</th>
<th>have an accent (s)</th>
<th>an accent (s)</th>
<th>speak a (s)</th>
<th>speakers of (s)</th>
<th>for example (s)</th>
<th>French accents (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TW</td>
<td>0.28</td>
<td>0.43</td>
<td>0.43</td>
<td>0.37</td>
<td>0.13</td>
<td>0.68</td>
</tr>
<tr>
<td>HK</td>
<td>0.29</td>
<td>0.37</td>
<td>0.36</td>
<td>0.38</td>
<td>0.17</td>
<td>0.57</td>
</tr>
<tr>
<td>BJ</td>
<td>0.20</td>
<td>0.23</td>
<td>0.30</td>
<td>0.30</td>
<td>0.15</td>
<td>0.44</td>
</tr>
<tr>
<td>NS</td>
<td>0.11</td>
<td>0.21</td>
<td>0.09</td>
<td>0.12</td>
<td>0.10</td>
<td>0.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>that accents (s)</th>
<th>but old (s)</th>
<th>without a (s)</th>
<th>lot of (s)</th>
<th>give up (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TW</td>
<td>0.37</td>
<td>0.39</td>
<td>0.19</td>
<td>0.24</td>
<td>0.34</td>
</tr>
<tr>
<td>HK</td>
<td>0.55</td>
<td>0.41</td>
<td>0.23</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>BJ</td>
<td>0.34</td>
<td>0.39</td>
<td>0.33</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>NS</td>
<td>0.20</td>
<td>0.24</td>
<td>0.15</td>
<td>0.08</td>
<td>0.20</td>
</tr>
</tbody>
</table>

than three speakers in one group. The number in the parentheses indicates the number of subjects who produced this pause. The Chinese groups produced many more pauses within the text than the NS group did. Most of the pauses produced by the NS group were located at places in which there is punctuation (e.g., commas); however, no specific patterns, such as syntactical boundaries, can be found among three Chinese groups.

Linking Duration

As Hieke (1987) observed, linking can occur in English between two consonants, between a consonant and a vowel (CV), or between two vowels. Due to its high frequency, this study was limited to CV linking between words only. For each participant, a computation was made based on the measurement of linking duration. Table 1 reveals that the TW, HK, and BJ groups had larger overall VI in linking duration (0.027 for TW, 0.033 for HK, and 0.024 for BJ), while the NS had the smallest VI (0.009). The ANOVA showed that the differences between these four groups were statistically significant (F\textsubscript{3,36} = 9.157, p < 0.001). The post-hoc LSD test did not show any significant difference within the three Chinese groups. This indicates that all three Chinese groups had not yet mastered CV linking like native English speakers.

Table 3 summarizes the mean of each CV linking durations for the four groups. The three Chinese groups used CV linking significantly less often than NS, and the quantitative differences were dramatic. The analysis of forms for linking showed that Chinese learners demonstrated a tendency to preserve word boundaries. The absence of linking between segments accentuates the impression of syllable timing in Chinese-accented English. In CV linking, the word boundaries were maintained by the insertion of pause and glottal stops. It is possible that the tendency to keep vowels intact may be related to a concern for intelligibility.

Consonant Cluster Duration at Initial (ICC) and Final Positions (FCC)

Table 1 presents the means, standard deviations, maximum values, and minimum values of VI of consonant clusters durations at initial and final positions for the four groups. The TW, HK, and BJ groups had larger overall VI in ICC duration (0.010 for TW, 0.011 for
HK, 0.010 for BJ), while the NS had the smallest VI (0.005). Further, the TW, HK, and BJ groups had larger overall VI in FCC duration (0.021 for TW, 0.023 for HK, and 0.022 for BJ), while the NS had the smallest VI (0.012). The ANOVA result shows that the duration of initial and final consonants produced by three groups and Chinese learners are significantly different from that produced by native speakers: ICC, $F_{3,36} = 10.540, p < 0.001$, and FCC, $F_{3,36} = 12.761, p < 0.001$.

Table 4 summarizes the percentage of deletion and epenthesis in initial and final consonants among the four groups. For consonant deletion, the NS (39%) and HK (46%) groups displayed more final consonant deletion than the TW (28%) and BJ (13%) groups, while for initial consonant clusters, the HK group (13.33%) showed more consonant deletion than other three groups (2.22, 1.11, and 2.22%). Based on the detailed observation of the strategy all groups used for initial and final consonant clusters, after summing up the values of ICC and FCC for each group, we found the dominant strategy for simplification for all groups was consonant deletion [TW: 30.22% (2.22 + 28), HK: 59.33% (13.33 + 46), BJ: 14.11%, (1.11 + 13), and NS: 41.22% (2.22 + 39)], but not epenthesis; the TW [11.11% (1.11 + 10)) and BJ (9% (0 + 9)]) groups used epenthesis more often than the HK group (0%).

The consonants deleted most often by all the groups were “t” and “d” in final position of accent and second, although this tendency was observed more often in the NS group. Instead of final /t/ air burst, /t/ became unreleased and merged with /n/. These consonants form a natural class owing to the same phonological feature [+coronal] (Chen and Chung 2008). However, the Chinese groups also deleted consonants with grammatical functions not deleted by the NS group, for example, plural /s/ and past participle /t/ in the final consonant clusters in accents and changed.

In fact, some deletion cases cannot be counted as errors. These instances of deletion (e.g., nt and nd) have long been observed in English phonology and have been confirmed in Chen and Chung’s (2008) study. In the present study, the native speakers demonstrated similar rates of deletion as the non-native learners, even though some instances differed in nature from those of the NNSs. This result challenges the findings of Weinberger (1994), Major (1994), Lin (2001), and Chen and Chung (2004), in which the deletion or insertion strategies for consonant cluster simplification were construed as learners’ errors. In these studies, the speech samples from native speakers were not used as the norm.

As expected, there was no epenthesis in the NS group. The performance of the HK group was similar to that of the NS group. However, the TW and BJ groups used more /s/ epenthesis (e.g., mustɔ; changeda) than the HK group. In Heyer’s (1986) study on Chinese speakers’ production of English final obstruents, the learners’ higher percentage of epenthesis to monosyllabic words than to polysyllabic words was construed as their preference for “bi-syllabicity.” The results also confirm the assumption that the CV syllable is the least marked, or the optimal syllable type (Carlisle 1994; Chomsky and Halle 1968; Hooper 1976). The reason why the HK group did not demonstrate a similar preference to the other two Chinese groups may be due to their greater exposure to the English language in Hong Kong.

Speech Rate

Table 1 presents the means, standard deviations, maximum values, and minimum values of VI of speech rate for the four groups. It was found that the BJ group performed the closest to native speakers for the duration of entire sentences. The ANOVA showed that the differences between these four groups were statistically significant ($F_{3,36} = 17.770, p < 0.001$). The post-hoc LSD test showed a significant difference between NS and the other three groups. Among the three Chinese groups, there is a significant difference between HK and BJ groups.
## Table 4  Frequency of deletion, insertion in initial consonants among the four groups

<table>
<thead>
<tr>
<th></th>
<th>Speak (%)</th>
<th>Guage (%)</th>
<th>Spea (%)</th>
<th>Glish (%)</th>
<th>From (%)</th>
<th>France (%)</th>
<th>French (%)</th>
<th>Pro (%)</th>
<th>Gress (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial position: deletion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.22</td>
</tr>
<tr>
<td>HK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>30</td>
<td>20</td>
<td>40</td>
<td>0</td>
<td>13.33</td>
</tr>
<tr>
<td>BJ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1.11</td>
</tr>
<tr>
<td>NS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>2.22</td>
</tr>
<tr>
<td><strong>Initial position: epenthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.11</td>
</tr>
<tr>
<td>HK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BJ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Final position: deletion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>20</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>HK</td>
<td>40</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>46</td>
</tr>
<tr>
<td>BJ</td>
<td>30</td>
<td>50</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>NS</td>
<td>60</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>90</td>
<td>39</td>
</tr>
<tr>
<td><strong>Final position: epenthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>10</td>
<td>0</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>HK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BJ</td>
<td>10</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>NS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
It must be noted that L2 speech rate is typically slower than native speech. Derwing and Munro (1997) indicated that the speech rate of Mandarin Chinese speakers in an English utterance is slower than that of native English speakers. A possible reason for this is that L2 speakers tend to not reduce function words, such as “the” or “and,” as much as native speakers (Aoyama and Guion 2007). In the current study, even though BJ speakers produced a fast speech rate, they still were not able to show reduced function words like native speakers did.

The Comprehension Study

Methods

Participants

To assess the intelligibility of Chinese accented speech, 24 listeners (raters) from 6 language backgrounds were invited and divided into 6 groups: the native English speaker group (American \(n = 2\), Canadian \(n = 1\), Australian \(n = 1\)), ESL speaker group (Indian \(n = 1\), Finnish \(n = 1\), Czechs \(n = 1\), Nigerian \(n = 1\)), EFL speaker group (Japanese \(n = 1\), Korean \(n = 3\)), Hong Kong group (Cantonese \(n = 4\)), Taiwan group (Taiwan Mandarin \(n = 4\)), and Beijing group (Beijing Mandarin \(n = 4\)). It must be noted that the four participants categorized as the ESL speaker group rather than the EFL speaker group were because India and Nigeria are countries where English is important for historical reasons and plays a central role in the nations’ institutions as an official language. Further, according to the report in 2013 of EF English Proficiency Index (EF EPI), Finland’s English level has been ranked as very high proficiency and that of the Czech Republic has been ranked as moderate proficiency, followed by India and Hong Kong. Based on the previous literature and the actual English proficiency of the listeners, it is reasonable to include the two listeners from Finland and the Czech Republic into the ESL speaker group. These listeners were either local university students or exchange students in Hong Kong universities. These participants were categorized as being at an intermediate-high level of English proficiency. Moreover, all the listeners were required to have a basic understanding of phonetics.

Procedures

A total of four typical speakers (one NS and three Chinese speakers from TW, BJ, and HK) from the forty participants whose speech samples were collected in the acoustic study were chosen as representatives for the four groups. When conducting the intelligibility task, the 24 listeners were required to complete 4 sets of dictation passages with 78 embedded blanks that were selected from the read-aloud transcripts. While the listeners listened, a blank-filled sheet, as shown in “Appendix 2”, was presented on a screen with some key words missing. Participants were required to fill in the blanks after listening to the recordings. Listeners could play the sound clips by themselves, but each piece of the recording was played only twice. To eliminate the effect of memory retention from listening to the similar-content recordings, at the end of each set of transcription, the participants were asked to do a 15-min accent-rating task for the purpose of distracting them. Thereafter, the error rates and error types of the dictation task were analyzed.
Table 5  Accuracy rates (Accu. R) of the dictation task

<table>
<thead>
<tr>
<th>Listener group</th>
<th>TW</th>
<th>HK</th>
<th>BJ</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS</td>
<td>82</td>
<td>89</td>
<td>85</td>
<td>98</td>
</tr>
<tr>
<td>EFL</td>
<td>83</td>
<td>87</td>
<td>78</td>
<td>90</td>
</tr>
<tr>
<td>ESL</td>
<td>90</td>
<td>93</td>
<td>82</td>
<td>97</td>
</tr>
<tr>
<td>TW</td>
<td>91</td>
<td>88</td>
<td>86</td>
<td>97</td>
</tr>
<tr>
<td>HK</td>
<td>93</td>
<td>96</td>
<td>87</td>
<td>98</td>
</tr>
<tr>
<td>BJ</td>
<td>90</td>
<td>89</td>
<td>89</td>
<td>97</td>
</tr>
<tr>
<td>Average</td>
<td>88</td>
<td>90</td>
<td>84</td>
<td>96</td>
</tr>
</tbody>
</table>

Results

Accuracy Rates

Table 5 illustrates the accuracy rates for the dictation assessment for six listener groups. All the listener groups achieved at least 78% intelligibility of the three Chinese accents, with the NS group achieving the highest accuracy rate for all six groups of listeners, from 90 to 98%. The intelligibility of the TW group ranged from 82 to 93%, that of the HK group ranged from 87 to 96%, and that of the BJ ranged from 78 to 89%. On average, the NS group is significantly better than the other three groups with an accuracy rate of 96%, while the HK group with an accuracy rate of 90% is significantly better than the BJ group with 84%. The ANOVA showed that the differences between these four groups were statistically significant ($F_{3,20} = 10.269, p < 0.001$). The post-hoc LSD test helps to conduct multiple comparisons of the intelligibility ratings among four groups. When rating the TW speaker group, TW ($p = 0.044$) and HK ($p = 0.016$) listeners’ dictation scores were found to be significantly better than the other four listener groups. When rating the HK and BJ speaker groups, no significant differences have been found among the six groups of listeners’ dictation scores ($p > 0.05$). It was commonly believed that listeners exhibit an intelligibility benefit for speech produced in their own accents. As found in this study, TW listeners demonstrated an intelligibility benefit for their own TW accent. However, we also found that—although to different degrees—a diverse group of listeners may share a response to L2 speech, as found in HK and BJ speaker groups.

To a great extent, this study echoes Munro, Derwing, and Morton’s (2006, p 111) findings:

The listeners didn’t consistently exhibit an intelligibility benefit for speech produced in their own accent. Properties of the speech itself are a more potent factor in determining how L2 speech is perceived, even when the listeners are from diverse language backgrounds.

Compared to the results of Chen (2011), with intelligibility rates of at least 70%, this study shows that the intelligibility reached at least 78%, on average 87%. One of the reasons for this is language proficiency. In this study, the three groups of Chinese speakers are
university students in Hong Kong, with language proficiency being much higher than those studying in community college, as in Chen (2011). Among listener groups, for example, all the ESL listeners in this study are exchange students in university, whereas the majority of the ESL listeners in Chen (2011) are Pilipino domestic helpers in Hong Kong, whose language proficiency levels were not well controlled. All these factors may make the intelligibility rates higher.

Error Types

An analysis on the error types of three Chinese speaker groups was conducted to provide further details for the reasons that may impede intelligibility. The most frequent errors of each group were calculated. The data only shows the frequency of error equal to or more than 50% judged by each group of listeners. Table 6 shows the most frequent error types and examples from the three groups of Chinese speakers’ speech. In TW speech, “word-stress shift” (e.g., accent) was found to make sounds unintelligible. Furthermore, the TW speaker’s replacement of can’t by can led listeners to be unable to decode the word correctly.

Other features such as “Insertion of schwa /ə/” (e.g., old: insertion of schwa /ə/ at the end), “Consonant replacement” (e.g., French: /ɛ̃/ → /s/), “Vowel replacement” (lots: /ɔ̃/ → /ʌ/), “Consonant cluster simplification” (e.g., exposure /ˈɛkspoʊsər/), and “mispronunciation” (e.g., afraid, theories, and exposure) caused some difficulty for listeners in decoding the words.

Word-stress shifts also occurred with HK speakers (e.g., accent), and this caused an intelligibility problem for the majority of the raters. Other common unintelligible words like linguists, theories, and recognize in HK speech have different features, which are given below:

a. linguists was produced without the final consonants “ts”.

b. theories was produced as “feories” with consonant substitution /θ/ for /θ/.

c. recognize was produced as “reCOgnize” (/ˈriːkɒnɪz/) with stress shift and deletion of consonant/g/.

Based on this result, we can infer that stress shift, deletion of final consonant(s), and replacement of /θ/ are possible reasons that impeded the intelligibility of the speech of the HK speaker, while other features such as “insertion of final consonants” (e.g., after foreign and progress) and “monophthongization” (e.g., /eɪ/ in native) seem to not have influenced intelligibility much.

For BJ speakers, except for complete mispronunciation (e.g., individuals, theories, and manage), the most obvious pronunciation features are stress shift, deletion of final consonant(s), monophthongization, consonant replacement, and vowel replacement. Among these features, vowel replacement (e.g., Spanish: /æ/ → /æ/ & /ə/ → /ə/) and monophthongization (e.g., applied: /æ/ → /æ/) indeed led listeners to be unable to decode the words.

Among the words most commonly mispronounced from all three Chinese speaker groups are accent(s), can’t, linguists, and theories, thereby suggesting that irrespective of what accents the listeners heard, they could not understand the words or spell them out. A majority of raters were not able to distinguish between “can” and “can’t,” even for native English speakers. Other problematic words were mostly multisyllabic words that were not easy to pronounce or perceive. There are two possible reasons for the errors in the dictation task. First, speakers’ accents were so strong that even native speakers were not able to ascertain the word (e.g., ACcent → acCENT). Second, raters’ insufficiency in listening skills or vocabulary can also lead to these errors. For example, as their command over vocabulary was insufficient,
Table 6  The top three errors types/examples in speech: TW, HK and BJ accents

<table>
<thead>
<tr>
<th>Error types</th>
<th>TW examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress shift</td>
<td>1. ….because of your ‘foreign accent (stress shift as acCENT).’</td>
</tr>
<tr>
<td></td>
<td>2. You also need accurate (stress shift and mispronunciation as ‘acCUrant’) information about the English sound system</td>
</tr>
<tr>
<td>Deletion of final consonant(s)</td>
<td>3 They may also be able (stress shift as aBLE) to identify Spanish or Arabic (stress shift as ‘aRAbic’) speakers over the telephone</td>
</tr>
<tr>
<td>Insertion of schwa /a/</td>
<td>1. But old (insertion of schwa /a/ at the end) habits won’t change without a lot of hard work.</td>
</tr>
<tr>
<td></td>
<td>2. Appears to be a combination of hard work, a good ear, and a strong (insertion of schwa /a/ at the end) desire …</td>
</tr>
<tr>
<td></td>
<td>3. Several theories (/ˈθiəri-/insertion of schwa /a/) address this issue.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error types</th>
<th>HK examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress shift</td>
<td>1. …people may have noticed (stress shift as noTICE)</td>
</tr>
<tr>
<td></td>
<td>2. Why do people usually have an accent (stress shift as acCENT)</td>
</tr>
<tr>
<td>Deletion of final consonant(s)</td>
<td>3 accurate (stress shift as aCUrate)</td>
</tr>
<tr>
<td></td>
<td>4. Recognize (stress shift ‘reCONize’/rɪˈkɒnɪz/)</td>
</tr>
<tr>
<td>Insertion of final consonants</td>
<td>1. …. because of your ‘foreign accent (deletion of final consonant cluster ‘nd’).</td>
</tr>
<tr>
<td></td>
<td>2. …without an accent (deletion of /k/), but applied linguists (deletion of final consonants) have reported (deletion of final consonant /l/) cases (deletion of final consonant /s/) of…</td>
</tr>
<tr>
<td></td>
<td>3. Influences (deletion of final consonant/s/)</td>
</tr>
<tr>
<td></td>
<td>4. But old (deletion of final consonant cluster /ld/) habits</td>
</tr>
<tr>
<td></td>
<td>5. …or will you just (deletion of final consonant /t/ give up?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error types</th>
<th>BJ examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress shift</td>
<td>1. address (stress shift as ADdress) this issue</td>
</tr>
<tr>
<td>Deletion of final consonant(s)</td>
<td>2. You also need accurate (stress shift as ‘accuRATE’)</td>
</tr>
<tr>
<td></td>
<td>3. Will you manage to make progress (stress shift as proGRESS)</td>
</tr>
<tr>
<td>Insertion of final consonants</td>
<td>1. …have noticed (deletion of final consonant /t/)…</td>
</tr>
<tr>
<td></td>
<td>2. Linguists (deletion of final consonant)</td>
</tr>
<tr>
<td>Monophthongization</td>
<td>3. your first language influences (deletion of final consonant /s/) your pronunciation</td>
</tr>
<tr>
<td></td>
<td>4. Does this mean that accents can’t be changed (deletion of final consonant cluster /ld/)?</td>
</tr>
<tr>
<td></td>
<td>1. If English is not your native (/eɪ/ → /e/) language,</td>
</tr>
<tr>
<td></td>
<td>2. A strong desire to sound (/aʊ/ → /a/) like a native (/eɪ/ → /e/) speaker</td>
</tr>
<tr>
<td></td>
<td>3. applied (/əʊ/ → /ə/) linguists</td>
</tr>
</tbody>
</table>
they were not able to recognize the words; even if they got the word, they could not spell out the words properly under time constraints (e.g., *linguists* and *theories*).

**Discussion and Conclusions**

Most of the past research on L2 timing patterns, foreign accents, and intelligibility assumed that listeners perceive speech holistically. There is no framework to explain what discrete components compose foreign accents or which factors weigh most heavily on listeners’ reactions. This study acoustically investigated the seven timing patterns of stressed/unstressed vowel reduction, linking duration, pause duration and consonant cluster duration in initial and final positions and speech rate produced by Chinese learners from Hong Kong, Taiwan, and Beijing, China. The question of whether the seven acoustic timing variables in Chinese learners’ output are similar to or deviate from those of native speakers of English (the norm) was explored in detail. By comparing the phonologies of three Chinese groups and the NS group, and the perceptions of NSs and NNSs, the major findings of both the production and comprehension studies are summarized below.

**The Production Study**

First, an analysis of the seven acoustic timing variables discussed in this study suggests that three groups of Chinese learners display specific speech patterns that deviate from those of native speakers of English. Among the three Chinese groups, the BJ group performed significantly closer to the NS group in speed rate and pause, while the HK group performed significantly better in unstressed syllable duration. Apart from this, few significant differences could be found in other timing variables among the three Chinese groups.

In fact, the findings suggest that Mandarin speakers from Beijing and Taiwan may fail to differentiate the duration between stressed and unstressed syllables. Vowel reduction occurred much more rarely amongst Chinese learners from TW and BJ; however, the HK group, who have had much English exposure during the past decades under the sovereignty of the United Kingdom, performed much better in unstressed vowels.

This study has demonstrated that a significant disparity exists in syllable duration and syllable-to-syllable variation between American English and English spoken by Chinese English learners by calculating the PVI. The acoustic timing patterns in American English, as expected, can be considered stress-timed, while those in English spoken by Chinese learners (particularly by the BJ group) are more syllable-timed.

Moreover, three Chinese groups of learners performed much longer linking duration and pause duration than native English speakers. In the absence of C-V linking, the word boundaries were maintained through the insertion of short pauses and glottal stops, which lengthened linking duration. The tendency to keep vowels intact may possibly be related to a speaker’s concern with intelligibility. The dominant strategy for speech simplification across the four groups (i.e., TW, HK, BJ, and NS) is consonant deletion, but the TW speakers commonly resorted to vowel epenthesis instead of consonant deletion.

**The Comprehension Study**

The results of the comprehension study revealed that all six listener groups achieved at least 78% intelligibility of the dictation task, with the NS accent achieving the highest rating, followed by HK, TW, and BJ accents. The NS, EFL, ESL, and HK listener groups yielded
higher accuracy rates when listening to HK speech, the TW listener group generated the highest rate for TW speech, and BJ listeners gave almost the same ratings for all three Chinese groups. All the listeners perceived inappropriate word-stress shift and consonant(s) deletion as most unintelligible in comparing the three Chinese foreign-accented speeches.

Is there a shared-L1 effect? Factors such as the properties of the speech itself might outweigh the effect of shared L1 familiarity, as explained in Munro et al. (2006) and confirmed in the current study. When evaluating foreign-accented speech, listeners are affected by the characteristics of the speech itself, their own language backgrounds, and their familiarities with speech varieties. This study considered different dialectal backgrounds of speakers as well as different language backgrounds of listeners in the process of production and perception of speech timing patterns. The possibility of a shared-L1 intelligibility advantage has been explored in this study. There is a theoretical foundation for a shared-L1 effect based on the principle that L2 accents are primarily characterized by transfer from the L1, and that listeners who share a speaker’s L1 will have an intimate familiarity with the phonological patterns of that speaker’s L2 accent. Moreover, in this study, the TW listener group’s rating echoed the findings of Flowerdew (1994) that L2 listeners find speakers who share their L1 most comprehensible, given their greater familiarity with that variety.

However, the HK speaker with better phonological production in this study was not only granted higher ratings from NS, EFL, and ESL listener groups but also from their own shared L1 HK listener group. The BJ speaker, with the poorest performance of phonological production, did not receive the highest intelligibility rating from their shared-L1 BJ listener group. The BJ listener group gave almost the same ratings to all three Chinese groups, which demonstrated that a shared-L1 background may have little or no impact on intelligibility. Therefore, the speech properties might prevail over the shared L1 effect.

All these findings have contributed to our understanding of the acquisition of ESL/EFL timing patterns and particularly to the manner in which ESL/EFL language learners develop new timing patterns and how such accented speech is used in communication with people from different parts of the world.

This study also makes several other contributions. First, this study took into consideration timing patterns over whole sentences. Prior studies have usually compared stressed and unstressed syllables using isolated words or highly selected syllables from utterances. This study examined complete sentences, thereby providing us a more realistic view of timing patterns.

This study also used a larger sample size (40 subjects) than most other studies. Due to the time required for measurement and analysis, it is typical for acoustic studies to have limited samples (e.g., usually under 10 subjects). Since larger individual variations are commonly observed in the study of inter-language, the reliability of results based on a larger sample, as in this study, should be higher.

Further, this study expanded the use of VI to make cross-sectional comparisons. VI took into account the characteristics of stressed/unstressed syllable durations, linking duration, pausing duration, and consonant cluster durations at initial and final positions. Thus, it captures the differences between the average performances of native English speakers and those of the other three groups of Chinese learners. An analysis of PVI further captured the rhythmic typology between stress-based and syllable-based varieties of English.

In conclusion, the relationship between acoustic timing patterns and perceptual judgments has been explored in this study. Word stress and consonant cluster deletions are two main features for Chinese learners of English that contribute to international intelligibility. Listener judgments of such variables as measures of intelligibility have helped us understand the manner in which people respond to particular accented speakers and to ascertain the difficulty
they encounter in understanding a particular phonological variable. This study could benefit ESL/EFL teachers by increasing their sensitivity to the timing difficulties experienced by Chinese learners of English. They should re-examine the ways in which timing features are introduced in current classroom practices and further develop appropriate teaching materials for Chinese learners. The establishment of the relationship between acoustic ratings and perceptual judgments in learners of English could also stimulate computer-assisted teaching, learning, and assessment of aspects of English timing patterns.

Acknowledgments This work was supported by the Internal Research Grant of the Hong Kong Institute of Education.

Appendix 1: The Major Pauses (MP) Among the Four Groups of Speakers

NS Group: Total no. of MP: 3
1. Why do people usually have an accent when they speak a second language?
2. Most native speakers of English can, for example, P(8) recognize people from France by their French accents
3. But old habits won’t change without a lot of hard work, P(4) will they?
4. Does this mean that accents can’t be changed?
5. Will you manage to make progress, P(4) or will you just give up?

TW Group: Total no. of MP: 31
1. Why do people usually have an accent P(10) when they speak P(4) a P(3) second language?
2. Most P(4) native speakers of English P(3) can, P(8) for example, P(10) recognize P(3) people P(3) from P(5) France P(9) by their P(3)
   French P(6) accents
3. But old habits P(3)bits P(6)won’t P(4)change P(4)without a lot of P(3) hard work, P(9) will they?
4. Does this mean P(4) that P(5) accents P(6) can’t P(5)be P(8)changed?
5. Will you P(5) manage P(6)to make P(10) progress, P(9) or will you P(3) just P(4) give P(3) up?

HK Group: Total no. of MP: 25
1. Why do people usually have P(3) an P(4)accent P(7) when they speak P(3)a second language?
2. Most P(7) native P(3) speakers P(4) of English can P(9), for example,
   P(10) recognize P(4) people from France P(8) by P(3) their
   French P(7) accents
3. But P(6) old habits P(3) won’t change P(5) without a lot of P(3) hard work, P(10) will they?
4. Does P(3) this mean P(3) that P(9) accents P(5) can’t be changed?
5. Will you manage to make P(9) progress, P(9) or will you just give up?

BJ Group: Total no. of MP: 13
1. Why do people usually have an accent P(5) when they speak a second language?
2. Most native speakers of English can, P(10) for example, P(7)
   recognize people from France P(5)by their French P(3) accents
3. But P(3) old habits P(4) won’t change P(5) without a lot of hard work, P(9) will they?
4. Does this mean P(4) that P(3) accents can’t be changed?
5. Will you manage to make progress, P(8) or will you P(4) just give up?

1. Major pause (MP) means the pause produced by 3 or more than 3 speakers in one group
2. The no. in the bracket indicates the no. of subjects who had this pause

Appendix 2. Dictation Passage

(From Celce-Murcia et al.’s (2001) Teaching Pronunciation, p. 398)
If English is not your native language, people may have noticed that you come from another country because of your “foreign accent.” Why do people usually have an accent
when they speak a second language? Several theories address this issue. Many people believe that only young children can learn a second language without an accent, but applied linguists have reported cases of older individuals who have mastered a second language without an accent. Another common belief is that your first language influences your pronunciation in a second language. Most native speakers of English can, for example, recognize people from France by their French accents. They may also be able to identify Spanish or Arabic speakers over the telephone, just by listening carefully to their pronunciation. Does this mean that accents can’t be changed? Not at all!! But old habits won’t change without a lot of hard work, will they? In the end, the path to learning to speak a second language without an accent appears to be a combination for hard work, a good ear, and a strong desire to sound like a native speaker. You also need accurate information about the English sound system and lots of exposure to the spoken language. Will you manage to make progress, or will you just give up? Only time will tell, I’m afraid. Good luck, and don’t forget to work hard!

References


Chinese learners’ acquisition of English word stress and factors affecting stress assignment

Hsueh Chu Chen

The Hong Kong Institute of Education, Hong Kong

ARTICLE INFO
Available online 24 September 2013

Keywords:
Lexical class
Word stress
Syllable structure
Perception and production

ABSTRACT
This study explores Chinese ESL learners’ acquisition of English word stress and compares the factors affecting their stress assignment with Guion et al.’s (2003) findings of three factors (syllable structure, lexical class, and phonologically similar words) affecting English speakers’ stress placement. Twenty Hong Kong advanced learners produced and perceived 40 real words and 40 pseudowords of varying syllable structures in noun and verb sentence frames. They also indicated words they considered to be phonologically similar to the pseudoword. The results show that the participants performed well in the tasks with real words, but there were asymmetrical abilities in the tasks involving pseudowords. Regarding the factors affecting the stress assignment, different from the findings by Guion et al., stress assignment was not significantly influenced by syllable structure or the stress patterns of phonologically similar real words. Only lexical class had an effect on main stress assignment.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

English was established as the official language of Hong Kong as early as the colonial period (Setter, Chan, & Wong, 2010). Hong Kong’s government has been trying to remove the English language from its prevailing position since the handover from the British government by imposing a new language policy called ‘Biliteracy and Trilingualism’ (Li, 2009, p. 72) to drive a balanced development of English, Mandarin, and Cantonese. However, English remains the gateway to higher education as well as to a better job and social position. Both parents and students believe that attending English medium schools will bring more opportunities for employment and further studies. Most elite secondary school students choose to be taught partly or entirely in English, even though they still expose themselves to the Chinese language outside of the classroom. Given the special status of English in Hong Kong, it would be interesting to investigate whether it is possible that differing patterns of stress assignment may exist for Chinese speakers immersed in both their native language and in English (i.e., Hong Kong), and native English speakers immersed in the English language environment only (i.e., the United States (U.S.)).

Guion, Clark, Harada, and Wayl (2003) found that three factors affect English speakers’ stress placement on bi-syllabic non-words: syllable structure (e.g., heavy syllable, vowel length), lexical class (noun vs. verb counterparts), and stress patterns of phonologically similar words. In this study, the influence of syllable structure, lexical stress, and the stress patterns of known phonologically similar words on native Hong Kong Chinese speakers’ acquisition of English stress patterns was investigated. Whether the factors Guion et al. (2003) found to affect the stress assignment differ significantly between L1 Chinese in Hong Kong and L1 English speakers was also examined.
The research questions were raised as follows:

1. What is the relationship between Hong Kong ESL learners’ perception and production of English word stress patterns?
2. What are the factors affecting Hong Kong ESL learners’ stress assignment for English pseudowords?

2. Review of the literature

According to Roach (2009), English word stress is highly complex because stress patterns for English syllabic structures and word affixes are unpredictable, although a few generalizations can be made (e.g., Chomsky & Halle, 1968; Fudge, 1984). Cruttenden (2008) also pointed out that attempts to formulate the rules of English lexical stress are “bedeviled by the existence of large numbers of exceptions to almost any rule,” implying that we cannot declare any absolute rules for the English stress system (2008, p. 238). As a result, many theorists and practitioners have suggested that students can learn the stress of individual words as part of the acquisition process for each new lexical item (Howard, 2010). Therefore, it is fundamentally difficult for learners to deal with the placement of lexical stress in English.

Guion et al. (2003) investigated possible influences on the acquisition of English stress patterns and found that three factors affect English speakers’ stress placement on bi-syllabic non-words: syllabic structure, lexical class, and stress patterns of phonologically similar words. Later, Guion and her colleagues (e.g., Guion, 2005; Guion, Harada, & Clark, 2004) conducted a series of experiments to validate their findings by testing learners of English as a second language (ESL) from different language backgrounds (e.g., Spanish and Korean speakers). Within their findings, all three factors were shown to significantly influence stress assignment patterns on non-words among native and nonnative English speakers. However, syllabic structure played a reduced role for late Spanish–English bilinguals. Furthermore, both syllabic structure and lexical class played a reduced role for late Korean–English bilinguals, while stress patterns of phonologically similar known words were good predictors of word stress for both late bilingual groups.

Several studies have suggested that ESL learners from a non-stressed language background, such as Chinese and Japanese speakers, often experience difficulties in English lexical stress acquisition and may not ever possess a knowledge base of English word stress in the same way as native speakers (Archibald, 1997; Peperkamp & Dupoux, 2002). Archibald (1997) conducted a research study on the acquisition of English stress by speakers of non-stressed languages, including Chinese and Japanese speakers. His research indicated that neither syllabic structure nor grammatical category significantly affect Chinese English speakers’ acquisition of English stress assignment. “They seemed to be treating stress as a purely lexical phenomenon—something that has to be memorized as part of the phonological representation of a word” (Archibald, 1997, p. 175).

This notion was reinforced by Wayland, Landfair, Li, and Guion (2006), who investigated the influence of syllabic structure, lexical class, and the stress patterns of known words on the acquisition of the English stress system in 10 native Thai speakers. The results suggested that syllables containing long vowels were more likely to be assigned stress than syllables containing short vowels and that nouns received initial stress more often than verbs. Thai participants’ patterns of stress assignment on non-words were significantly influenced by the stress patterns of phonologically similar real words.

As the aforementioned studies by Guion and her colleagues were mostly conducted in English-speaking countries with English-rich environments, the participants’ exposure to English was quite likely to have helped these bilinguals to internalize stress rules via the clues of syllabic structure, lexical class, and stress patterns of phonologically similar words. It is therefore doubtful that these results can be applied to those who are learning English word stress in second or foreign language environments such as China.

The pronunciation problems of Chinese speakers in Hong Kong are quite prominent and distinctive, and they have captured the attention of linguists within the second language learning circle. Various segmental problems such as initial and final consonant cluster deletion or simplification and the merging of /n/ and /l/ are frequently reported. Chinese and English represent two contrasting prosodic types: tonal languages and stressed languages. English has a system of culminate word stress, but Chinese, a tonal language, has no system of word stress; rather, it has a system of four distinctive tones in which pitch is used to distinguish individual lexical items. As a result, ESL learners in China experience considerable difficulties in correctly perceiving and producing English stress.

Archibald (1997) and Wayland et al. (2006) also suggested that the stress patterns of phonologically similar words contribute significantly to Chinese and Thai speakers’ stress placement, and they minimized the roles of syllabic structure and grammatical category in English word stress assignment. Their proposed explanation of lexical stress assignment attempted to use the idea of an analogy, developed from Skousen’s Analogue Model of Language (Skousen, 1989) to explain how speakers determine linguistic behaviors such as stress placement based on the relevant examples stored in their minds. An analogy is a mechanism that can extend the knowledge of lexical phonology to novel forms. As an attempt to study phonology in terms of analogy, Bybee (1999, 2001) provided sound support for the analogy hypothesis proposed by Archibald (1997). Bybee suggested that phonology and the lexicon are inextricably linked. This serves as the basis of a likely analogical effect; in other words, Bybee claimed that speakers store phonological patterns like stress lexically in their mental lexicon. Phonologically similar items are interrelated and interconnected, and if stored together, can help speakers quickly process and search for the analogs.

Based on the previous literature, it seems that students can apply analogies as a strategy to deal with the unpredictable English stress in novel words. The ability to generalize is part of linguistic competence, which means that ESL learners can
only apply this strategy after possessing a certain level of ability in the target language. It is doubtful that ESL learners in the ESL/EFL context, like Hong Kong, can easily attain this high level of competence and build a sufficient bank of vocabulary in their second language to make the analogy immediately upon encountering a new word.

The extension of stress patterns from known real words to non-words was still a strong predictor for stress placement among late Korean–English and Spanish–English bilinguals in Guion and her colleagues’ work (Guion, 2005; Guion et al., 2004). The results from the adult learners in these studies, combined with the results from the adult Thai–English bilinguals in Wayland et al. (2006), suggested that adult learners of English may rely more heavily on word-by-word learning of stress patterns and are less likely to abstract generalities about stress placement by syllabic structure and lexical class. However, as Wayland et al. mentioned, such an interpretation would run counter to the findings by Kelly and Bock (1988) and Arcuilli and Cupples (2003, 2004), in which late learners demonstrated knowledge of English stress patterns based on the lexical classes noun and verb” (p. 298). It therefore remains controversial whether it is difficult for ESL learners to acquire knowledge of lexical classes and whether this plays a limited role in stress assignment.

It is of great interest to examine the knowledge of English main stress patterns possessed by Hong Kong advanced ESL learners, who have a non-stressed first language (L1) and live in an English-as-a-second-language (L2) environment. My comparative study explores the relationship between their perception and production of English word stress patterns and the factors affecting stress assignment for English words. A number of previous studies have investigated either the perception or production of L2 stress. The most thorough was conducted by Altmann (2006), who investigated the perception and production of L2 stress among seven distinct L1 groups: Arabic, Chinese, French, Japanese, Korean, Spanish, Turkish, and native English speakers. All of the participants took part in both perception and production experiments. The results indicated that, on one hand, learners with predictable stress in their L1 (i.e., Arabic, Turkish, French) had problems perceiving the location of stress, but they performed better in production. On the other hand, learners without word-level stress in their L1 (i.e., Chinese, Japanese, Korean) or with unpredictable L1 stress (Spanish) showed almost perfect perception scores, but their production was quite different from that of the control group. Thus, “Good perceptibility of the location of stress does not imply good production of stress, and bad perception does not imply bad production” (Altmann, 2006, p. 162). However, Altmann did not give a detailed illustration of the factors influencing stress placement.

Concerning the stress acquisition of Chinese speakers of English, most research has concentrated on the stress patterns of English produced by Chinese speakers or on the acoustic characteristics of English lexical stress. For example, Chen (2012) investigated the acquisition of English word stress by ESL learners in Hong Kong. Ten students were given an online program as a supplement to a 30-h English course, and were requested to practice on their home computers. Their development in spoken English was compared with a control group of ten students who did not use the online resource. Results indicated that the online pronunciation tutoring was fairly effective in improving the ESL learners’ ability to produce and perceive correctly different stress patterns in words, phrases, and sentences. Practice with the program was beneficial to those students who began the course with a strong foreign accent, but it was of limited value to students who began the course with better pronunciation. Zhang, Nissen, and Francis (2008) compared the use of frequency (F0), duration, intensity, and vowel quality in the production of lexical stress contrasts by 10 Mandarin and 10 native English speakers. The stress patterns produced by Mandarin speakers differed significantly from L1 English speakers, although they did use all four acoustic correlates to distinguish stressed from unstressed syllables.

At present, no systematic studies have investigated the influence of the three factors—syllabic structure, lexical class, and stress patterns of phonologically similar words—on Chinese speakers’ acquisition of English stress principles. A systematic study of the factors that influence Chinese English learners’ acquisition of stress, therefore, will fill the gap in this field of research.

3. Method

3.1. Participants

Twenty advanced ESL learners from Hong Kong, China, were paid to participate in the experiment. They were native speakers of Cantonese between 19 and 25 years of age and had never studied abroad in an English-speaking country. They were English major undergraduates or postgraduates who had studied English since first grade and achieved at least a 6.5 score on the IELTS or an equivalent score on another English proficiency exam (e.g., TOEFL CBT 225/PBT 560/IBT 85). None of them reported being diagnosed with any hearing disorders. The reason English majors were recruited was that they had taken an Introduction to English Phonetics and Phonology course and their basic knowledge on supra-segmental phonology was ensured, which provided the prerequisite for the experiment.

3.2. Materials and procedures

There were five tasks in total for the whole experiment. Among these, three tasks were perception tasks and the other two were production tasks. All of the tasks were programmed into an online platform and conducted in a language lab at a university in Hong Kong. The tasks were sequenced in order from production to perception; that is, Task 1: Real word production task was followed by Task 2: Pseudo-word production task, Task 3: Production of phonologically similar words, Task 4: Pseudoword perception task, and finally Task 5: Real word perception task. The participants wore a head-mounted
microphone and earphone, and their responses were recorded by the voice recorder embedded in the website. They adjusted the presentation volume to a comfortable level before testing began. A briefing session lasting approximately 10 min was conducted before each experiment task to ensure that all participants fully understood the instructions for each task. The materials in the present study were chosen to represent the stimuli used by Guion et al. (2004) as closely as possible. The participants were tested in groups. For the production tasks, the results were coded by two researchers specialized in phonology. Both of them had extensive experience in phonological research and in coding phonological and acoustic phonetic data in several research projects. Ten responses from the data pool were arbitrarily chosen from 10 randomly chosen participants for trait ratings. With close discussion of the perceptual differences in pitch, volume, and duration on the stress placement of each token, they obtained 95% agreement to achieve high inter-rater reliability.

3.3. Perception task

**Task 1: Real word production task.** This task was conducted to test participants’ ability to correctly assign stress to real English words. A list of 36 real English words was adopted from Guion et al. (2004). These words were selected for participants to read during the task. Among the words, 18 had stress patterns consistent with regular English stress patterns (e.g., agent (o’CVCC). In addition, 18 words of the same syllabic structure, but which were not consistent with the regular patterns (e.g., percent), were selected (for more details, please see the appendix in Guion et al., 2004).

The participants were asked to read each of the syllables in the sentence frame, “I said _ _ this time.” The words were presented orthographically in English spelling on the screen. The participants were informed to click the recording button before reading the wordlist. The words were later coded for correct stress placement. Those containing mispronunciations that were not related to stress were not considered.

**Task 2: Pseudoword production task.** In this task, participants were asked to produce two-syllable pseudowords in both noun and verb sentence frames. The effects of lexical class (noun and verb) and syllable structure (four types of non-word stimuli) were investigated. The stimuli in Task 2 consisted of 40 two-syllable pseudowords presented as isolated stressed syllables. There were four types of pseudowords, each with a different syllabic structure, as shown in Table 1.

A native English speaker from the United States was invited to produce the individual syllables making up the pseudowords in the frame, “Now I say _.” Thus, each syllable was produced with sentential pitch accent and stress, creating an equal degree of stress and clarity in the production of all syllables. All productions were recorded on a sound recorder. The syllables were then excised from the frame sentence. They were normalized to 50% peak intensity by using a waveform editor (Cool Edit Pro, 2003), which created individual stimuli of the same intensity. That is, during the process of peak amplitude normalization, the peak amplitude of each waveform was set to 50% of the quantization rate, and all other sampling points were scaled proportionally.

The same speaker also recorded the phrases “I’d like to” and “I’d like a.” The final “to” and “a” were produced in reduced form (i.e., [ts] and [a]). For example,

\[
\begin{align*}
\text{I'd} & \quad \text{like} & \quad \text{to} & \quad \text{be} & \quad \text{tst} \\
\text{I'd} & \quad \text{like} & \quad \text{a} & \quad \text{tst} \\
\end{align*}
\]

Participants heard two separated stressed syllables with the same degree of intensity and stress for each trial. After listening, they were asked to make a single word from the two separated and stressed syllables, determine the stress of the word, and say the word in the carrier phrases that had just been presented both aurally and visually.

Two frame sentences were used to indicate different word classes. The phrase “I’d like to _” will be referred to as the verb frame, and the phrase “I’d like a _” will be referred to as the noun frame. Participants were instructed to reproduce all sounds of the syllables they heard in the same order as presented.

Each of the pseudowords was presented twice, once with the noun frame and once with the verb frame. Presentation of the frame was both aural and visual. After a 500-ms delay, the two stressed, isolated syllables composing the non-word were presented with a 500-ms inter-stimulus interval. Presentation of non-words was aural only; no visual (i.e., no orthographic) representation was given. Participants could replay the trial if they wished. After responding, they pushed a button to continue to the next trial. Two pseudo-randomized, counter-balanced blocks were used, making a total of 80 trials. Each non-word was presented only once in each block. Half of the productions in each block were in the noun frame, while the

---

**Table 1**

Pseudowords used in Tasks 2–4 (retrieved from Guion et al., 2004, p. 51).

<table>
<thead>
<tr>
<th>Type 1 CVV CVCC</th>
<th>Type 2 CVVCVC</th>
<th>Type 3 CVVC</th>
<th>Type 4 CVVCVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>bet tst</td>
<td>de kips</td>
<td>ni let</td>
<td>ni list</td>
</tr>
<tr>
<td>tu kips</td>
<td>ni gept</td>
<td>de sin</td>
<td>de gust</td>
</tr>
<tr>
<td>tai gept</td>
<td>ki minz</td>
<td>se lin</td>
<td>br tens</td>
</tr>
<tr>
<td>psu tist</td>
<td>se tist</td>
<td>bi tes</td>
<td>bi taoos</td>
</tr>
<tr>
<td>gi kips</td>
<td>bi bekt</td>
<td>se get</td>
<td>ki gin</td>
</tr>
<tr>
<td>psu bekt</td>
<td>se bekt</td>
<td>de let</td>
<td>se list</td>
</tr>
<tr>
<td>tu mlnz</td>
<td>de mlnz</td>
<td>ni sin</td>
<td>ni gust</td>
</tr>
<tr>
<td>tai mlnz</td>
<td>ni kips</td>
<td>ki get</td>
<td>ki tens</td>
</tr>
<tr>
<td>bet bekt</td>
<td>ki gept</td>
<td>bi lin</td>
<td>de taoos</td>
</tr>
<tr>
<td>gr gept</td>
<td>bi tist</td>
<td>ki tes</td>
<td>se gin</td>
</tr>
</tbody>
</table>
other half were in the verb frame. There was a short break between the two blocks (Guion et al., 2004). The recordings were coded as having either first or second syllable stress by the researcher and her research assistant.

Task 3: Production of phonologically similar words. The same 40 pseudowords used (10 pseudowords for each syllabic type) in Task 2 were also used as the stimuli in Task 3. They were presented as isolated syllables with a 500-ms inter-stimulus interval. There was a 15-s interval between each stimulus, during which participants were asked to produce any phonologically similar words they could think of in the 15 s. The non-words were presented in 40 trials in a single, randomized block.

The participants were asked to produce any phonologically similar words they could think of in 15 s (at least one word for each stimulus) after hearing two isolated syllables (40 trials in total). The stimuli were only presented aurally. After playing the stimulus for 15 s, the webpage jumped to the next page and automatically played the next word. It was not possible to repeat the trial.

In Task 3, the pseudowords were transcribed orthographically. Only the stress of words with more than one syllable was recorded. Most words produced were disyllabic, but a few were longer. For disyllabic words, the stress was transcribed as either initial syllable stress or final syllable stress. For longer words, the stress was transcribed as initial syllable stress for stress on the initial syllable and as final syllable stress for stress on the second syllable. The placement of stress was determined by the way in which the participants actually produced it. Participants sometimes responded with more than one word per trial. In the majority of cases, the second and third word had the same stress as the first word. If the stress assignment was shown on the second word, then the stress of the first word was recorded (Guion et al., 2004).

Task 4: Pseudoword perception task. In Task 4, 20 pseudowords within the four syllabic structure types (five for each type) were read inside of the two sentence frames, “I'd like a...” and “I'd like to...” In this task, the same American English speaker was used as in the previous tasks. Each pseudoword was read in pairs that varied only in stress placement. The first 20 non-words used in Task 2 (listed in Table 1) were produced with stress on the initial and final syllable in each of the carrier frames, “I'd like a” and “I'd like to,” making a total of 80 sentences. The “a” or “to” of each phrase was produced in a reduced manner as [ə] and [tə].

All participants were asked to listen to the pre-recorded pair sentences that varied only in the stress placement on the pseudoword. In a given trial, the same sentence frame (noun or verb) was presented. Participants were instructed to listen to the two sentences and indicate which one they believed to sound the most like a real English sentence. For example, consider the two representative trials below.

Example Trial 1:
[ aim ə1aik ə bertst] —— the first one will be identified as A
[ aim ə1aik ə berstst] —— the second one will be identified as B

Example Trial 2:
[ aim ə1aik tə mulet] —— the first one will be identified as A
[ aim ə1aik tə melet] —— the second one will be identified as B

Each target pseudoword was presented in two trials, once in a noun frame and once in a verb frame. Two pseudo-randomized, counter-balanced blocks were used, making a total of 80 trials. Each pseudoword was presented only once in each block. Half of the productions in each block were in a noun frame, while the other half were in a verb frame. The order of initial or final syllable stress within a trial was also controlled.

For a given pseudoword, the order of stress presentation remained constant for both the noun frame and verb frame trials. For each of the four pseudoword types, five of the pseudowords were presented with initial stress and then final stress. Conversely, the other five were presented with final stress and then initial stress. There was a break of a few minutes between the two blocks. Participants were given two practice trials using non-test items before the first block. For each trial, the participants were presented with two sentences with a 1500-ms inter-stimulus interval. They responded by pushing a button labeled “A” or “B” to indicate the sentence they thought sounded the most like a real English sentence. Then, after participants made their choices, the webpage would jump to the next page with a 2000-ms delay. No repetition of the trial was allowed (see Guion et al., 2004). The answers from participants were in form of “A” or “B” and were coded as either initial or final stress by the researcher.

Task 5: Real word perception task. This task was designed for investigating whether the perception of stress placement of real English words was consistent with their production. In Task 5, the same 36 real English words used in Task 1 were reproduced by the same native English speaker using the frame sentence, “I said __ this time.” Participants were instructed to listen to the 36 sentences with the target words. After listening to each sentence pair, they were instructed to indicate the stress of the target word that they heard by clicking “1” (if they thought the first syllable was stressed), “2” (if they thought the second syllable was stressed), or “3” (if they thought the third syllable was stressed). No repetition of the trial was allowed in this task. The webpage would jump to the next page with a two-second interval after participants made their choices. Participants’ answers were transcribed as first, second, and third syllable stress.

For Tasks 1 and 5 (real word tasks), the correct percentage of stress assignment was calculated. For Tasks 2, 3, and 4 (pseudoword tasks), the percentage of words that have initial stress was calculated. These analyses followed Guion’s (2003) study.

4. Results

In this section, the results of the production and perception tasks of real words are presented first. The results of the production and perception tasks of pseudowords are then reported. There is also a detailed analysis investigating the effects
of syllable structure and lexical class on main stress. Next, the effects of analogical extension of stress patterns from individual lexical items to pseudoword stimuli are discussed. Finally, analyses of variance on the production and perception data are presented in which the predictive power of syllable structure, lexical class, and stress patterns of phonologically similar words is assessed.

**4.1. Production task and perception task of real words (Tasks 1 and 5)**

The results indicate that participants performed well both in perception (89% correct rate) and production tasks (90.85% correct rate), showing that they have been equipped with appropriate knowledge of English stress. For real word production and perception, a paired samples t-test showed no significant difference between their means (p > .05). Among the words frequently stressed incorrectly, the top two in the perception task were “introduce” (the stress rule for σCV(C), e.g., indicate) and “calendar” (the stress rule for σCVCC(V), e.g., agenda), both of which are not consistent with the rules. In the production task, the top three words frequently stressed incorrectly include “descend” (the stress rule for σCVVC, e.g., aspect, which is not consistent with rules), “prestige,” and “origin” (the stress rule for prestige is σCVVC, and the stress rule for origin is σCVVC, which are perfectly consistent with the rules). The random pattern of misperception and mispronunciation in consistent and inconsistent groups indicates that little influence from explicit stress rules has been found.

As Altmann (2006) argued, L2 speakers are often able to produce real words correctly because they have learned or memorized how to pronounce these words even if they are not aware of where the word is stressed. Therefore, the participants were then assessed on their production and perception tasks of pseudowords to verify their ability to encode and decode unfamiliar words and further find out the possible factors affecting their word stress assignment.

**4.2. Production task and perception task of pseudowords (Tasks 2 and 4)**

The 40 pseudowords were produced and perceived by the 20 participants: once in a noun frame and once in a verb frame. The productions were then coded as having main stress on the first or final syllable. Table 2 shows the means and standard deviations of the proportions of initial syllable stress in the dataset of production and perception. The results indicated that 57.00% of the items in the production task were assigned to the words on the initial syllable, while 68.25% of the items in the perception task were assigned to the initial syllable. The participants performed differently in assigning the initial stress on pseudowords in production (Task 2) and perception tasks (Task 4).

Three possible factors affecting stress assignment for English pseudowords (syllable structure of the word, lexical class, and phonologically similar words) are reported below.

**Factor one: syllabic structure.** The last column of Table 2 shows the means and standard deviations of the proportions of initial syllable stress on four types of syllabic structure in production and perception. For the overall performance (production and perception), the four types of syllabic structure had mean proportions of 64.50%, 56.17%, 63.67%, and 58.67%, respectively. Type 1 (CVVCVC) had a higher mean on overall performance than Types 2 (CVVCV) and 4 (CVVCVC).

Table 2 shows the proportion of stress assignment on four types of syllabic structure between production and perception tasks. In the production task, Types 1 and 3 (60.00% and 62.75%) presented more preference in initial stress assignment than Types 2 and 4 (50.50% and 54.75%). In the perception task, the participants showed a stronger preference for assigning the initial stress than the final stress across the four syllable types (all of them were 65% or above), especially in Type 1: CVVCVC (73.50%). It seemed difficult for Chinese learners to differentiate the stress patterns from different types perceptually, and
initial stress was usually prioritized. This finding was consistent with Archibald's (1997) claim about the tendency among native English speakers to stress the initial syllable for most items.

According to a prediction by Guion et al. (2003), the long vowel in the first syllable of Type 1 attracted more stress than the short vowel in the first syllable of Type 2. The initial syllable of Type 3 attracted more stress than that of Type 4 due to its final long vowel. The different types of final super heavy syllables, such as the final consonant cluster in Type 2 or the long vowel in Type 4, were predicted to condition more final stress.

Comparing the findings in this study with the predictions by Guion et al. (2003), it was found that the participants' performance in the production task was aligned with Guion's predictions, except in Type 4. The long vowel in Type 4, CVVCVC (e.g., ni li:t), was predicted to condition more final stress, but in fact, Table 2 shows that 54.75% of participants assigned initial stress (that is, 45.25% assigned final stress) in Type 4, which was much higher than the 50.50% who assigned initial stress (that is, 49.50% assigned final stress) in Type 2. On the contrary, the participants' performance in the perception task was mostly not aligned with the predictions. Only the long vowel in the first syllable of Type 1 was assigned more initial stress than the short vowel in the first syllable. Table 3 summarizes the asymmetry of initial stress assignment between production and perception.

**Factor two: lexical class.** The first two columns of Table 2 show the means and standard deviations of the proportions of initial syllable stress on word class in the dataset of production and perception. For the overall performance (production and perception), the noun frame received more initial stress than the verb frame. The mean proportions of initial stress in the noun frame were 82.58%, twice as much as the proportions in the verb frame, 38.92%. The difference between two types of lexical class suggested the type of lexical class influenced the assignment of word stress.

![Fig. 1. Mean proportion of initial syllable stress of four syllable types across two word classes (noun and verb) in production (a) and perception (b).](image-url)
Table 4
Means and standard deviations of the proportions of initial syllable stress in the dataset of phonologically similar words.

<table>
<thead>
<tr>
<th>Syllable structure</th>
<th>Total responses</th>
<th>Percentage with initial stress</th>
<th>SD</th>
<th>Most frequently presented words</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVVCVC</td>
<td>136</td>
<td>54.35%</td>
<td>0.19</td>
<td>dentist(6), Baptiste(3), debate(2), protect(4), detect(2), forget(2), desktop(2), kidnap(2), decades(2)</td>
</tr>
<tr>
<td>bet test</td>
<td>143</td>
<td>50.40%</td>
<td>0.12</td>
<td>dentist(6), basin(3), listen(2), beaten(2), nephew(2)</td>
</tr>
<tr>
<td>de kips</td>
<td>154</td>
<td>64.59%</td>
<td>0.16</td>
<td>neglect(8), bitter(2), Baptiste(2), betray(2)</td>
</tr>
<tr>
<td>CVCVC</td>
<td>138</td>
<td>62.01%</td>
<td>0.24</td>
<td>Salute(2), elite(2)</td>
</tr>
<tr>
<td>Total</td>
<td>571</td>
<td>57.84%</td>
<td>0.18</td>
<td></td>
</tr>
</tbody>
</table>

Note: N is the number of items.

Looking closely at the differences between production and perception tasks, as shown in Table 2, the average initial percentage of words in the noun frame was 79.75%, twice as much as the words in the verb frame, 34.25% in production tasks. The same trend was found in perception tasks. The average initial stress percentage of words in the noun frame was 88.25%, while the percentage of words in the verb frame was 48.25%. In both the noun and verb frames, the proportion of initial stress in production tasks (57.00%) was lower than that in perception tasks (68.25%). As expected, however, verbs were predicted to condition more final stress. The participants perceived that almost half of the words in the verb frames (48.25%) were stressed on the first syllable. These findings imply that the participants were weaker in perception than production on stress assignment of verbs.

Fig. 1 shows that in the production task, among the four syllable types, the noun frame received 79.75% initial stress (ranged from 75% to 84.5%), whereas the verb frame received around 33% initial stress (ranged from 22% to 42.5%). In the perception task, the noun frame received around 88.25% initial stress (ranged from 82% to 93%), whereas the verb frame received around 48.25% initial stress (ranged from 43% to 54%). In sum, pseudowords produced in a noun frame were more often produced with initial syllable stress than those in a verb frame regardless of the influence of syllable types.

After comparing the current data with those of native English speakers (Guion et al., 2003), we found some similarities in the stress assignment in production tasks; for example, Type 1 (CVVCVC) was stressed more often in the initial position both in noun frames and verb frames, and words in noun frames received much more initial stress than those in verb frames.

Factor three: phonologically similar words (Task 3). Another question addressed in this study is whether Chinese learners of English will show analogical extension of stress patterns from individual lexical items to pseudoword experimental stimuli and whether such analogical extension is used. For this factor, stress was predicted on the initial or final syllable based on the stress placement of a phonologically similar real word collected from each participant for each of the pseudowords. It was likely that different participants would consider different words to be phonologically similar to the pseudowords. Thus, this task was employed to empirically determine phonologically similar words for each of the participants individually. According to Guion et al. (2003), in the case of native English speakers, 258 responses were recorded out of a possible 400 (40 pseudowords, 10 participants). For the participants in this study, they produced 571 out of a possible 800 responses (40 pseudowords, 20 participants). In the 571 valid responses, the percentage of initial syllable stress was 57.84%, whereas that of final syllable stress was 42.16%. The response percentages of initial syllable stress in each syllable structure were CVVCVC: 54.35%; CVVCVC: 50.40%; CVCVC: 64.59%; CVVCVC: 62.01. In the 229 no responses, the rates in each syllable structure were CVVCVC: 32.00%; CVVCVC: 28.50; CVVCVC: 23.00; CVVCVC: 31.00.

Table 4 summarizes participants’ performance in producing phonologically similar words with four types of syllable structure. The most frequently presented words, total responses, and percentages of initial stress are presented in the table, but the detailed examples are coded in Appendix 1. There were a total of 136 responses in Type 1 syllable structure, and 54.35% of the responses assigned initial stress. Take the pseudoword stimulus/bet test for example: the most frequently presented word was “dentist.” There were a total of 143 responses in Type 2 syllable structure, and 50.40% of the responses assigned initial stress. Take the pseudoword stimulus/de kips for example: the most frequently presented words were “detect, forget, desktop, kidnap, and decades.” It was observed that even though Chinese ESL learners were able to produce as many words as the native English speakers did in Guion et al.’s (2003) study, the relative random patterns of syllable structure could be found. There were 154 responses in Type 3 syllable structure (64.59% of the responses assigned initial stress) and 138 responses in Type 4 syllable structures (62.01% of the responses assigned initial stress). It was interesting to find that in Type 4, CVVCVC, (e.g., the pseudoword stimulus/bt test), a long vowel was in the second syllable and stress was expected to be assigned. However, 62.01% of the responses were assigned initial stress. The random assignment on the word stress in the task was evident.

After conducting Pearson correlation between the stress assignment of pseudowords, as in Task 2, and their phonologically similar counterparts in this task (task 3), the mean scores on the 80 items of Task 2 and the 40 items in Task 3 were computed for the 20 respondents. The mean and SD in Task 2 were 57.00% and 10.71%, respectively, and 58.19% and 14.00% in Task...
3, respectively. The correlation was .45 (p = .045), suggesting a statistically significant correlation between them, and the strength of association was positive and moderate. This implied when the participants pronounced a pseudo- or unfamiliar word, most of them read it as being not totally independent, but only fairly associating it to known words with similar phonological structures.

4.3. Summary of the factors affecting stress placement

There were two kinds of data, one was production and perception (containing 120 items), and the other was phonologically similar words (containing 40 items). For the first dataset of 120 items, a three-way analysis of variance was conducted, in which the three independent variables were syllable type (four levels: CVCCVC, CCVCV, CVCVC, and CVCCV), word class (two levels: noun and verb), and task (two levels: production and perception), and the dependent variable was the initial percentages of stress assignment. It was found that the main effects of word class and task were statistically significant (p < .001) and the others (including the main effect of syllable type and all interaction effects) were not (p > .05); R² = .82, suggesting very strong main effects of word class and task.

For the second dataset of 40 items, only the four syllable types were involved. Thus, a one-way analysis of variance was conducted, in which the independent variable was syllable type (four levels) and the dependent variable was the initial percentages of stress assignment. No statistically significant effect was found (F(3, 36) = 1.30, p = .29).

5. Discussion and conclusion

This study examined Chinese ESL students’ acquisition of English word stress and the factors affecting stress assignment for English pseudowords in both production and perception tasks. The major findings are summarized as follows.

The participants were approximately 90% correct in both production and perception tasks of real words. In the production and perception tasks of pseudowords, the initial percentage of words in a noun frame was 82.58%, twice as much as the words in verb frames (38.92%). Among the four word types, Type 1 and Type 3 received more initial stress in production tasks. However, this result was not evident in perception tasks.

In both the production and perception tasks, the noun frame received more initial stress than the verb frame, regardless of the influence of syllable types. There was a moderately significant correlation between the stress assignment of pseudowords, as in production tasks, and their phonologically similar counterparts. Predictions of stress placement made by lexical class contributed significantly to the prediction of main stress assignment.

The Chinese students in this study could more accurately produce rather than perceive word stress. These advanced Chinese ESL learners have more or less implicitly learned the stress rules during their primary and secondary school years, and they usually imitate the speech of their teachers or from what they hear in the media. For them, it may be easier to produce the correct stress patterns, but they may not be able to notice or be aware of the differences in these stress patterns. Since both perception and production are considered vital for communication, it is suggested that ESL teachers incorporate a period of perception training before or after production training.

Among the three factors determining word stress, only word class has been found to have predictive power. In contrast with the findings from Guion et al., syllable structures for Chinese ESL learners did not have a significant effect on word stress. In fact, syllable structure rules, such as super heavy syllables bearing stress, might not be difficult to learn, but applying them seems to be difficult in real contexts when unfamiliar words are encountered. In the four types of syllable structures, Chinese ESL learners exhibited more initial stress overall; one reason, according to Archibald’s study (1997), may be that even native English speakers tend to stress the initial syllable for most items reported. Whether it is a universal tendency or a language-specific phenomenon, further studies of cross-language analyses are needed.

Another possible explanation is learners’ first language transfer, that is, the role of tone in the Chinese language. The phonetic properties used to distinguish suprasegmental features, such as stress and tone, vary from language to language. Chinese words in general are monosyllabic, and there is a one-to-one mapping between characters and syllables. The tone of each Chinese word contributes to lexical meanings. Tone is an integral part of word meaning, which is arbitrary and cannot be manipulated at the speaker’s free will. Therefore, whenever Chinese learners are not able to perceive or produce the stress in multisyllabic words, initial stress will be considered first. The differences between the Chinese and English languages affect how ESL learners perceive and produce phonetic qualities.

In respect to the finding of phonologically similar words, Chinese ESL participants’ patterns of stress assignment on pseudowords in Task 2 were only marginally significantly correlated with the stress patterns of phonologically similar real words in Task 3. One plausible explanation is the nature of the task. Participants were asked to listen to an unfamiliar disyllabic word and then to produce any real words they could associate with the unfamiliar word. This kind of task was a new experience and was possibly more challenging to participants than analogy tasks that are commonly used in printed texts. ESL learners in China are usually more familiar with using analogy strategies via reading tasks, visualizing the printed letters, and associating the unknown words to known words with similar word formations. Compared to the participants in Guion’s study (e.g., the English speakers or Spanish bilinguals), even the participants of this study identified as advanced English learners found the task too difficult to come up with appropriate words within the given time limit. This may also have undermined participants’ real performance.
To further discuss the importance of this associative learning process, we must ask whether analogy plays a role in English word stress assignment. It is commonly believed that analogy is an effective phonics strategy for ESL learners to predict sound patterns. However, in reality, a significant number of real words are consistent and inconsistent with stress rules; the random pattern of learners’ misperceptions and mispronunciation in consistent and inconsistent groups, as shown in the findings of Tasks 1 and 5 (real word production and perception tasks), indicate that little influence from explicit stress rules has been found. As Archibald (1997) and Flege and Bohn (1989) found, learners tend to store word stress lexically. Jenkins (2002) even claimed that some of the phonological features, such as word stress, are not teachable as the rules are too complex.

However, learning stress on a word-by-word basis is not an efficient way to acquire an understanding of word stress rules. The current study has supported the notion that Chinese learners could benefit from learning general word stress rules based on grammatical category. Lexical class has been found to be easier to acquire than syllable structures and phonologically similar words, and thus priority could be placed on teaching it.

Van Rees, Ballard, McCabe, Macdonald-D’Silva, and Arciuli (2012) recently conducted a study examining whether children improve their independent production of lexical stress based on orthography with explicit training versus no training and if children improve in this skill, whether improvement is sustained post training. Their results suggested that the training group learned to produce appropriate lexical stress for the pseudowords with strong maintenance and generalization to related untrained stimuli. Findings revealed that typically developing children can learn to produce lexical stress patterns for orthographically biased pseudowords via explicit training methods. This approach is typically used to encourage generalizability rather than item-by-item learning. Therefore, the approach could be explored in the area of English instruction for non-native speakers.

Arciuli, Monaghan, and Seva (2010) examined the process of stress assignment in learning to read through a combination of three methods. The triangulation of corpus, behavioral, and computational modeling techniques within this paper reflected the gradual process of learning the statistical properties of written input and provide key constraints for adequate models of reading aloud. The written stimuli of beginnings and endings of English bisyllabic words in this study were highly predictive of stress position, which could be helpful in the instruction of English in assisting accuracy of stress production.

On a basis of the present study’s findings, several suggestions for further study can be made. First, L2 learners from a non-stress language background, such as Japanese speakers, may not possess a system of stress in the same way as native speakers; therefore, we may expand the study to include more subjects with different language backgrounds to increase the validity of the research results. In addition, different learning environments (L1, ESL, or EFL contexts) may influence learners’ quality and quantity of language input, and thus either facilitate or impede the acquisition of language, especially for the abstract suprasegmental features like word stress. Compared with native English speakers or bilinguals in English-speaking countries, ESL or EFL learners living in their home countries may have different learning obstacles. Further studies on this aspect are worthy of investigation.

Moreover, while counting the valid tokens of phonologically similar words in Task 3 of the current study, all of the disyllabic words with obvious stress preference were included regardless of their syllabic structures. It is difficult to define which kind of words or syllabic structures could be included as phonologically similar words. Therefore, clearer criteria to define “phonologically similar words” are needed. Next, tasks appropriate for ESL or EFL learners should be redesigned to closely measure their effect on word stress assignment. Finally, in addition to the three factors influencing English stress placement investigated in this study, several other factors, such as the morphological structure of words and the number of syllables in words, could be further considered in future research.

Appendix A. Summary of phonologically similar word production

<table>
<thead>
<tr>
<th>Syllabic structure</th>
<th>Examples</th>
<th>Most frequently presented words</th>
<th>Total responses</th>
<th>Percentage with initial stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial stress</td>
<td>Final stress</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVVCVCC</td>
<td>forget(2), setback, desktop(2), kidnap(2), decades(2), liquid, deadly, select,</td>
<td>detect(2), Tibet, decline, delight,</td>
<td>143</td>
<td>50.40%</td>
</tr>
<tr>
<td>bet tist</td>
<td>dentist(6), Baptism(3), daily, weekends, banquet, dentals, prospect,</td>
<td>debate(2), betray, respect, aspect, rotate, predict, protect(4), possess, persist, select, hotel, obey, protest, suspect</td>
<td>dentist(6) Baptism(3), debate(2), protect(4),</td>
<td>136</td>
</tr>
<tr>
<td>psubekt</td>
<td>homepage, backpack, Baptist, basket, scientist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVVCVCC</td>
<td>forget(2), setback, desktop(2), kidnap(2), decades(2), liquid, deadly, select,</td>
<td>detect(2), Tibet, decline, delight,</td>
<td>143</td>
<td>50.40%</td>
</tr>
<tr>
<td>de kips</td>
<td>dentist(6), keeping, artist, gadget, sentence, settings, scientist, city, Baptist, scientist, satisfy, physics, stylist</td>
<td>transcribe forget, certificate, police, tonight, in’sist, re’leased</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se tist</td>
<td>dentist(6), keeping, artist, gadget, sentence, settings, scientist, city, Baptist, scientist, satisfy, physics, stylist</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Phonologically similar word production

<table>
<thead>
<tr>
<th>Syllabic structure</th>
<th>Examples</th>
<th>Most frequently presented words</th>
<th>Total responses</th>
<th>Percentage with initial stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVCVC</td>
<td>basin(3), listen(2), beaten(2), nephew(2), beaten, design, recent, divorce, neglect neglect(8), basin(3), listen(2), beaten(2), nephew(2)</td>
<td>154</td>
<td>64.59%</td>
<td></td>
</tr>
<tr>
<td>sn sn</td>
<td>deepen, dancing, reading, napkin, minute, lesson, dialect, nickname, islands, leaflet, needy, happy, sorry, sandy, daylight</td>
<td>bett(2), Baptista(2), keyboard, biscuit, beatles,</td>
<td>138</td>
<td>62.01%</td>
</tr>
<tr>
<td>nl let</td>
<td>silly, sandy, basin, letter, leader</td>
<td>bet(2), rephrase, dictation, café salut(2), elite(2), retreat, forgive, believe massage, delete,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVCVVC</td>
<td>basin(3), listen(2), beaten(2), nephew(2), beaten, design, recent, divorce, neglect neglect(8), basin(3), listen(2), beaten(2), nephew(2)</td>
<td>154</td>
<td>64.59%</td>
<td></td>
</tr>
<tr>
<td>bt tesi</td>
<td>bitter(2), Baptista(2), keyboard, biscuit, beatles,</td>
<td>bett(2), Baptista(2), bet(2), salut(2), elite(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>se list</td>
<td>silly, sandy, basin, letter, leader</td>
<td>bet(2), rephrase, dictation, café salut(2), elite(2), retreat, forgive, believe massage, delete,</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 571 57.84%

Note: The numbers in the brackets indicate the times the words were produced by the speakers.

References


Matching vocabulary learning process with learning outcome in L2 academic writing: An exploratory case study

Qing Ma

Department of Linguistics and Modern Languages, The Hong Institute of Education, 10 Lo Ping Road, Tai Po, N. T., Hong Kong

ARTICLE INFO

Available online 6 December 2012

Keywords:
Vocabulary learning
Academic writing
Metacognition
Lexical quality
Case study

ABSTRACT

This exploratory case study of two undergraduates links vocabulary learning approaches with lexical quality measured in academic writing. Employing an array of qualitative data, it is shown that in a “semi-language-rich” learning context, Chinese learners may dispense with rote learning and engage in a more natural learning approach in which vocabulary items are discovered from reading newspapers and consolidated in writing. A shift in language assessment is another key factor that contributes to this rarely reported learning approach among Chinese learners. Measuring lexical quality shows that the two students made impressive but differentiated lexical achievement. Underlying these findings is the two students' highly explicit metacognitive awareness of the learning process in initiating, selecting and consolidating the vocabulary items to be learned in academic learning contexts.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Traditionally, L2 vocabulary acquisition focuses on the learning process, i.e. documenting how vocabulary items are learned incidentally from reading or intentionally from list memorisation or explicit instruction. Another established approach to researching into the vocabulary learning process is to examine what learning strategies are self-reported by learners. A number of such studies have attempted to establish the link between the learning process and the learning outcome (Fan, 2003; Gu, 2003; Gu & Johnson, 1996; Kojic-Sabo & Lightbown, 1999; Lawson & Hogben, 1996). However, when both are addressed, the outcome is often measured in terms of vocabulary size or general language proficiency. A growing body of studies has shown that lexical quality is tightly related to L2 writing quality (Baba, 2009; Engber, 1995; Laufer & Nation, 1995; Lee & Mune, 2006), but no studies investigating both learning process and learning outcome have, to our knowledge, measured lexical use in free essay writing. Given the importance of academic writing in most programmes for undergraduate or postgraduate courses, it becomes urgent to examine how the vocabulary learning process accounts for lexical quality in academic writing.

This article begins with reviews of lexical studies that investigate learners’ vocabulary strategy use, the aim being to show how learning process is related to learning outcome. A stereotypical image of Chinese learners (mainly non-language specialists) emerges from this review process. The importance of measuring lexical use in essay writing is then highlighted. This is followed by a case study of two high achieving Chinese students’ vocabulary learning experience as well as their marked lexical achievement as measured in academic assignments. Both similarities and differences will be pinpointed with a view to showing how individual learning paths can account for their similar but nevertheless differentiated learning outcome.

* Tel.: +852 2948 8593,
E-mail address: maqing@ied.edu.hk

0898-5898/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.linged.2012.11.002
It is shown how the two learners’ vocabulary learning accounts diverge from the stereotypical image of Chinese learners; this different picture is deemed to be attributed to the nature of the learning environment and the shifting assessment formats.

2. Vocabulary learning strategies and vocabulary achievement

Most studies that investigate learners' vocabulary strategies have tried to discover how strategy use could contribute to vocabulary knowledge or general language proficiency, although this is not the sole focus of the studies. Findings generally show that different levels of vocabulary achievement are associated with combinations or clusters of strategies (Gu & Johnson, 1996; Kojic-Sabo & Lightbown, 1999; Sananui, 1995), i.e. metacognitive regulation strategies such as the efforts made in initiating learning opportunities and selecting words (Gu & Johnson, 1996) as well as certain cognitive strategies involving the use of dictionaries or guessing (Fan, 2003). However, learning outcome in these studies is always measured in size tests, or in general English proficiency with commonly used test formats. For example, in Gu and Johnson’s study (1996), learning outcome was measured in (1) vocabulary size and (2) general proficiency composed of discrete language skills: listening, reading, vocabulary, grammar and translation. Writing, a key factor in measuring lexical use, was excluded from the measures. One reason is that only non-English majors were included in the study. Non-English majors’ proficiency is often measured in standardised tests where writing only counts for a small proportion. On the other hand, writing should be accorded much more importance for language majors or postgraduate students as their assignments frequently require essay submissions.

In a subsequent case study, Gu (2003) examined two high-achieving Chinese university students’ vocabulary learning using think-aloud protocols in conjunction with interviews. One female student read extensively, and was hence a “reader,” and the other male student employed a wide range of strategies, this making him an “active strategy user.” The active strategy user exclusively focused on reading textbooks and treated most new vocabulary items meticulously, noting down various lexical aspects for the new item: pronunciation, meaning, synonyms, usage and example sentences. Such careful learning was followed by memorisation, i.e. oral repetition till the word had been memorised; he used both self-made vocabulary lists and commercial ones. By contrast, the reader confessed to being a lover of English and was drawn to the beauty of the language. She read novels and other works of prose, deriving great pleasure from the emotions, sounds and images evoked from reading these works. Apart from extensive reading for pleasure, like the other student, she also made great efforts to memorise vocabulary lists, mainly commercial ones; she went through the lists alphabetically and tested herself on each word in turn.

3. Chinese students: bound to be rote learners?

Gu’s study (2003) confirmed a stereotype of Chinese learners, namely, that, regardless of being good or poor learners, they relied heavily, if not exclusively, on rote. Gan, Humphreys, and Hamp-Lyons (2004) also came to the conclusion that both successful and unsuccessful learners regarded rote learning as the most valuable strategy, though the former would make efforts to consolidate the memorised vocabulary whereas the latter would not. While these studies involved only non-language majors, that of Ding (2007) constructed the same image of rote learners among elite English majors in China, who regarded text memorisation and imitation as the most efficient way to learn English. Chinese students’ rote learning style is assumed to be heavily influenced by Confucianism (Hu, 2002; Marton, Dall’Alba, & Tse, 1996; Watkins & Biggs, 1996). Ma (2011) provides more contextualised explanations for Chinese learners’ rote learning style: (1) the culture of learning; (2) vocabulary learning difficulties; (3) the exam-centred Chinese education system; (4) profitable commercial vocabulary lists which encourage Chinese learners’ rote learning style. Recent studies show that strategy use is not necessarily culture-bound but may be context or situation-dependent; such contextual mediating factors are complex and dynamic, including instructional pedagogy, language proficiency (Cheng, 2000), language policy (Hu, 2003), learning experience at home and abroad (Gao, 2006), and historical perspectives (Jiang & Smith, 2009). Immersed in the L2 countries (UK, USA, Canada or Australia) or a comparatively rich language-input environment (such as Hong Kong or Singapore), students can learn outside, as well as inside, the classroom, by communicating in the target language. More importantly, writing academic assignments for courses pushes them to think and organise ideas in the L2; they thus engage in a high level of cognitive processing, which would allow them to maximise their language use. Given such a fast changing scene in the language learning context, it is necessary to construct an updated account of how Chinese learners learn L2 vocabulary in a different context rather than adhere to the traditional stereotype.

4. Lexical use in L2 writing

Along with the growing interest in vocabulary, lexical use has been shown to be crucial in developing various language abilities, including L2 writing. Vocabulary is now considered one of the most important indicators of the overall quality of L2 writing (Baba, 2009; Engber, 1995; Laufer, 1994; Lee & Muncie, 2006). Laufer and Nation (1995) put it this way: “A well-written composition, among other things, makes effective use of vocabulary. This need not be reflected in a rich vocabulary, but a well-used rich vocabulary is likely to have a positive effect on the reader.” (p. 307). The literature on L2 vocabulary frequently distinguishes two types of knowledge, i.e. size and depth. Vocabulary size or breadth refers to the number of words of which at least some aspects of the meaning the learner knows (Anderson & Freebody, 1981). Passive or
meaning recognition tests are frequently used to measure vocabulary size (Ma, 2009; Milton, 2009). Depth refers to the more qualitative aspects of word meanings and use as measured in Read’s (1993, 1998) Word Associates Test. Words produced in L2 writing can be measured in terms of size, the range of vocabulary at different frequency levels, as well as depth, i.e. lexical quality, which is frequently equated with lexical richness. Common measures of lexical richness include lexical variation, density, sophistication and error rate (Read, 2000). Lexical variation refers to the type/token ratio (TTR); lexical density is the ratio between the number of lexical words and the total number of words; lexical sophistication is the ratio between the number of infrequent words and the total number of words; lexical error rate is the ratio between the total number of errors and the total number of word tokens. However, not all of these indices have been shown to be equally related to writing quality. Engber (1995) employed timed compositions (around 300 tokens) as part of a language test in order to examine the relationship between lexical richness and overall rating scores of L2 writing. Fairly high significant correlations were found between the overall rating scores and the lexical variation with error (.45) or without error (.57), and the lexical error rate (−.43), but no significant correlation was found between the overall rating and the lexical density. Engber’s study suggests that having a wide range of vocabulary as well as being able to use these items accurately are two key factors in determining high writing quality.

Lauffer and Nation (1995) proposed an alternative measure of lexical richness – Lexical Frequency Profile (LFP). A computer programme screens the input texts and measures the range of learners’ productive vocabulary divided into four frequency levels: the first 1000 words (K1), the second 1000 words (K2), the University Word List (UWL) and the off-list words (OL) not in any of the previous three lists. Using LFP, they investigated three different proficiency groups’ lexical use in short compositions produced during class time. The results showed that the three proficiency groups demonstrated significant differences in K1, UWL and OL, but not K2. A productive vocabulary test yielded good correlations (ranging from .6 to .8) with OL and with UWL, and negative correlations with K1 (−.7). This suggests that more proficient learners tend to use more vocabulary in the UWL and OL but less at the most basic K1 level.

Using a modified version of LFP, Morris and Cobb (2004) presented a programme called Vocabprofiler, providing lexical percentages for K1, K2, AWL (the academic Word List, developed by Coxhead, 2000) and OL. They used Vocabprofiler to examine the relationship between learners’ lexical knowledge in essay writing and their subsequent academic grades in a grammar course. The two authors argued for a strong connection between use of AWL and academic success in tertiary education; they claimed that language skills, particularly lexical skills, are related to “higher level thinking skills that university students and future teachers need to be successful” (Morris & Cobb, 2004, p. 78). The correlation obtained between the course grades and the AWL and K1 were moderate but yet significant (.37 and −.34 respectively). These studies indicated that making use of academic words, as well as showing a wide range of vocabulary beyond K1 in academic writing, are associated with academic success.

All the studies cited above employed timed essays in a testing or similar environment. Studies by Kenworthy (2006) and Muncie (2002) have shown that learners’ lexical features behave differently in timed essays and at-home essays. As a corollary, these timed essays were quite short, around 300–500 words. Measuring learners’ lexical use in free at-home longer essays is called for.

5. The current study

To date, there is no record of any study that looks into the relationship between learning process and learning outcome as measured in terms of productive knowledge in free L2 academic writing. In learning environments such as Hong Kong universities where English is used as the medium of instruction and most courses require essay submission, measuring learners’ lexical use in academic writing would make much more sense than simply administering passive vocabulary tests such as the size test. It would also be meaningful to look into learners’ vocabulary learning process by identifying the key features or strategy use related to the learning outcome. Linking vocabulary learning process with lexical quality in academic writing could inform both learners and teachers of the key actions or pedagogy to improve the quality of academic writing. In addition, fathoming how the learning process contributes to the learning outcome for successful L2 Chinese learners in underreported contexts will lead to a deeper understanding of by far the largest world-wide legion of English learners.

This study focuses on two high-achieving students who were enrolled in a Bachelor of Education programme, while studying for a degree in English Language, in a Hong Kong tertiary institution. Efforts were made to document both their learning process, i.e. accounts of learning experiences in terms of strategy use, and learning outcome, i.e. the lexical quality as measured by two lexical indices of two academic essays: lexical profiles (size of productive vocabulary measured at different frequency levels) and lexical variation (TTR). The first was considered by Lauffer and Nation (1995) as one important measure for lexical richness in writing and the second placed at the top among the four essential features for lexical richness constructed by Read (2000). Of these four measures, it is lexical variation, or TTR, that has been widely employed in studies that investigate lexical use in written or oral language (Javis, 2002; Lemmouh, 2008; Vermeer, 2000). The other three, i.e. lexical sophistication, density and error rate, will not be reported due to the limited space and their relatively less frequent occurrence in empirical studies. There are two research questions that guide this exploratory study:

1. How do the two students approach vocabulary learning at the tertiary level?
2. What are their lexical profile and lexical variation as measured in their academic writing?
A case study approach is adopted in this study for several reasons. First, case studies have a long history documenting learners’ language learning process. Second, the current study is exploratory in nature and a case study can fulfil this purpose (Yin, 2003). Thirdly, many quantitative studies have been conducted to investigate either learners’ vocabulary learning process or their lexical quality separately, but findings do not shed light on the interconnections between the two. Thus, a small-scale qualitative study is needed prior to a larger scale investigation. Data are collected from multiple sources: questionnaire, interview, vocabulary notebooks, self-reflections revealed in email correspondence, and academic assignments, which is in line with Yin’s (2009) claim that using multiple sources of data is to ensure data triangulation; multiple sources of evidence are “highly complementary” (p. 101) and could “corroborate and augment evidence from other sources” (p. 103).

5.1. Participants

Two high-achieving students, K and A, in the English Department of a tertiary institute in Hong Kong were chosen from 110 fourth year students to participate in the present study. They were placed in first and third position upon graduation in their cohort, both receiving a first class Honours, with a GPA higher than 3.4.

5.2. Data collection and analysis

5.2.1. Learning process

A self-designed vocabulary learning strategy questionnaire based on previous research (Fan, 2003; Gu & Johnson, 1996; Schmitt, 1997) was given to the two students; this was followed by a semi-structured interview aimed at probing deeply into K’s and A’s vocabulary learning process. The interview questions were organised based on the three learning stages as identified in Gu’s (2003) study: how learners discover new words and their initial handling of these words, how they commit the words to memory and how they make use of the words. The interview was conducted in English as the researcher is not fluent in Cantonese and K and A are both fluent in English. During the interview, it was discovered that both A and K had the habit of keeping vocabulary notebooks. After the interview, they agreed to submit their vocabulary notebooks as additional evidence for their vocabulary learning. After reading through the transcripts and studying the vocabulary notebooks, the researcher sent follow-up emails to K and A, requesting more detailed information regarding their vocabulary learning approaches as a way of self-reflection which unveiled their motivation and purposes of their strategy choice. Since the self-report questionnaire contained only Likert-type questions and what was revealed from the questionnaire was consistent with the findings from the interview, only the qualitative data from the interview, the vocabulary notebooks and email correspondence will be reported later. The data analysis went through a two-stage process: initial coding and categorising types in order to “identify key features and relationships in the data” (Richards, 2003, p. 273). Emerging themes that resulted from initial coding were merged to produce patterns in the data by techniques as informed by Grounded Theory developed by Strauss and Corbin (1998).

5.2.2. Learning outcome

In order to measure their lexical achievement, each student submitted two essays they had written as assignments for two common courses they took in the previous semester. One essay was about how to deal with learning diversity (or students with learning disabilities) in the classroom and the other was a discussion of Henrik Ibsen’s plays. Thus the essays were on two markedly different topics: special education and drama. Further, a length of 3000 words was specified for the first essay and 1500 words for the second. The essays were put into the programme “Range” developed by Paul Nation (available at http://www.victoria.ac.nz/lals/staff/Publications/paulnation/Range_GSL_AWL.zip); it is an updated version of the LFP discussed in Laufer and Nation (1995). This tool was used to analyse the lexical features of the text and display a number of indices such as the proportion of 1K, 2K, AWL and OL. So far, there are no established standards or a threshold level of proportion for each of these four frequency levels for academic writing. A few studies have alluded to some important indices regarding learners’ lexical profiles, one being the study by Laufer and Nation (1995), which suggested that more proficient learners use more words in the AWL and the off-list words. For example, the most proficient group (second year Israeli English majors) achieved a range of 8.1% to 10.1% and 7.5% to 8.7% at these two levels respectively. By correlating profile scores with academic grades, Morris and Cobb (2004, p. 83) arrived at a “fairly modest writing standard” for vocabulary profile: K1 < 85%; AWL > 5%. But the essays were compiled by only first year TESL students of mixed L1-background (English, French, Italian, Greek, etc.). In addition, these indices were obtained from texts of around 300 words in length. It would be difficult to employ these indices as standards. As other factors may come into play: differing proficiency levels, varying L1 backgrounds and fixed text length.

In order to overcome the lack of any standard score to evaluate K’s and A’s lexical profile, two academic essays written by well educated native speakers on similar topics and of similar length were chosen. By using key words and searching in relevant journals; two articles were selected from two different international journals; one related to special education and the other to drama which also discusses Ibsen’s plays. The author information provided on the title page shows that both authors were native speakers of English and were working at universities in English-speaking countries. Lexical variation or type/token ratio (TTR) was measured in WordSmith (version 5).
6. Results

In this section, qualitative findings are described under “learning process” while the analysis of the students’ two essays are compared with scholarly equivalents quantitatively under “learning outcome.”

6.1. Learning process

6.1.1. Similarities

Two patterns were generated for both students’ vocabulary learning process, one reflecting the similarities and the other the differences. A clear and similar path of vocabulary learning emerged for K and A: both students regarded English newspapers as the primary source to encounter new vocabulary items; they would put down important words in a notebook; they considered using newly learned words directly to be the most effective means to retain these words in memory. Reading newspapers was frequently mentioned during the interview and both said they read a newspaper almost every day, usually the South China Morning Post. This was confirmed by the vocabulary notebooks and additional learning evidence provided by the two students. A’s vocabulary notebooks comprised essentially pasted news cuttings; unknown words in each text were marked (underlined or circled) and annotated (mostly in English and occasionally in Chinese). As for K, his vocabulary notebook was a classic one, composed of lists of lexical words or phrases, but he also supplied a number of text cuttings from newspapers and in subsequent email correspondence explained why he kept news cuttings:

I always think the usage of English in newspaper is very helpful as it reflects a common practice and usage of English in the real world so a study of it can definitely help to improve my mastering of the language. (K-email)

Newspapers were not only a source for them to discover new words but also provided good opportunities for them to consolidate newly learned words:

Uh, yes, particularly my source of new words is mainly from newspapers, so I think they are quite frequently used in the newspaper as well, so once I go into the newspaper again, I will catch the newly encountered words and I know the meaning now so I can read it smoothly. (K-interview)

Both mentioned that academic lectures could provide a source for vocabulary learning, but it was their impression that they did not find new vocabulary items very often from academic lectures. Lecturers would usually explain those important academic words and four years of exposure to academic lectures and readings have consolidated their mastery of the academic vocabulary; this has been particularly the case for A, as will be shown later. Frequent contacts with academic vocabulary encouraged a rather unconscious “acquisition” process instead of a conscious “learning process”.

After finishing reading the whole newspaper, both K and A would note down important or interesting words, together with the Chinese meaning and sometimes an example sentence or synonym. Both would review the words in their notebook but not on a regular basis. K explained later in email correspondence his metacognitive strategy:

My belief is that when this is the new knowledge, keeping it in the notebook can refresh my mind on that and help me to be familiar with its application. But if I keep on doing that after some time, I may develop a tendency to rely too much on it without the confidence of using it freely. Therefore, after some time, the notebook will no longer be referred to. (K-email)

Regarding the most effective means for them to memorise vocabulary, both affirmed it was by using newly learned vocabulary in writing or speaking. Although K had tried to use different memory strategies, he claimed that those memory strategies were “not as useful as directly using it [the new word]”. A said he used some memory strategies in the past (primary or secondary school) but not now. When it comes to how they used newly learned words, both K and A would prefer to use words in writing rather than speaking as they said: “we tend to use simple vocabulary” in oral English.

6.1.2. Differences

Though it appears that K and A were similar in many ways regarding their vocabulary learning approaches, the data also disclosed a number of noticeable differences between them. First, K reported to have occasionally used a number of memory strategies though he did not think these strategies were more helpful than using the words directly. For example, he reported dividing words into syllables, grouping words according to the word class, using actions to help to remember the meaning of words or phrases. By contrast, A only reported using rhyming or similar sounds for memorising words on a few occasions, without mention of any other memory strategies. Second, A was more systematic and strategic in managing his learning than K. For example, A collected news cutting once a week; this habit was formed when he was a Form 4 student in the secondary school and attributed to his teacher’s advice. A also showed the strategic efforts he made in dealing with the news so as to achieve maximum benefit from learning vocabulary items:

Sometimes, I would listen to the Chinese version of a piece of news on TV and then find the corresponding/relevant news in the newspaper, so I didn’t need to spend time on understanding the context but on vocabulary and sentence structures. (A-email)
Table 1
General information pertaining to students’ essays after editing.

<table>
<thead>
<tr>
<th></th>
<th>K Special education</th>
<th>Drama</th>
<th>A Special education</th>
<th>Drama</th>
</tr>
</thead>
<tbody>
<tr>
<td>Token Type</td>
<td>2796</td>
<td>1339</td>
<td>3090</td>
<td>1488</td>
</tr>
<tr>
<td></td>
<td>746</td>
<td>408</td>
<td>909</td>
<td>564</td>
</tr>
</tbody>
</table>

Table 2
Lexical profiles of native authors’ articles (in token and type).

<table>
<thead>
<tr>
<th>Word List</th>
<th>Special education (Native 1)</th>
<th>Drama (Native 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Token/% Type/%</td>
<td>Token/% Type/%</td>
</tr>
<tr>
<td>K1</td>
<td>2258/72.63 410/52.36</td>
<td>1155/78.31 351/55.45</td>
</tr>
<tr>
<td>K2</td>
<td>240/7.72    73/9.32</td>
<td>71/4.81 61/9.64</td>
</tr>
<tr>
<td>AWL</td>
<td>361/11.61   155/19.80</td>
<td>91/6.17 79/12.48</td>
</tr>
<tr>
<td>OL</td>
<td>250/8.04    145/18.52</td>
<td>158/10.71 142/22.43</td>
</tr>
<tr>
<td>Total</td>
<td>3109         783</td>
<td>1475          633</td>
</tr>
</tbody>
</table>

Thirdly, although both A and K kept vocabulary notebooks, the way in which A arranged his news cuttings and vocabulary items shows that there is a qualitative difference in the two’s approaches: A arranged news cuttings and relevant vocabulary items in themes; he divided all news cuttings according to five themes in his notebook: education, environment, health, comics (short dialogues accompanied with pictures) and miscellaneous (sports, music, entertainment, etc.). By contrast, K simply jotted down items alphabetically.

Finally, K and A preferred to use words in different written registers: written conversations (emails) for K and formal writing (academic assignments) for A who stressed the deliberate efforts he made to expand his range of vocabulary in writing academic assignments:

When I try to use [newly learned vocabulary], it’s mostly, I think, in writing, and perhaps now sometimes when I put down my understanding of my argument in a sentence, in an assignment, sometimes I’d try to ask myself ‘Can I use another word to substitute, or write or use this word, or can I use another one?’ and like that. (A-interview)

6.2. Learning outcome

6.2.1. Lexical profiles

The student essays were edited before being put into the Range programme. Different types of information were removed from the essays: (1) title page, (2) proper nouns, (2) abbreviated forms, (3) direct citation with question marks and (5) references. The articles written by the two native speakers went through the same procedure of editing. After such careful editing, it is fair to say that what remained to be compared could truly reflect the lexical repertoire of all the four authors in their writing. Each essay/article was put into Range separately. Range is similar to LPP or VocabProfile, showing the four frequency levels of the text input: the first and second 1000 word families (K1 and K2), the AWL and OL. Table 1 shows the general information pertaining to K’s and A’s essays after editing.

A wrote longer essays than K in terms of tokens and this remained unchanged after editing when both token and type were taken into consideration. In order to establish the standards to which the students’ lexical profiles could be compared, Table 2 provides lexical profiles of the two native authors at the four frequency levels.

Table 2 shows that if tokens are counted, both articles relied predominantly on the first two levels (K1 + K2), which accounted for more than 80% in both cases. The two remaining more sophisticated levels account for less than 20% in both cases. However, when types are counted, the distribution of the four frequency levels shows a notable difference. There is a decrease of at least 20% for both articles in K1 and a substantial increase in AWL and OL, boosting the combined two more sophisticated levels to more than 30%. Although the table shows that the running words (tokens) in special education are twice as many as those for drama (3190 vs. 1475), the difference in word type is much smaller with only a difference of 150 words. This indicates that there tends to be a considerable repetition of words in longer texts. Given that we are only interested in the productive size of authors, counting word types would be more appropriate than counting word tokens. For this reason, the analysis will be presented in word types in the ensuing sections.

When the native authors’ lexical profiles are examined for types, some differences are observed due to the genre of the two articles. Overall, while the two articles remain similar at the K1 and K2 levels (61.12% and 65.09% respectively), the one in special education used a considerably higher proportion of words in AWL (19.80% vs. 12.48%) and a lower proportion in OL (18.52% vs. 22.43%) than in drama. These differences could be attributed to the genre differences in the two articles. It is expected that this trend would be maintained in the essays produced by the two student authors if they were to have had a similar lexical achievement to that of the two native authors.

Table 3 presents the students’ lexical profiles together with the native authors’ profiles for reader’s convenience. At the more basic levels, i.e. K1 combined with K2, A’s essays range from 66.12% to 69.51%, similar to the combined K1 and K2
percentages for the two native authors (61.12% and 65.09%). At the more sophisticated levels, A achieved 19.69% and 13.83% in AWL in special education and drama respectively. This indicates that A’s essays have shown a similar if not identical trend at AWL level to that of the native authors’ essays: a higher proportion of AWL and a lower proportion of OL in special education than in drama. The actual percentages of these two levels are also quite close to those of the native authors (19.80% and 12.48%). This demonstrates that A has achieved a fairly satisfactory command of academic vocabulary in terms of productive size, almost comparable to that of native authors. By contrast, A’s percentages in OL (14.19% in special education and 16.67 in drama) are less satisfactory when compared to those of the native authors (18.52% in special education and 22.43% in drama).

We will now turn to K’s lexical profiles. The two combined proportions at the basic levels (K1 and K2) were 73.59% and 82.35% in the two essays respectively, which was considerably higher than those of the two native authors and that of A. K’s percentages for AWL were 16.22% in special education and 10.78% in drama, this being considerably lower than the native authors’ range as well as A’s. Regarding OL, the percentages were 10.19% in special education and 6.86% in drama, this too being much lower than those of the native authors and A’s. Apart from these notable differences, K, unlike the two native authors and the other student author, has only demonstrated a greater proportion of AWL in special education but not of OL in drama. On the other hand, when compared with the standards set by Morris and Cobb (2004), i.e. K1 < 85% and AWL > 5%, K’s lexical profile should be considered quite satisfactory, his lexical indices being far above these two levels.

6.2.2. Type/token ratio (TTR)

Researchers frequently remind us that TTR suffers in longer texts, i.e. the longer the text, the lower the TTR. In this study, all student essays and native authors’ articles were divided into segments of 300 words and the TTR based on each segment was calculated. Finally, a general mean was obtained, resulting in the Mean Standardised TTR (MSTTR), calculated using WordSmith version 5. See Table 4 for the TTR and MSTTR of all authors.

The table shows clearly that TTR is considerably lower for longer texts on special education (almost twice as long as drama). The MSTTR shows less deviation across the two different genres than the non-standardised TTR. Not surprisingly, the table also shows that the drama texts tend to show higher MSTTR than in special education (except K’s drama essay) as writing in social science requires language to be precise, accurate and technical. On the other hand, using a wide range of lexis may be a key feature in quality writing in the humanities. This is evident in the two native authors’ articles (51.30 in special education and 58 in drama). It should be noted that the MSTTR of the student essays on special education (K: 53.59; A: 55.87) has greatly exceeded that of the native author. This, on the other hand, suggests the two students used considerably varied lexis compared with the native author. On the other hand, such varied lexis may not be always desirable in writing in social science. Perhaps what matters more is to use relevant academic or technical vocabulary in an accurate and appropriate manner.

7. Discussion

The results delineated the two students’ vocabulary learning paths measured by an array of qualitative data: retrospective interview, notebooks and self-reflections. Their learning outcomes were measured by two indices, lexical profiles and variation, in academic writing. Although measuring the learning outcome in essay writing is not a novel idea, the current study differs from previous studies in two prominent ways. First, much longer academic essays, instead of timed short essays were examined. Second, essays were provided by more advanced learners (high-achieving English majors) instead of low-intermediate learners who often make mistakes that render the text incomprehensible. Putting things together, a number of themes emerge that are worth further discussion.
7.1. Updated account of Chinese learners’ vocabulary learning approaches

The literature has constantly portrayed Chinese L2 learners to be rote learners regardless of their learning styles, strategies, motivation or achievement (Gan et al., 2004; Gu, 2003); this image remains unchanged in the general context of learning and testing (Marton et al., 1996; Tang & Biggs, 1996). Central to all these studies is the concept of “memorisation” employed by Chinese learners, though it does not necessarily equate with mindless repetition. Some argue that Chinese learners’ memorisation is associated with mental activeness or open-mindedness (Hu, 2002; Watkins & Biggs, 2001), thus leading to a deeper understanding of the information to be retained (Marton et al., 1996). The current study shows this is not necessarily true for Chinese learners of an L2 in Hong Kong. Both students took a natural approach to vocabulary learning, namely, noticing new vocabulary from reading newspapers, noting down new items, practicing using the items in writing. What is clearly lacking in this process is reliance on deliberate efforts made to commit the items to memory, either in mechanical repetition or more elaborated techniques such as mnemonics. There are several reasons which may explain the different learning approaches these two students adopted.

First, they were studying in a “semi-language-rich” environment. As Hong Kong is a former colony of Britain, English is one of the official languages and plays a key role in education, especially at the tertiary levels. However, unlike Singapore, where English is spoken in many sectors of the society as a national language, English in Hong Kong remains the language to be spoken with a relatively small number of English-speaking or other L1 expatriates or on limited occasions among local Chinese when non-Chinese are present. Hong Kong is largely a Cantonese-speaking society. Thus “semi-language-rich” is the label for this specific language learning environment. Compared with their counterparts on the Chinese mainland, the two students had been undoubtedly exposed much more frequently to English. In a study that investigates students’ vocabulary learning beliefs (Gao & Ma, 2011), both mainland Chinese and Hong Kong students perceived Hong Kong to be a language learning site that is much richer in learning resources and provides many more opportunities for using the language than in mainland China.

Second, the two students’ major learning strategy, i.e. using words directly in writing instead of relying on memory strategies, is related to the major assessment formats of their tertiary education. In previous studies, learners’ heavy reliance on memory strategies for retaining vocabulary is frequently associated with examinations where vocabulary is assessed in multiple-choice questions or more contextualised formats such as gap filling or translation. However, in the case of K and A, when longer academic take-home assignments were mostly required, they simply needed to put newly learned words into use by completing the written assignments rather than memorise them by heart to cope with examinations, which is not the case anymore. Thus, the change of assessment from language testing to essay submissions is the key factor that changed A and K’s learning approaches; they both reported they had used memory strategies quite a lot in secondary school when they needed to face high-stakes examinations. This provides further evidence that contextual factors may play a key role in mediating learners’ strategy use (Gao, 2006). Using words is perhaps the best way to consolidate words met previously: it forces the learners to pay attention to various aspects, such as meaning, spelling, part of speech, grammatical constraints, collocations, etc. All these types of lexical information can be integrated into a lexical item when the learner attempts to use it in writing.

7.2. Learning approaches and the learning outcome

The results suggest that there is a connection between the approaches to vocabulary learning and the two students’ lexical achievement. Overall, K and A took a similar, but not identical learning approach. Though both have achieved overall satisfactory lexical profiles and MSTTR, A’s lexical achievement was superior to K’s in a number of different ways, as reported above. Although many other factors can play a role, the differences in their learning approaches can partially explain their differentiated learning outcome. Firstly, K preferred to use newly learned words in informal email writing. Unlike K, A favoured practising newly learned words in formal academic writing. A also showed a higher metacognitive awareness in making deliberate efforts to substitute words of similar meaning in writing. All of this may suggest that A is likely to pay more attention to academic vocabulary which is required less in daily email writing; the analysis of his essays showed this is indeed the case: his proportion of AWL was almost as good as that of the other two native authors in writing on two markedly different genres whereas there was a fairly large gap between K and the native authors in terms of AWL. Secondly, though both reported that they read English newspapers almost every day, A was more strategic in processing news and systematic in noting down new lexical items as evidenced in the two notebooks he provided: there were more than 120 cuttings of news texts arranged semantically. In contrast, K only submitted one thin vocabulary notebook which he started to keep in secondary school. This may explain why A’s proportions in off-list words were considerably higher than K’s, though there was still a gap between A’s and the two native authors’ off-list words.

7.3. Explicit awareness and metacognition of successful learners

Good and poor learners are distinguished mainly by their approaches to learning tasks, such as a deep and surface approach to learning (Biggs, 1993; Rao, Gu, Zhang, & Hu, 2007) or having a repertoire of a wide range of strategies and knowing how to select learning strategies for specific learning tasks (Cohen, 1998; Gu, 2003). This study provides evidence for the view that successful learners are also marked by their explicit awareness and metacognition of the processes and
contexts of learning; it adds to our understanding of how successful learners plan and engage in learning tasks. Both A and K fully perceived the learning environment and learning resources available to them; they had clear learning goals for L2 vocabulary and made plans in conjunction with sustained efforts to achieve the goals. In a “semi-rich-language” learning environment and when academic studies gave them full exposure to English, their efforts for vocabulary learning were concentrated on news reading and practising using vocabulary output in writing, which was in sharp contrast with what they did in the secondary school, i.e. memorising vocabulary lists. It is obvious that the shift in learning context made them re-evaluate their strategy use and re-regulate their learning behaviour and efforts. This exemplifies the essential traits for being good learners; they are “learning theorists” (Gu, 2003, p. 73) and metacognition of their learning situation helps them select the strategies that are congruent with their learning tasks and contexts.

8. Conclusion

Chinese learners are not necessarily innate rote learners for language learning as portrayed in the literature. In a ‘semi-language-rich’ learning environment, Chinese learners, as shown by this study, have ample opportunities to discover new vocabulary and put it directly into use. Essay submissions further pushed them to practise using the learned vocabulary in academic writing. This explains why the two students undertook a natural learning approach and made remarkable achievements compared with their peers. At the surface level, the two students took a similar learning approach; at the deep level one was more systematic in managing his vocabulary learning and put more efforts into using learned vocabulary in academic writing. This difference can partially explain their differentiated vocabulary learning outcome as measured in their lexical profiles and variation. A key finding from this study is that learners’ explicit awareness and metacognition of the learning situation initiate, direct and sustain their learning efforts which are the key to success.

Teachers should be aware of learners’ shifting learning contexts, provide appropriate guidance to raise their awareness of the new learning context, perceive the learning sources available and adjust their learning approaches accordingly. Since, as this study shows, English newspapers could become an important source for vocabulary learning, further research should focus on how English newspapers could turn into useful learning materials both inside and outside the classroom learning, particularly when the language learning context is not as rich as in the L2 countries. Teachers may need to push learners to see the value of using news reading, provide them with appropriate guidance regarding how to select news and keep a record of new vocabulary items, prompting them in particular to make early use of these newly learned items.

Due to limited space, the measure focused only on quantitative aspects by obtaining figures regarding learners’ lexical profiles and variation. Future studies that measure students’ essay writing could examine more qualitative aspects of learners’ lexical use, such as lexical phrases, turns of phrase, etc.; these may constitute the construct of lexical idiomaticity which has rarely been explored in L2 studies. A further step is to raise learners’ awareness of lexical idiomaticity as this may push them to make additional efforts to advance their language use.

Lastly, this study illustrates how learners’ productive vocabulary can be measured in writing; this measure is particularly useful in tertiary education when examinations are replaced by academic assignments. Well-educated native authors’ academic writing can be used to set up the standards to which learners’ writing is to be compared, but the topics should be matched as genre is likely to affect the proportions at different frequency levels. Future studies may make use of more native authors’ writing in different genres to establish the standards. Different from other studies involving short essays (300–500 words), the current study measured only types (but not token) in creating the two students’ lexical profiles in much longer essays (1500–3000 words) for the reason that words tend to be repeated in longer essays. In future, researchers may compare both tokens and types in essays of varying length in order to find out how to measure lexical quality more accurately and appropriately.

Acknowledgements

I am grateful to the Hong Kong Institute of Education who funded this article as part of an IRG project (R3139). Thanks also go to Dr. Paul Stapleton and Professor Peter Kelly for their valuable comments on the earlier versions of this article.

Appendix A.

Interview guideline:
How do you normally discover new English words to be learned?
When you find a new word and think it is important, how do you find out the word meaning?
When you think the new word is important and is worth studying, will you record it? If so, how?
How do you memorise the new words?
How do you make use of newly learned words?

References

Teaching Young Learners English Vocabulary with Reading-Based Exercises in a Real Classroom Situation

MA, QING
The Hong Kong Institute of Education, Hong Kong

SIN, CHING HA
Christian Alliance S. Y. Yeh Memorial School, Hong Kong

ABSTRACT: This classroom study investigated how English (L2) vocabulary was taught to primary school pupils in a series of reading lessons under two conditions: (a) reading plus receptive learning exercises or (b) reading plus both receptive and productive learning exercises. The results showed that the condition involving both receptive and productive exercises led to more vocabulary retention than the condition with receptive exercises only. Furthermore, it is tentatively proposed that young learners’ productive knowledge of vocabulary items can be assessed based on acceptable collocations instead of sentence writing which relies on a fully developed grammar system that young learners may lack.

Keywords: vocabulary acquisition/instruction, young learners, receptive exercises, productive exercises, vocabulary learning difficulties

La enseñanza de vocabulario inglés a jóvenes aprendices con ejercicios diseñados a partir de textos de comprensión lectora en una situación real de aula

RESUMEN: Este estudio práctico en el ámbito del aula investigó cómo se enseñaba el vocabulario inglés (L2) a alumnado de primaria en una secuencia de unidades didácticas de textos de comprensión lectora estableciendo dos situaciones: (a) lectura más ejercicios de aprendizaje a nivel receptivo, o (b) lectura más ejercicios de aprendizaje tanto a nivel receptivo como productivo. Los resultados de la investigación muestran que la situación que incluye ejercicios tanto a nivel receptivo como productivo lleva a una mayor retención de vocabulario que la que considera únicamente ejercicios a nivel receptivo. Además, se propone de manera tentativa que el conocimiento de vocabulario a nivel productivo de los aprendices jóvenes se puede evaluar basándose en la producción de colocaciones correctas en lugar de frases escritas que dependen de un sistema gramatical altamente desarrollado del que los aprendices jóvenes pueden carecer.

Palabras claves: adquisición/instrucción de vocabulario, aprendices jóvenes, ejercicios a nivel receptivo, ejercicios a nivel productivo, dificultades de aprendizaje de vocabulario

1. INTRODUCTION

In recent years FL or L2 vocabulary acquisition tends to be investigated in a reading context (Pellicer-Sánchez and Schmitt, 2010; Waring and Donkaewbua, 2008; Horst, 2005; Hill and Laufner, 2003), which is frequently associated with incidental learning modelled on L1 acquisition. It is widely believed that children acquire most L1 vocabulary through
extensive listening or reading and their main focus is on understanding the message being conveyed rather than on the new words encountered (Nagy, Herman and Anderson 1985; Sternberg 1987). Extending this notion to FL or L2 acquisition, researchers have investigated to what extent adults, especially university students or the equivalent, learn vocabulary in different reading contexts (see Pitts, White and Krashen, 1989; Mason and Krashen, 1997; Horst, Cobb and Meara, 1998; Horst 2005; Pellicer-Sánchez and Schmitt, 2010; Laufer and Rozovski-Roitblat, 2011). Although children’s L2 acquisition is supposed to be closely associated with their L1 acquisition, there has been relatively little interest in how they acquire L2 vocabulary in reading, particularly in classroom settings.

It is generally agreed that pure incidental L2 learning from reading on its own is not as effective as previously expected (Huckin and Coady, 1999; Laufer, 2006; Sonbul and Schmitt, 2010). Reading plus explicit focus on target items, i.e. using vocabulary exercises, has been proposed as an alternative to address the drawbacks in incidental learning by a number of researchers (Paribakht and Wesche, 1997; Hill and Lauffer, 2003; Min, 2008; Laufer and Rozovski-Roitblat, 2011). Vocabulary knowledge is invariably classified as receptive and productive, i.e. retrieving the meaning of a word in listening or reading as opposed to producing it in speaking and writing (Nation, 2001). Vocabulary exercises can be broadly divided accordingly, depending on which type of knowledge is aimed for. Previous studies on reading-based vocabulary exercises typically employ both types of exercises or focus on only one type, whereas very few studies have investigated the learning effect of each separately for either adult or young learners. This teacher-led classroom research can shed light on young learners’ L2 vocabulary acquisition as well as offer practical suggestions for primary school English teachers to make vocabulary instruction more efficient in the most common classroom learning context when reading materials are used.

2. THEORETICAL FOUNDATIONS

2.1. Vocabulary Acquisition through Reading

Similar to L1 acquisition, reading is believed to be one of the most important resources for acquiring L2 vocabulary in an incidental manner. Incidental learning occurs as a result of a number of exposures to unknown items during the reading process, meaning being obtained by either guessing or consulting a dictionary (Hulstijn, 2005). Although the role of reading in second language vocabulary learning is widely accepted, there are several limitations to this approach. First, guessing from the reading context may be inaccurate, time-consuming and cognitively demanding (Huckin and Coady, 1999). Even when a dictionary is consulted, the learner may soon forget more than half of the words (Hulstijn, Hollander and Greidanus, 1996). Studies suggest that incidental learning typically yields vocabulary gains in meaning recognition only but not in production (Pitts et al., 1989; Horst et al., 1998). Secondly, researchers generally agree that vocabulary retention rate is rather low; the average gain ranges from 1 to 5 words per text (Pitts et al., 1989; Knight, 1994; Paribakht and Wesche, 1997; Horst et al., 1998). One of the reasons why the learning process is so slow is that multiple exposures to target words are required. Waring and Takaki (2003) proposed 8 encounters with the target item as the minimum figure to ensure retention. In a similar vein,
Brown, Waring, and Donkaewbua (2008) concluded that 7–9 encounters are necessary for acquisition and to reduce attrition. A recent study by Pellicer-Sánchez and Schmitt (2010) shows that 5–8 exposures to target items are needed for incidental learning to occur and 10–17 can accelerate the learning.

2.2. Text-based Vocabulary Learning Exercises

In response to the main drawbacks associated with pure incidental learning, additional vocabulary activities/exercises are considered an important means for enhancing learners’ vocabulary learning and have been traditionally valued in classroom teaching. In reading, the learner might notice an unknown word and try to access its meaning, but may pay little attention to the word form because it is meaning that is important for text comprehension. While doing vocabulary exercises, learners’ attention can be more directly attuned to the form of target items. There are two types of text-based vocabulary exercises: one is fully contextualized and the other semi-contextualized. The difference lies in the length of the reading context on which the exercises are based.

2.2.1. Fully contextualized reading-based exercises

Given the limitations of acquiring vocabulary through reading only, a number of researchers have suggested an alternative approach: reading plus exercises. Paribakht and Wesche (1997) compared the effects of two different kinds of vocabulary instruction, reading only and reading plus word-focused exercises, ranging from noticing to producing the target item. A group of university students learning English as an L2 participated in the study and experienced both conditions. The Vocabulary Knowledge Scale (VKS), measuring five types of knowledge from no knowledge to being able to use the word both grammatically and semantically correctly in a sentence, was used to measure the vocabulary retention. It was found that, although both approaches resulted in important gains in learners’ vocabulary knowledge, the reading plus condition yielded a greater vocabulary retention. More specifically, the words learned in the reading only condition tend to remain at the recognition level, whereas many words learned in the reading plus condition moved to a higher level of knowledge, e.g. from receptively known to productively known. Min (2008) adopted a similar procedure and obtained similar results.

A recent study by Laufer and Rozovski-Roitblat (2011) adopted a more complex design in which the effect of both reading type and number of word occurrences was measured. There are two task conditions: the “focus on form” condition is to expose readers to the target words through reading, and the “focus on forms” condition is reading exposure plus different word-focused exercises. The target words occurred 2–3, 4–5, 6–7 times in the reading and exercises. Only the word meaning was tested in this study, showing that the reading plus word-focused exercises outperformed the reading alone condition when words occurred four times or more in both meaning recognition and call. However, even when readers were exposed to the target words 6–7 times, the maximum meaning retention was only 63.5% for recognition and 34.5% for recall. This shows that reading plus a considerable number of encounters with the target words in word-focused exercises mainly leads to meaning recognition and less favorably to meaning recall. It is assumed that the retention rate for form production would be even lower.
2.2.2. Semi-contextualized exercises

Not all vocabulary exercises are based on a fully contextualised reading text. Sometimes the target items to be practised are simply embedded in one or two sentences while the contextual information is much less rich than that in the reading text. For example, Folse (2004) designed three types of such sentence-based exercises: meeting the target word in one fill-in-the-blank exercise, meeting the target word in three fill-in-the-blank exercises and writing a self-created sentence containing the word, all types being productive in nature. A modified VKS was used to measure the knowledge gain of university students, showing that the three-exercise condition was superior to the other two conditions, leading to the conclusion that word retention hinges upon the number of word retrievals while the efforts made in creating sentences, associated with depth of processing, is less crucial.

Webb (2005) compared university students’ receptive and productive learning effects. In the receptive task the learner was given a word pair (L2 with L1 definition) plus three sentences containing the target word, which serve as additional glosses and the learner were asked to learn the L2 words. In the productive task the word pair was followed by a sentence-making exercise. A range of tests were employed to measure learners’ vocabulary gain in form, meaning (basic and associative) and use (receptive and productive). A between-subjects design, in which sufficient time was given to each condition, showed that the receptive condition was superior to the productive condition. A second within-subjects design, in which only the time needed was given to each condition (but longer time for the productive task due to its nature), revealed that the productive condition was more effective. Webb (2005: 50) argued that the second experimental design represents more authentic learning and claimed: “productive learning is superior to receptive learning not only in developing productive knowledge but also in producing larger gains in receptive knowledge.”

2.3. Receptive vs. Productive Vocabulary Learning

Despite the complexity involved in degrees of word knowledge, a number of researchers (Henriksen, 1999; Meara, 1990; Nation, 2001) proposed a two-category framework for vocabulary knowledge, i.e. the reception/production distinction, which can be viewed as a practical pedagogical application in assessing vocabulary rather than as representing how vocabulary knowledge is acquired in reality. A number of studies suggested a few attributes regarding the relationship between reception and production: (1) the receptive vocabulary is larger than the productive vocabulary (Melka, 1997; Laufer, 1998); (2) reception precedes production (Melka, 1997); (3) production is more difficult than reception (Griffin and Harley, 1996; Waring, 1997; Mondria and Wiersma, 2004).

It is generally agreed that the learning of receptive or productive knowledge can be explained by the transfer-appropriate processing (TAP) theory that suggests that words processed receptively result in better receptive knowledge and those processed productively result in better productive knowledge (Bransford, Franks, Morris and Stein, 1979). A number of lexical studies on either individual words (Griffin and Harley, 1996; Waring, 1997; Mondria and Wiersma, 2004; Shintani, 2011) or lexical phrases/idioms (Steinel, Hulstijn and Steinel, 2007) lend support to this view, i.e. receptive learning is more compatible for receptive knowledge and productive learning more suitable for productive knowledge.
In parallel grammatical studies, similar results have been obtained, as shown by Dekeyser (1997), DeKeyser and Sokalski (2001), Tanaka (2001) and De Jong (2005). However, all the lexical studies cited above, except the one by Shintani (2011) that investigated the receptive and productive learning effects, simply employed de-contextualized word-pair list learning as the sole learning activity. More specifically, learning the L2–L1 list is treated as receptive learning as it requires the learner to retrieve a more familiar L1 word cued by an L2 equivalent while learning the L1–L2 list is a productive task which requires the learner to retrieve a less familiar L2 word cued by a more familiar L1 word. In addition, all the learning tasks were completed in rather a short fixed period of time of up to 15 minutes. Such de-contextualized list learning does not represent the authentic way in which most vocabulary is learned, thus being of limited value for instructional pedagogy. Examining the receptive and productive learning effect with more contextualized vocabulary learning activities/exercises would be more meaningful and pertinent. Although the study by Shintani (2011) employed both receptive and productive exercises rather than the list learning, the exercises are either aural or oral in nature but not based on a reading text.

2.4. Research Questions

As we can see, the studies reviewed above mainly involve adults (university students) or young adults (senior secondary school pupils) and a lack of attention to young learners is evident. If reading plus exercises can prove to be effective with pupils, teachers are provided with a common yet very useful instructional pedagogy as an alternative to teaching vocabulary. In the current study, following fully contextualized reading, different types of vocabulary exercises are designed for L2 young learners to practise target items. Receptive exercises include giving or matching definitions for new items while productive exercises comprise gap filling and sentence writing. This study seeks to provide answers to the following research questions:

1. Do reading-based exercises help L2 young learners acquire new vocabulary?
2. How do receptive learning and productive learning affect L2 young learners’ vocabulary retention?

3. Method

3.1. Participants

The participants in this quasi-experimental research were a class of 25 primary three (P.3) pupils taught by one of the researchers in a Hong Kong primary school where Chinese is used as the medium of instruction except for English courses. There were 14 boys and 11 girls, aged 8 to 9. Most of these pupils would have learned English as their first L2; they usually had an English lesson of half an hour to an hour every weekday. Due to practical constraints, it was difficult to access other groups of pupils of the same level, so a within-subjects design was adopted, i.e. each pupil experienced both learning conditions (reading followed by receptive or receptive + productive exercises).
3.2. Instruments

There were two types of instruments in this research: four short reading texts with vocabulary exercises developed by the researchers as the instructional instrument (the independent variable) and a test adapted from the Vocabulary Knowledge Scale (VKS) as the testing instrument (the dependent variable). The four reading texts were chosen from supplementary learning materials which were semantically related to the topics of the textbook used by the pupils during the experimental period. The texts were slightly modified (e.g. change of complex sentence structures or substitution of difficult words) to ensure that pupils could understand these texts. Each of the texts was about 100 words long, matching the normal length of the texts used in their textbooks. 20 new content words (nouns, verbs and adjectives) were inserted or substituted in the four texts, 5 words being in each text. All these words were assumed to be unknown to the pupils based on the teacher’s knowledge of the pupils. In addition, all these words, except one, occurred beyond the first 1000-word frequency. No pre-test was given which might alert pupils to the nature of the study, which in turn might affect the results.

The Vocabulary Knowledge Scale (VKS) developed by Wesche and Paribakht (1997) was adapted and translated into the L1 to measure pupils’ vocabulary development in the current study. The original 5-point scale was modified to a 3-point scale for two reasons. First, as Bruton (2009) has pointed out, some of the scales may not be well differentiated, e.g. “I think I know it…” (Scale III) and “I know this word…” (Scale IV). Secondly, the participants were fairly young (8–9 years old) and not cognitively mature enough to distinguish the 5 statements from each other. Due to the tight teaching schedule to follow the required school curriculum, pupils could not finish the test on all the 20 items in one lesson (35 minutes), only 16 target items appeared in the test, i.e. 1 out of 5 target items in each text was randomly removed. See Table 1 for the summary of word class information of the 16 words tested in both conditions. Previous research indicated that word class might be a variable that affects word learning difficulty but how exactly is unclear (Folse, 2004). Laufer (1990) tentatively suggested that, among all content words, nouns are the easiest to learn followed by verbs and adjectives while adverbs are the most difficult. Mainly two word classes – nouns and verbs – occurred in the test and they were more or less equally allocated to each condition. See Table 1 for the words and their word class information. After calculating the scores for each participant, a reliability test yielded a fairly high Cronbach Alpha: .914, which is higher than the figure (.82 – .89) reported for the original version by Paribakht and Wesche (1997). This indicates that the test had been appropriately adapted and was fairly reliable. See below for an example of a test item translated into English.

Harvest
1. I don’t know this word.
2. I know this word. It means _________________________ (Chinese translation or English definition).
3. I know how to use this word to make up a sentence.
Table 1. Information for words tested in each condition

<table>
<thead>
<tr>
<th>Word class</th>
<th>Receptive learning</th>
<th>Productive learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>5 (harvest, lantern, carnival, exhibition, almond)</td>
<td>5 (pumpkin, witch, recipe, bacon, lettuce)</td>
</tr>
<tr>
<td>Adjective</td>
<td>0</td>
<td>1 (wonderful)</td>
</tr>
<tr>
<td>Verb</td>
<td>3 (peel, slice, sprinkle)</td>
<td>2 (threaten, scare)</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Following a similar modified VKS adapted by Folse (2004), pupils’ knowledge of each vocabulary item in the current study was scaled from 0 to 2. “0” is the lowest scale, i.e. failing to recognize the word that had been encountered previously. “1” would be awarded if the pupils could provide the word meaning (in L1 translation or L2 definition) but failed to produce a sentence. “2”, the highest mark, would be awarded to those who could make a sentence using the vocabulary items correctly as well as provide the correct meaning. Later it was discovered that these pupils often could not produce fully grammatically correct sentences. The degree of appropriate meaning expression determined by the collocation involving the target word was the focus and not the grammar of the whole sentence. 2 marks would still be awarded if the meaning could be correctly expressed in the collocation, regardless of other grammatical mistakes in the sentence.

3.3. Procedure

This research was conducted over a period of 5 weeks in 5 lessons of 35 minutes each. Two reading-based methods were used to help pupils learn the 20 new vocabulary items. After they had read the texts, they either learnt through receptive learning or through productive learning in different weeks.

In the first week, the receptive learning method was employed. Pupils read a short passage about Festivals and then they were asked to answer three comprehension questions to make sure that they had understood the meaning of the passage. After that, they were asked to read the 5 underlined new words in the text. Pupils were asked to provide the Chinese meaning for each target word in the exercise sheet with the teacher’s help. After elicitation of initial guessing of the meanings, pictures shown on PowerPoint and actions performed by the teacher were used to illustrate the meaning of the target items. Then the Chinese definition for each item was shown on the PowerPoint to ensure the correct meanings were understood by the pupils. They were told to match the 5 target words with the corresponding English explanations. Pupils’ answers to the exercise were then checked by the teacher.

One week later, pupils were asked to read the second passage, also about Festivals, but this time the productive learning method was adopted after reading. After reading the passage, pupils needed to answer the comprehension questions and provide the Chinese meaning of the 5 target items. Instead of doing the matching definition activity as in the
receptive condition, they did two productive exercises: gap filling and sentence writing using each new item. These two exercises were used to focus on the productive aspect of vocabulary knowledge. The productive learning condition used one additional exercise and the time needed for completing all tasks was in average 5-6 minutes more than the receptive learning condition. As it is generally agreed that productive learning is much more difficult than receptive learning (Griffin and Harley, 1996; Waring, 1997; Mondria and Wiersma, 2004), it is understandable that the productive learning condition used a little more time than the receptive learning condition.

In the third week, pupils read a passage on food and the same procedure carried out in week 1 was repeated. In the fourth week, pupils read another passage on food and the procedure in week 2 was then repeated. In week 5, an unannounced delayed post-test adapted from the VKS was distributed to pupils who were required to complete the test during the 35-minute lesson. Since the test was administered one week after all instructional sessions, it is fair to assume that the test score represented stable, long-term vocabulary retention.

3.4. Data analysis

All pupils’ scores were entered into SPSS for statistical tests. A paired samples $t$-test was carried out to measure the within-subjects effect of the two conditions. In addition, the pupils’ vocabulary knowledge gain was classified into three types: no knowledge, receptive knowledge and productive knowledge, based on which a chi-square test was performed. These two statistical tests aim to answer the first and second research questions.

4. Results

4.1. Statistical tests

Table 2 presents the results of the test, showing the mean score and the standard deviation of the vocabulary retention score for each condition.

*Table 2. Vocabulary retention of the two conditions*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean</th>
<th>Retention rate*</th>
<th>Standard Deviation</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>receptive learning</td>
<td>7.68</td>
<td>48%</td>
<td>3.67</td>
<td>16</td>
</tr>
<tr>
<td>productive learning</td>
<td>11.64</td>
<td>73%</td>
<td>4.21</td>
<td>16</td>
</tr>
</tbody>
</table>

*Retention rate is obtained by dividing the mean score by the maximum score.

From Table 2, it can be seen that both conditions yielded overall satisfactory vocabulary retention (7.68 – 11.64 out of 16) when reading was followed by exercises, thus providing
the answer to the first research question, i.e. do reading-based exercises help L2 young learners acquire new vocabulary? In order to answer the second research question, i.e. how do receptive learning (with only receptive exercises) and productive learning (combining both receptive and productive exercises) affect L2 young learners’ vocabulary retention? a comparison of the mean scores show that score for receptive learning (7.68) is considerably lower than that for productive learning (11.64). Accordingly, the retention rate is much higher for productive learning (73%) than for receptive learning (48%). This suggests that pupils, in general, retain more vocabulary knowledge in the productive learning condition than in the receptive learning condition. A paired samples t-test shows that the difference is significant (mean difference = 3.96; p < 0.001). The Cohen’s d is 1.01, showing a large effect size. It is evident that pupils performed much better in the test items after doing both receptive and productive exercises than only the receptive exercises.

The second research question can also be answered by looking at the type of vocabulary knowledge gained in each condition. All answers were classified into three types: unknown (i.e. scored “0”), receptive (i.e. scored “1”) and productive (i.e. scored “2”). See Table 3 for the results. This shows clearly that when receptive exercises are used, students tend to obtain receptive knowledge (52%) but fall short of productive knowledge; by contrast, when both receptive and productive exercises are used (productive learning), productive knowledge is more likely to be obtained (59%). A chi-square test indicates that the distribution patterns for both conditions differ significantly from each other [χ² (2, N = 400) = 56.87, p < 0.001]. See Figure 1 for the different distribution patterns for the receptive learning and productive learning conditions. Notably, the productive learning condition had fewer instances of unknown and receptive knowledge but far more for productive knowledge than the receptive learning condition. If we average the knowledge gain for all pupils, the productive learning condition resulted in 4.64 (118 divided by 25 participants) words productively for each participant while the receptive learning condition resulted in only 1.76 words productively out of the total of 8 words. The number of words that were learned productively in the productive learning condition was almost three times that in the receptive learning condition.

Table 3. Knowledge types in vocabulary retention between the receptive learning and productive learning conditions.

<table>
<thead>
<tr>
<th></th>
<th>receptive learning</th>
<th>productive learning</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>52 (26%)</td>
<td>26 (13%)</td>
<td>78 (19.5%)</td>
</tr>
<tr>
<td>Receptive</td>
<td>104 (52%)</td>
<td>56 (28%)</td>
<td>160 (40%)</td>
</tr>
<tr>
<td>Productive</td>
<td>44 (22%)</td>
<td>118 (59%)</td>
<td>162 (40.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>200 (100%)</td>
<td>200 (100%)</td>
<td>400 (100%)</td>
</tr>
</tbody>
</table>
5. DISCUSSION

5.1. Effectiveness of reading plus exercises on young learners

Previous research has demonstrated the effectiveness of using reading-based exercises on adult or young adult learners’ L2 vocabulary acquisition (Paribakht and Wesche, 1997; Min, 2008; Laufer and Rozovski-Roitblat, 2011). The results of our study have shown that reading-based exercises can also be effective for young learners and the retention rate is quite satisfactory, pupils being able to retain 48% – 73% of the vocabulary in the delayed post-test. Thus the current research adds to our understanding of the usefulness of reading-based exercises in terms of younger learners’ L2 vocabulary acquisition. This is encouraging news for the vast number of primary school teachers in that the reading-based vocabulary instructional method can also work for young learners in primary schools as an alternative to other popular methods for teaching vocabulary such as repetition or dictation. As demonstrated in this study, the teacher played a key role in guiding and instructing pupils during the whole learning process; it was the teacher who drew pupils’ attention to the unknown items in the reading texts and helped them understand the word meaning as well as practise using the words with various exercises. Young learners, who are not cognitively mature, need clear instructional guidance from the teacher when they are required to learn new vocabulary items from reading-based lessons.

5.2. The effects of reading-based receptive and productive learning

This research also investigated how reading-based receptive and productive learning, implemented by different types of exercises, affect pupils’ vocabulary retention and knowledge
types. Both the mean scores and the paired samples $t$-test show that the productive learning condition outperformed the receptive learning condition in the delayed post-test, the score for the former being almost twice that of the latter. The chi-square test demonstrates clearly that when both receptive and productive exercises follow reading, this is much more likely to lead to productive knowledge than receptive knowledge alone; each pupil has on average learned almost three times more words productively in the productive learning condition than in the receptive learning condition. This suggests that pupils generally have acquired many more vocabulary items and far better productive knowledge in the productive learning condition than in the receptive learning condition. This may be due to the fact that pupils had to use the target items in gap filling or sentence making in the productive exercises; they were forced to pay attention to various lexical aspects, such as spelling and collocations, in addition to meaning. Our study employed contextualized reading to investigate the effects of receptive and productive learning; the results corroborate what Mondria and Wiersma (2004: 98) suggested: “When the learning aim is both receptive and productive word knowledge, it is recommended to learn the words both receptively and productively”. The results of our study are also in line with those of Webb’s (2005) second experiment in which a within-subjects design (the same as in our study) was used; Webb’s university participants demonstrated more satisfactory vocabulary retention in the productive condition than in the receptive condition. If Mondria and Wiersma (2004) and Webb (2005) supported the transfer-appropriate processing (TAP) in de-contextualized learning studies, our study lends evidence to support the same view in fully contextualized reading-based vocabulary learning. In addition, our research focused on much younger L2 learners and demonstrated convincingly that they can benefit more if productive learning follows contextualized reading. Primary school teachers are encouraged to adopt both receptive and productive exercises after reading rather than simply receptive exercises which are likely to lead to vocabulary recognition only.

5.3. Limitations of the study

Due to the practical constraints faced by the current study, teachers or researchers who would like to conduct similar research in future need to be alerted to a number of limitations. First of all, this is a small-scale investigation and the results may be biased due to the small sample size. A study involving more participants is desired in order that the results may be generalized more confidently. Secondly, time is not strictly controlled in this quasi-experimental study, i.e. pupils in the productive learning condition spent slightly more time completing all the exercises than in the receptive condition. On the other hand, this perhaps reflects the different nature of receptive and productive learning, namely, the latter is more difficult and hence needs a little more extra time. Our primary concern is not to compare and find out which type of exercise, receptive or productive, is more effective, as in a strictly controlled experiment, but to offer teachers a clear understanding of how receptive and productive exercises can complement each other to promote pupils’ vocabulary acquisition to a productive level in a real classroom situation.

In addition, it is found that pupils made grammatical mistakes frequently when using the target items to write sentences. Although some of the sentences were comprehensible, they contain grammatical mistakes to varying degrees. This suggests that not all primary three
pupils (at least in our sample) were able to produce fully grammatically correct sentences when they attempted to use the newly learned item, possibly due to their under-developed grammatical system. In this sense, asking them to make up a sentence with the target word may not be the ideal way to test their productive knowledge. Despite a rather high reliability reported by the modified version of the Vocabulary Knowledge Scale (VKS) used in the current study, it is tentatively suggested that the highest scale – sentence writing – is not suitable for young learners or low level learners and may be replaced by collocating the target words.

6. Conclusion

Using five naturally occurring lessons for a group of Primary three pupils, this teacher-initiated classroom study documented how reading-based lessons could be conducted to teach young learners L2 vocabulary from meaning recognition to using the word in a sentence context. To ensure pupils’ mastering of target items to a productive level, both receptive and productive exercises are needed. Young learners at this stage have great difficulties in learning and using verbs accurately. Such difficulties can be partially attributed to learners’ L1 influence (as in this study) and partially to the L2 verbs’ own complexity. Considering young learners’ relatively lower cognitive capacity than that of adult learners, teachers should teach verbs in chunks as well as create opportunities for young learners to meet the target word in contexts different from the original context. Attention should be paid to variant morphological forms of verbs and the context in which they should be used. Reading-based vocabulary text exercises prove to be an effective way to help young learners learn vocabulary, but the importance of more communicative oral activities/tasks in facilitating vocabulary use should not be underestimated. Future studies should focus on how text-based exercises and oral activities can be synergized to promote young learners’ L2 vocabulary learning.

7. References


Researching and teaching China and Hong Kong English

DAVID C. S. LI

Issues, problems and prospects

The English curriculum in China – including the Hong Kong Special Administrative Region (HKSAR) – has traditionally been dominated by native-speaker (NS) based pedagogical models. This is a source of many problems, ranging from learning outcome to teaching performance, and from cultural inappropriacy to speaker identity. Research in World Englishes (WE), in English as a lingua franca (ELF) and an international language (EIL), and to a lesser extent in second-language acquisition (SLA) has shown that a curriculum informed by a deficit model (by measuring learner performance using the yardstick of native-speaker-based standards) is by its very nature disempowering, and should be replaced with a model of difference, whereby learners’ L1 identities and ownership of English are both respected.

Introduction

To put theory into practice – and let WE and ELF insights inform local English curricula – much remains to be done, including basic research in local(ized) English features across all linguistic levels. The main challenge is intelligibility in international communication, currently the focus of intensive ELF research. Identity, ownership and intelligibility are important issues that bear principally on social interaction, and so should receive more attention in the teaching of speaking and listening. Learners’ needs that relate to reading and writing are rather different. Contrary to such radical proposals as replacing the NS-based pedagogic model with local(ized) models, it is argued that Standard (written) English lexicogrammar (hereafter Standard English) is indispensable.

As Trudgill (2002:165) has noted, Standard English, with all its idiosyncrasies (pp. 167–68) is ‘by far the most important dialect in the English-speaking world from a social, intellectual and cultural point of view; and it does not have an associated accent’. Its role in writing for speakers of all dialects is ‘unassailable’ (p. 169). In the context of China, learners need it to absorb information in this ‘social dialect’, particularly in printed sources, notably including the Internet. It is also crucial for upward and outward mobility. True to the spirit of empowerment, therefore, Standard English should remain an important part of a newly restructured curriculum in the territory.

China English

Without doubt, when compared with other nations, China has the largest number of learners of English in the world. McArthur (2003:22) estimates the number of its English speakers to be around 300 million. The fact that English is now an important medium of instruction in schools has generated a large number of learners of English. In addition, the high number of learners of English in China is also due to the fact that English is a required subject in Chinese higher education institutions. As a result, Chinese learners of English face many challenges in learning this language, such as the pressure to achieve high scores in English exams, the need to learn a large vocabulary, and the difficulty of mastering the grammar of English. To address these challenges, many Chinese institutions have implemented various strategies to improve the teaching and learning of English, such as offering more resources for learners, providing more opportunities for practice, and incorporating more language learning technologies.

DAVID C. S. LI is an associate professor at the Department of English and Communication, City University of Hong Kong. He obtained his BA (English) in Hong Kong, MA (Linguistics and Applied Linguistics) in France, and PhD (Linguistics) in Germany. He has published in three main areas: Cantonese–English code-switching, Hong Kong English vis-à-vis English as a global language, and Hong Kong Chinese EFL learners’ learning difficulties and correction strategies. He is currently working on two funded projects: comparing the motivations of Chinese–English code-switching in Hong Kong and Taiwan; investigating the preferred norms and pedagogic model(s) of English in Hong Kong.

<http://personal.cityu.edu.hk/~endavidl/>
users as between 200 and 300 million, which is comparable to the number in India (around 250 million, p. 21). In recent years, ‘China English’ has been ‘attracting growing attention from linguists, researchers and educators’ (Hu 2004:27). Conceived of as a distinct new variety of English, China English has its own distinctive features, ranging from phonology to syntax and from lexis to discourse pragmatics: features characterized above all by cross-linguistic influence from Chinese. As such, China English is seen as an autonomous variety not to be confused with ‘Chinglish’, a pidginized variety situated at the other end of the lectal continuum (cf. Chinese pidgin English: Bolton 2003).

The pedagogic model of English in China is currently dominated by NS-based models, especially American English (Hu 2004, 2005; Jin 2005). In line with the stance of WE and ELF scholars, more and more ELT researchers and teaching professionals in China are in favour of developing localized norms of English, with the goal of ultimately replacing the NS-based models with a local model. Their stance is justified by a number of inter-related facts and observations similar to those in the recent WE and ELF literature:

1. **NS-based competence is not needed:** most Chinese learners learn English for intranational purposes, or to interact with other NNSs in the region. A NS-based pedagogic model, especially pronunciation, is therefore largely irrelevant to their needs; For the majority of learners, a NS-based pedagogic model is unattainable: ‘If L2 learners feel that the chief measure of L2 success is passing for native, few are going to make it. Both teachers and students become frustrated by setting themselves what is in effect an impossible target’ (Cook 2002:331). So from the point of view of learning outcome, the continued use of a NS-based pedagogic model is impractical.

2. **The majority of Chinese learners learn English from local teachers:** most Chinese teachers of English learned English as a second or foreign language; their English is naturally characterized by cross-linguistic influence by their mother tongue to different extents, which should be seen as a resource rather than a deficit (Cook 2002).

3. **Chinese English teachers would feel greatly empowered:** using an ELF or local(ized) pedagogic model has great potential for liberating and empowering Chinese English teachers by allowing them to capitalize on their strengths which are usually not shared by native-speaking teachers, namely their sound knowledge of the learners’ language and culture (Braine 1999, Medgyes 1994).

4. **Cultural elements in the curriculum should reflect learners’ needs:** aspects of foreign culture are remote to learners and users of English in China; most learners need English to talk about aspects of their own culture rather than Anglo-American culture (Kirkpatrick 2006; McKay 2002, 2006).

5. **Learners will become confident owners of China English rather than frustrated imitators of a NS-based standard:** whereas targeting a NS-based model typically produces failure and results in frustration and disappointing learning outcomes, using China English as the pedagogic model empowers its users by enhancing their purpose of learning English, and boosting their confidence and sense of ownership of a local(ized) variety that better reflects their communicative needs.

On the basis of these facts and observations, researchers in favour of China English appeal for concrete action, to codify the localized English features and have them incorporated into the national English curriculum (Hu 2004, 2005; Kirkpatrick 2006; cf. Kirkpatrick & Xu 2002; Li 1998). Awareness-raising regarding World Englishes is seen as an important strategy to promote the legitimacy of non-native varieties of English, and, at the same time, reduce the linguistic capital of NS-based models of English (Jenkins 2006:174; Jin 2004).

Recent research on China English is encouraging, as more and more local Chinese teachers of English have heard of World Englishes and identify with the basic premises associated with it (Hu 2005), although more work needs to be done in this regard to raise the consciousness of learners (Hu 2004, Jin 2004). The key is to make learners aware that non-native features such as their accent, and the accents of speakers of other varieties of the language, are perfectly legitimate, in the same way as people in English-L1 countries such as Ireland and Scotland take pride in their local standard varieties. Apart from the selection and codification of local linguistic features with widespread appeal, and concerted efforts to raise learners’ awareness of varieties of English other than NS varieties, winning official support is also important (Hu 2005:38). Teachers’ efforts to expose learners to linguistic and discourse-pragmatic features of different varieties of
English will be in vain if learners do not regard this as a worthwhile exercise. Clearly, such initiatives will not bear fruit without first finding their way into the national English curriculum, which in turn is impossible without the blessing of the Ministry of Education.

**Hong Kong English**

Hong Kong Special Administrative Region (HKSAR) was de-colonized and re-nationalized as a Special Administrative Region of China on July 1, 1997. To my knowledge, Bolton (2003) provides the most systematic and coherent account of ‘Hong Kong English’ (HKE). His approach, which is sociohistorical and political (supplemented by econocultural analysis) traces the development of HKE back to the early days of colonization. Bolton argues that, despite international recognition of the autonomy of HKE (as shown in certain references since the 1980s), locally based academics have tended to regard English as an untypical second language ('auxiliary language', Luke & Richards 1982; ‘value-added language’, Li 1999) and, as such, it has lacked the sound sociological basis necessary for an autonomous variety.

Bolton maintains, however, that previous reservations about according Hong Kong English the status of an autonomous variety are no longer valid. The most often cited argument against such autonomy is the reluctance of Hong Kong Chinese bilinguals to use, or switch to, English entirely, for intra-ethnic communication – unlike Chinese Singaporeans in this regard (cf. Li 1999). Bolton argues, however, that social change in the last two decades has rendered the need for such a switch obsolete. There are five reasons for this:

1. Census figures in the last two decades show that the number of people claiming to be able to speak English as a ‘usual’ or ‘another’ language has increased from 9.7% in 1961 to 43% in 2001 (Figure 2.3, p. 87).

2. For over a decade, many households in Hong Kong employ an English-speaking domestic helper from the Philippines (among other countries, such as Indonesia), of which there are over 160,000 at the present time, allowing English to penetrate into the home domain of many Hongkongers.

3. There has been a sizeable number of returnees from English-speaking countries after 1997 (such as emigrants who have secured foreign passports, and students returning after studying in an English-speaking country).

4. As a correlate of the last point, many Hongkongers have friends and relatives in English-speaking countries, which gives them plenty of ‘natural’ opportunities to use English.

5. English is widely used by educated bilingual Hong Kong Chinese for electronic communication, such as emails, chat rooms, and ICQ ['I seek you'].

On the basis of such observations, Bolton (2003) believes that the ‘essential conditions necessary for the emergence of such a variety may already be present in the community’ (p. 116). He also reviews a number of studies of language attitudes and cites some studies of syntactic structures that are specific to HKE. He argues that the body of recent research on Hongkongers’ attitudes toward English convincingly shows that, unlike similar studies prior to the 1980s, speaking English no longer makes them feel un-Chinese (Pennington & Yue 1994), while studies of syntactic structures reveal a number of HKE syntactic characteristics, widely perceived by local academics as ‘errors’ rather than local(ized) innovations. Bolton further provides a detailed analysis of the relevance of Butler’s (1997:106) five criteria for determining the emergence of a localized variety of English (see also Bolton 2003, ch. 4). On the basis of such evidence, Bolton argues that HKE is gradually evolving in the same direction as Philippine English, a more fully nativized variety also in the Outer Circle.

Bolton’s (2003) position is manifestly not shared by some colleagues in the field of English-language teacher education. The preference for a pedagogic model of English in Hong Kong, from the points of view of various stakeholders, has been the focus of research by Amy Tsui and David Bunton (Bunton 2000) who analysed discussions held on TeleNex, a computer network that supports the work of local teachers of English. After analysing 1,234 teachers' messages on the e-forum ‘Language Corner’, Tsui and Bunton found that those resources which teachers looked on as authorities of correctness and appropriateness were reference works such as dictionaries, grammar books, and guides to good usage, all of which were based on native-speaking and especially British models.

In contrast, locally published textbooks and specific usages heard in the local media (such
as ‘shopaholic’) were received with suspicion or sometimes criticized. Furthermore, the authors note that there was no mention of ‘Hong Kong English’ as such; rather, the term ‘Chinglish’ occurred twice in reference to student errors. Tsui and Bunton therefore conclude that English is perceived as exonormative by local teachers of English – not endonormative.

In a separate study, Bunton and Tsui (2002) analyse a body of documents from various sources (including the HKSAR government, business leaders, and the examinations authority), descriptors of benchmarks being developed for English teachers and the teaching profession at large. They again found that the implicit standards were largely exonormative. Only benchmark developers (cf Coniam & Falvey 2002) distanced themselves from a native-speaking model, favouring an ‘educated Hong Kong model’ on which teaching could be based. Even so, however, remarks relating to ‘grammatical accuracy’ at Level 5 (the highest level) frequently make reference to ‘errors’, while descriptors of ‘speaking’ and ‘classroom language assessment’ at Level 5 specify that the candidate’s English should be ‘completely error-free with no noticeable first language (L1) characteristics’ (p. 70).

Very similar findings were obtained in Andrews’ (2002) small-scale survey of 98 in-service, mostly NNS secondary-school English teachers’ views regarding the most appropriate choice of pedagogic model in Hong Kong. All this suggests that, even though the benchmark developers purposely distanced themselves from NS-based models, in practice the ideal speaker of ‘Hong Kong English’ should possess native-like competence regardless of linguistic levels.

In sum, while advocacy by WE scholars of recognition of HKE as an autonomous variety has potential merit (arguably both acquisitional and socio-psychological) in terms of empowering local learners of English, empirical evidence to date suggests that their advocacy is not welcomed by such key stakeholders as local teachers of English, business leaders, the examinations authority (including the benchmark developers), and indeed the government itself. In addition, parents are likely to resist the idea of using a non-native model of English in school, thinking that it is much less prestigious compared with a NS-based model. And the question arises: What about the learners themselves, the end-users or consumers of the English taught to them?

At present, WE scholars seem to assume that replacing the NS-based model with a local(ized) model will be well received by English-L2 learners. But this tacit assumption seems to have been called into question by the results of several surveys (see: Hu 2004; Jin 2005; Kirkpatrick & Xu 2002; Timmis 2002; Li forthcoming). Is there a way to reconcile the two positions using a sound curriculum that would empower both local learners and teachers of English by making it possible for them to benefit from ‘the best of both possible worlds’? This paper is an attempt to address this question.

**Toward a sound curriculum of China English and Hong Kong English**

The starting point for a sound curriculum as proposed here is a list of language-learning goals compatible with the needs of English-L2 users in both mainland China and Hong Kong. In general, the EIL/ELF goals should include the following competency areas, among others:

1. **Maintaining speaker identity:** communicating with others effectively (whether L1 or L2 users of English) – without compromising their L1 identity
2. **Ensuring intelligibility in ELF communication:** to convey speaker meanings that are maximally free from intelligibility problems
3. **Developing a sense of ownership and pride in the local(ized) variety:** to make learners proud of and confident in the local(ized) variety
4. **Being equipped with Standard English (cf. Trudgill 2002) as a prerequisite for life-long learning:** to be literate in, and conversant with, lexi-co-grammatical features of the written standard variety in order to absorb all kinds of information in print or on the Internet.

Of these areas, 1 to 3 require insights and input from research in WE and EIL/ELF. Regarding 1, as many have observed, if the covert norms governing pronunciation, lexi-co-grammatical choices, and discourse-pragmatic patterns of social interaction are locally based rather than NS-based, there is no pedagogically sound reason why English-L2 learners should be expected to conform to NS-based norms, especially pronunciation skills. This is where input from well-researched and codified local(ized)
linguistic features may inform the local English curriculum.

Goal 2 is more challenging, largely because research on intelligibility in ELF is still at its embryonic stage. Given what we know (for Lingua Franca Core features in phonology, see Jenkins 2000; for a provisional list of lexico-grammatical LFC features, see Seidlhofer 2004:220, Jenkins 2006:170), to achieve intelligibility in ELF, the speaker should not only rely on a set of linguistically well-defined LFC features, but should be ready to use meta-communicative strategies such as speech accommodation and paralinguistic cues.

It might be difficult to teach these strategies comprehensively in class, but so long as the language of accommodation and the effective use of paralinguistic cues are included in the learning objectives, learners will appreciate that mutual intelligibility depends crucially on a readiness for both sides to seek clarification with each other when communication seems to be impeded by unfamiliar phonological and/or lexico-grammatical features in the interlocutor’s variety. Intelligibility is obviously one area in which ELF research using the corpus-based core approach has much to offer. This has significant implications for Chinese researchers who support the goal of developing and codifying local(ized) norms in China English and Hong Kong English.

Goal 3 amounts to inviting English-L2 learners to consider abandoning the deep-seated link between a NS-based model and linguistic capital (Jenkins 2006). Such an association between NS-based competence and its perceived instrumental value is typically inculcated into the psyche of learners through repeated teacher admonition – that any linguistic features which deviate from the ‘standard’ are ‘errors’ in need of correction. To counteract this mentality, one important strategy is to raise English-L2 learners’ awareness of the legitimacy of other varieties of English than the ‘NS-based standard’ (Jenkins 2006; Kirkpatrick 2006). This objective could be achieved both implicitly and explicitly. The explicit move consists of including a few marked and well-researched linguistic features of specific regional varieties of English in the curriculum for awareness-raising purposes. At the same time, where possible, the presence of non-native English-speaking teachers (non-NESTs) from different L1 backgrounds is one implicit means to bring home the message that there exist regional varieties of English which are clearly different from the NS-based models but which are no less valuable. This is clearly one area where research into specific features of individual varieties of English may inform the local curriculum.

The focus of goals 1 to 3, while not exclusively confined to speaking and listening, is concerned primarily with verbal communication in speech. Goal 4, on the other hand, is closely related to reading and writing, that is, English for academic purposes (EAP) or Standard English in short. “The print standard is the first thing many people think of when issues of ‘Standard English’ are discussed” (McArthur 2001:6). Since the bulk of formal information conveyed in writing, both in print and on the Internet, is written in Standard English (Trudgill 2002), it is too valuable a tool to be dispensed with, as is often implied in the debate on NS-based vs localized English curriculum.

The argument that most English-L2 learners use English primarily with other English-L2 users is well taken. However, given that we are living in an increasingly globalized world, it may be short-sighted to remove Standard English from the local English curriculum. Standard English may be difficult to define (Seidlhofer 2005; Trudgill 2002), but there are sufficient and well-documented lexico-grammatical commonalities across the regional standard varieties (Gupta 2005, 2006; Trudgill & Hannah 2002) that make the task of teaching a standard variety (and its variants) reasonably well defined and delimited. It is the learning that is less predictable, and this is a second reservation, that relatively few English-L2 learners are able to fully master Standard English lexico-grammar through classroom input and instruction.

This is true to a large extent (Cook 2002). But unlike NS-based phonological systems, to withhold Standard English on the grounds that relatively few learners will be able to master it seems to ignore individual learners’ aspiration and possibly aptitude as well. It also goes against the spirit of empowerment (Li 2006): in view of the learners’ possible needs in future, withholding Standard English in the curriculum is tantamount to placing a glass ceiling on how much the learner could hope to achieve.

Further, in light of the general research finding that learners stand the best chance of
developing proficiency in a second language before puberty, such a curricular decision is counterproductive and not in the learners' best interest. Finally, unlike other linguistic areas of competence, grammar is usually one of the strengths that many non-NESTs are proud of and good at doing compared with, for example, the teaching of NS-based pronunciation (Medgyes 1994, Prodromou 2006). If non-NESTs are required to teach a variety of English characterized by marked simplifications, as indicated in the lexico-grammatical LFC features (Seidlhofer 2004), they may feel disempowered for not doing their best to their students' benefit.

Conclusion

In sum, what is proposed here is a radically restructured curriculum that incorporates the strengths and insights of WE and ELF research and the empowering potential of Standard English. Hu (2004: 28) characterizes China English as having, on one hand, a 'common core' which 'renders it as intelligible to speakers of other varieties of English as Hiberno-English or Australian English', and, on the other hand, culture-specific expressions which 'transmit many of the cultural references that Chinese speakers of the language will want to communicate', such as 'the four modernizations'. The same analysis essentially applies to HKE, where a common core is similarly supplemented by an indigenized vocabulary that is incomprehensible to those who are unfamiliar with Hong Kong culture: e.g.  astronaut, for 航天 for 航天 for 航天 (literally 'wife-empty-person'), denoting 'an émigré who lives away from his spouse and family and who often flies between the host country and Hong Kong' (translation adapted from Bolton 2003:288). An important component of Hu's 'common core' is arguably Standard (written) English.

The learners' needs for English mainly fall into four areas, which correspond with four teaching and learning goals in China and Hong Kong. The first three goals essentially concern speaking and listening skills. This is broadly one area of research where insights derived from WE and ELF research may serve an instrumental purpose. The fourth and last goal concerns the need for lexi-co-grammar in Standard English, which is crucial for absorbing information in knowledge-based economies where much of the information in higher learning is written (encoded) in this variety. It is absolutely crucial that Standard English should have a place in the local English curriculum in order that the learner will exit school equipped with a valuable tool for lifelong learning and, where necessary, communicate with English-L1 speakers – even though some learners will achieve this goal far better than others. For a minority of learners, however, it is conceivable that life circumstances might take them to work positions where they assume the role of generators of knowledge, rather than being forever condemned to the consumption end of the knowledge production process.

The curricular blueprint outlined above departs significantly from the current English curricula in both China and Hong Kong, which are dominated by NS-based pedagogic models. It is only possible when many researchers put in concerted efforts to produce empirical evidence of what constitutes a set of sound and widely shared local(ized) features at all linguistic levels: phonology, lexis, grammar and discourse-pragmatics. From where we are to where we want to be is a considerable distance. The goal of this paper is to attempt a theoretically informed road map, hoping to stimulate discussion and debate, and attract ELT researchers and teaching professionals in Greater China to work toward a more enlightened English curriculum which better reflects the ELF needs of the largest group of English-L2 users in the world.

References


Coniam, David & Peter Falvey. 2002. 'Selecting models


—. 2005. 'China English, at home and in the world.' In English Today 21:3, pp. 27–38.


—. (forthcoming). 'Researching NNS’ views toward intelligibility and identity: bridging the gap between moral high grounds and down-to-earth concerns.'


When does an unconventional form become an innovation?

David C.S. Li

Introduction

A lingua franca is needed to facilitate ever-expanding cross-border communication on a global scale. For historical reasons, that role has been and is increasingly assigned to English (McArthur 1998; Crystal 2003; Kirkpatrick 2007), including ‘postcolonial English’ (Schneider 2007). This has direct implications for language education in countries big and small, rich or poor. For the vast majority of ESL/EFL (hereafter: English-L2) learners who have no choice but to study English, typically as a school subject, the coming of age is hardly complete without developing an acute awareness of how important, and yet how difficult, it is to speak and write ‘good English’. English is not at all learner-friendly, especially to learners whose L1 is linguistically unrelated to English (e.g. Altaic languages Korean and Japanese; Sino-Tibetan languages Chinese and Thai). In the learning process, various kinds of cross-linguistic influence from features in the learners’ first language(s) have been shown to be major acquisition problems. Less well-known is the fact that Standard Englishes – the varieties of English being targeted for teaching and learning through education – are fraught with untidiness at different linguistic levels. This is not surprising, given that English, like all natural unplanned languages, evolved over time, rather than being consciously designed for meaning-making purposes – unlike artificial, planned languages such as Esperanto (cf. Li 2003). The untidiness is of two main kinds: (a) inconsistencies in various linguistic subsystems; and (b) considerable variation within each of the standard varieties of English (McArthur 1998; Trudgill and Hannah 2002; Kirkpatrick 2007). These two types of untidiness account for a large number of learner-unfriendly features rooted in standard varieties of English, in particular British English (BrE) and American English (AmE). For practical reasons, we will use ‘Standard English’ to refer to features which are true of one or more standard varieties of English.

In this chapter, I will first illustrate various kinds of learner-unfriendliness by examining some examples of untidiness in Standard English. Non-standard features will be exemplified using data collected from Hong Kong Chinese English-L2 learners and users. The important distinction between errors and innovations will be discussed.
Sources of learner-unfriendliness

Standard English is inconsistent

As a semiotic, meaning-making system, Standard English is inconsistent at various linguistic levels. This is especially clear with regard to orthography and grammar. Take the case of BrE. One of the best known criticisms of irrational English spelling was made by the British playwright George Bernard Shaw in the 1900s. He argued that ‘fish’ might well be spelt as \( \text{GHOTI} \), where the \([f]\) sound of \( \text{gh} \) is attested in a word like \( \text{laugh} \), the \([i]\) sound of \( \text{o} \) in \( \text{women} \), and the syllable-final sibilant \([ti]\) of \( \text{ti} \) in \( \text{nation} \). Another oft-cited example of inconsistent sound-spelling relationship is the various pronunciations (e.g. in RP) associated with \( \text{ough} \), for \( \text{thought} \) \( \text{[˧˧]} \), \( \text{though} \) \( \text{[ou]} \), \( \text{rough} \) \( \text{[˧]} \), \( \text{cough} \) \( \text{[˧]} \), \( \text{drought} \) \( \text{[a˧]} \) and \( \text{thorough} \) \( \text{[锦标]} \). Less eye-catching, nonetheless (or none the less) vexing problems of variation occur across British and American spellings (e.g. \( \text{programme} \) vs \( \text{program} \); \( \text{towards} \) vs \( \text{toward} \)) and word choices (e.g. \( \text{different from} \) vs \( \text{different than} \); see, e.g., Trudgill and Hannah 2002: 85–8; cf. Jenkins 2003: 71–2). No wonder ‘proper spelling’ is sometimes a problem even among English-L1 learners and users.

Paton (2008) reports that ‘Standards of spelling among university students [in the UK] are now so bad that lecturers are being urged to turn a blind eye to mistakes’. Among the high-frequency misspellings are \( \text{arguement (argument)} \), \( \text{Febuary (February)} \), \( \text{Wensday (Wednesday)} \), \( \text{ignor (ignore)} \), \( \text{occured (occurred)} \), \( \text{opertunity (opportunity)} \), \( \text{que (queue)} \), \( \text{speach (speech)} \), \( \text{thier (their)} \), \( \text{truely (truly)} \) and \( \text{twelth (twelfth)} \). A number of principles appear to be at work in these misspellings:

- Silent letters are dropped as spelling reflects pronunciation: \( \text{ignor; Febuary, opertuity, twelth, que, Wensday} \).
- Regularization or simplification: \( \text{truely, arguement, occured} \).
- Orthographic analogy: \( \text{thier (cf. the rule of spelling ‘i before e, except after c’ for the [i:] sound); speach after the productive model of beach, peach, reach, teach, etc.} \)

At the level of grammar, perhaps no other subsystem is more inconsistent than the choice of singular pronouns for designating indefinite reference, which is more or less equivalent in meaning to ‘everyone’ or ‘anyone’. Traditional grammars allow for the use of the male-gender set of pronouns (\( \text{he, him, his and himself} \)) to designate that meaning (e.g. \( \text{let everyone make his own choice} \)). One consequence is that, unlike Buddhists or bird-lovers who can consciously avoid using such unwanted culture-specific idioms as ‘killing two birds with one stone’, a Hong Kong tycoon-philanthropist like Mr Li Ka-Shing could not help being seen as gender-biased in English: ‘“While an individual has the duty to reach his highest potential, to be the best that he can be, in his mind, he must not delude himself to think that he is better than who he really is”, Li said’ (excerpt of speech delivered to all graduates of Shantou University, China; \textit{The Standard}, 27 June 2008: 2).

The original speech was almost certainly delivered in Chinese (Putonghua or the local dialect), which was rendered into Standard English by some bilingual journalist. That journalist should not be blamed for the sexist overtone, however. As Erving Goffman has observed in his celebrated (1981) monograph \textit{Forms of Talk}, unlike other frames of speech such as lecturing or drama performance, the sexist use of male pronouns to express indefinite reference in English (for academic purposes) is a rare sort of frame which is immune from any ‘frame break’.
He who lectures on speech errors and its correction will inevitably make some of the very errors he analyzes…. he who lectures on discourse presuppositions will be utterly tongue-tied unless unself-consciously he makes as many as anyone else…. [This] is not to say that other sorts of frame break might be as clearly doomed; for example, a reference at this point to the very questionable procedure of my employing ‘he’ in the immediately preceding utterances, carefully mingling a sex-biased word for the indefinite nominal pronoun, and an unobjectionable anaphoric term for someone like myself.

(Goffman 1981: 163)

Owing to inconsistencies in the pronominal system in Standard English, the use of he and his to designate ‘anyone’, as shown in this revealing quotation, is ‘unobjectionable’, however ‘questionable’ it might be in the eyes of gender-conscious users of English, including Goffman himself. He or she who feels unhappy about the status quo may try to get around the problem by adopting one of three options: (a) an ‘inclusivist’ stance (as in this sentence, i.e. using ‘he or she’, ‘his or her’, ‘himself or herself’), which sounds clumsy and cumbersome to say the least; (b) a ‘pluralist’ stance (e.g. saying those who do it instead of he who does it); and (c) an ‘exclusivist’ stance, i.e., reversing the discriminatory stance by using the female set of pronouns to designate ‘indefinite reference’, as Cameron et al. have done in their book on critical sociolinguistic research methods, as illustrated in the following example: ‘Circumventing the Observer’s Paradox often involves the researcher in concealing herself and/or her purposes from those she is studying’ (Cameron et al. 1992: 7, emphasis added).

What is interesting is that in some books published in the 1980s, when feminism was on the rise and gendered language use increasingly a concern to sociolinguists, inserting a disclaimer in the front matters was considered a necessary and useful strategy to distance the writer(s) from a perceived gender-insensitive stance. For example:

Finally, whenever I have needed to use a pronoun to refer to the nouns ‘learner’ and ‘teacher’, I have used ‘he’, ‘him’ or ‘his’. This is purely a linguistic convention and does not imply that the person is more likely to be male than female.

(Littlewood 1984: 3)

The need for such a disclaimer is itself strong evidence that Standard English is an untidy system that leaks. Grammatically embedded gender bias is not universal. For example, a sexist orientation is also found in Chinese writing by the male-gendered pronoun 他 (Putonghua/Mandarin tā), but not in speech, for the third-person singular pronouns are pronounced identically in all Chinese varieties (Chao 1968). In French, the choice of singular possessive pronouns (masculine son; feminine sa) depends on the grammatical genre of the common noun rather than the sex of the possessor. Thus the film Chacun son cinéma is rendered into English as ‘To Each His Own Cinema’, a gender bias not found in the original title.

Another inconsistency is the use of the same form to designate semantically incompatible meanings. This is clearly the case of using the same morpho-phonological exponent ‘-s’ (and its allophones and allomorphs) to mark ‘third-person singular’ present tense verb forms, and the plural forms of regular count nouns. Consequently, young English-L2 learners who are taught simple sentences such as Tom likes dogs and Sue likes cats have to grapple with rather different reasons why ‘-s’ is grammatically indispensable:
DAVID C.S. LI

suffixed to the verb like, it is required for marking the ‘third-person singular’ meaning ‘one and only one’; suffixed to the count nouns cat and dog, ‘-s’ is needed for signaling the meaning ‘necessarily more than one’. Since the two meanings are mutually exclusive, such a semantic discrepancy amounts to logical inconsistency. No wonder in the learning process, the ‘third-person singular’ and the plural morpheme are among the most slippery grammatical subsystems for English-L2 learners. This is empirically supported by research in ELF communication: detailed analysis of the Vienna Oxford International Corpus of English (VOICE) shows that the ‘third-person singular’ tops the list of emerging ELF lexico-grammatical features (e.g. you look very sad, he look very sad, Seidlhofer 2004, 2005; see also Breiteneder 2005, 2009; Jenkins 2003: 131; for the use of singular noun forms where plural forms are preferred in Standard English, see example 8 below).

Considerable variation in Standard English

Another source of learner-unfriendliness is considerable variation internal to Standard English. Despite being the most highly codified varieties, there continues to be considerable variation within Standard English. Thus the gradual demise of the subjunctive as a verb form (e.g. we suggest that she go) has reached a stage where it is generally seen as a stylistic variant of the verb phrase marked with should (e.g. we suggest that she should go). Guided by the principle of regularization, the explicit marking of this modal function or meaning using ‘should’ is a welcome development.

Another example of variation in Standard English is the prescriptive rule against ‘dangling modifiers’. Accordingly, in a complex sentence made up of two clauses – the first one a dependent (subordinate) clause with no apparent subject, the second one an independent (main) clause – the subject in the independent clause (overt or covert) should also be the antecedent of the missing subject in the dependent clause. This rule is for instance not respected in (1) (source: http://personal.cityu.edu.hk/~encrproj/dangling1.doc):

1 Entering the stadium, the size of the crowd surprised John.

Here the subject (‘the size of the crowd’) could not be interpreted as the subject in the first clause (‘entering the stadium’), thus leaving it ‘dangling’. One way to overcome this seemingly illogical sentence structure is to put ‘John’ in the subject position (e.g. ‘Entering the stadium, John was surprised by … ’). As Huddleston and Pullum (2005: 207–9) have pointed out, however, such a rule is by no means observed by all users of Standard English; some appear to find nothing wrong in a sentence like (2), which was collected from authentic print media data in an ENL country:

2 Jennifer Lopez stars as Marisa, a maid in a fancy New York City Hotel. While trying on a wealthy woman’s dress, a handsome and rich politician mistakes her for a society woman.

(Huddleston and Pullum 2005: 208)

Other synchronic variations within Standard English are arguably results of more or less recent diachronic changes: witness the neutralization of what used to be a clear functional division of labour between ‘compared with’ and ‘compared to’, which was triggered by a gradual shift of the former’s functional load to the latter (e.g. compared to...
my situation used to be considered substandard, when compared with NP was widely held to be the norm, which was not to be confused with, e.g., Cio-Cio-San was compared to a butterfly. Or, consider the collocation between the amount of and count nouns like books, which used to be seen as substandard about three decades ago when the number of was the norm. These examples, barely the tip of the iceberg, are indicative of perennial language change, including in standard varieties of English (Milroy and Milroy 1985).

In the face of the many learner-unfriendly features exemplified above, coupled with cross-linguistic influence at various linguistic levels in the learning process, it is not surprising that deviations from Standard English norms tend to occur at all stages of the English-L2 learning process.

Non-standard lexico-grammatical features

In general, an error is an error if it deviates from the norm. But given that language change takes place all the time, the question arises as to when a deviation may stop being seen as an error and start being considered as (the onset of) an innovation. Before discussing this issue in detail, let us first look at some salient examples of non-standard features which are commonly found among Cantonese-L1 users of English in Hong Kong. Most of the data cited below were collected from undergraduate students’ written output, including emails, supplemented by some authentic data from English language print media. Being undergraduate students, their English proficiency level may be characterized as either intermediate or upper-intermediate.

Some deviations from Standard English are clearly due to overgeneralization resulting from the principle of analogy. This is arguably the case with, for example, the use of widespread as a noun after the model of the nominal use of spread, as in the widespread of American culture; the widespread of Singlish. Or, consider the use of the to-infinitive as the preferred pattern of complementation after the verbs suggest and recommend (e.g. He suggested me to do it; we recommend you to stop), which deviates from the normative use of a that-clause (i.e., He suggested that I do it; we recommend that you stop). Given the dominant pattern of complementation required for many other verbs (compare: She asked/expected/told me to do it; they order/persuade/want you to stop), it is understandable why the to-infinitive is regarded by so many English-L2 learners/users as the preferred pattern of complementation for suggest and recommend. Indeed, there is some evidence that such a trend has been spread to proficient English-L2 users (3) as well as English-L1 users (4):

3 As a linguist who worked recently on the matter of how spatial notions of uchi (inside) and soto (outside) relate to language and culture, I would like to recommend you very strongly to read Dr James Stanlaw’s [2004] book on loanwords as a fascinating case study of interiorization of exterior things and words from English language and culture.

(Seiichi Makino, Princeton University; promotional flyer for a new book, 2004; emphasis added)

4 [Sir Brian Fender] observed that institutions might not have thought sufficiently about the reasons for carrying out knowledge transfer, and as a result might not have accorded sufficient priority to such ‘third mission’ activities. He recommended institutions to conduct more detailed forward planning, and gather
comparable and comprehensive management data with respect to knowledge transfer so that progress can be better monitored.

(Annex to letter by Mr Michael V. Stone, Secretary-General of the University Grants Committee, to the President of the Hong Kong Institute of Education: ‘Proposed Funding & Reporting Mechanism for Strengthening “Knowledge Transfer” in UGC-funded Institutions’, 6 March 2009, p. 2)

Sometimes variation in Standard English may give rise to disagreement. One such case that happened to me concerns the correct complementation pattern of the verb report (report using vs report to use). In response to my query on the grammaticality of reported to use in a draft paper, the author of that paper did a Google search and obtained some interesting results, which are worth quoting at length:

5 I couldn’t find any hard and fast grammar rules relating to this, but came across two websites:

- www.iei.uiuc.edu/structure/structure1/gerinfvbs.html
- www.tlumaczenia-angielski.info/angielski/gerund-infinite.htm

While the first clearly indicates that ‘report’ can only take a gerund object, the second seems to suggest that it can take both gerund and infinitive complements … I also did a Google search for ‘reported to use’ (where ‘reported’ is in active rather than passive voice) and noted that this usage is found in credible texts, such as published journal articles, although the gerund is more often used. Some of the contexts are as follows:

‘… respondents’ distribution according to how often they reported to use different pain control …’ …

‘… only one in five men and one in ten women reported to use no drugs at all’

Of interest here is the indeterminacy of correctness after several rounds of a Google search: while the gerund appears to be the normative pattern of complementation of report in active voice (reported using), the to-infinitive (reported to use) is also attested in some credible web pages on grammar and correct usage.

In extreme cases, both sides would contest what the other side regards as the correct usage. This is clearly the case of one email request I received in April 2008 from a former student (MD), a novice NET (native English-speaking) teacher of English in a well-known secondary school, who felt there was something wrong in the fill-in-the-blank question ‘How well do you know – this little animal?’ set by the head of English, with about being the intended answer. Below (7) is what MD wrote to me after receiving my affirmative response (6):

6 I did a quick Google search using ‘How well do you know about … ’; guess what: no websites were returned (from 1–10). I see this as confirmation of our shared intuition: ‘about’ collocates best with ‘How much … ’, not ‘How well … ’. I suppose the best way forward is to explain this to your students, and convince them that the so-called ‘model answer’ is inaccurate … You could instruct them to do a similar Google search to bring home this message I think.
The problem isn’t [sic] with my students [sic] the problem is with my panel head [of English]. And she used yahoo … and searched it using inverted commas and came up with a screen full of sites using how well and about. When I explained it to my colleagues they all agreed but my panel head doesn’t. She says that it is a common usage. But I disagree. I am not very sure what to do … I am going to search grammar books over the weekend, and collocation books too. I hope I can get some ‘evidence’ to show her.

Examples (5) to (7) are instructive in that the internet is increasingly resorted to as a means to determine to what extent a particular lexico-grammatical usage is legitimate or acceptable. Given that the ever-expanding internet has emerged as a de facto repository or huge English language database, the popular practice of checking for grammatical correctness on the web is thus gradually altering if not revolutionizing our perceptions of what constitutes correct and normative English usage. One crucial point here is that often it is difficult to tell whether the authors of internet texts are English-L2 or English-L1 users.

In the domain of ‘grammar proper’, one of the most slippery grammatical subsystems in Standard English is the distinction between singular and plural forms of a count noun. It is therefore not surprising that even highly proficient English-L2 users sometimes fail to use the appropriate plural form of a count noun. In the following quarter-page advert placed by a prestigious English-medium secondary school in Hong Kong for ‘the post of English teacher’, three count nouns – application, requirement, and purpose – are in singular form whereas Standard English usage would have them in plural:

XXX College invites application from qualified candidates for the post of English teacher (native speaker) as from September 1, 2008.

**Requirement**
- BA major in English
- Willing to help organizing activities and creating a rich language environment in school
- Salary: negotiable $25,000-$40,000 per month …

[In small print] (All information provided will only be used for recruitment related purpose)

*(The Standard, Careers Page, 13 June 2008: 23)*

Keen readers will have noticed that the verb forms after the verb help – organizing and creating – are also non-standard, since verbs that follow help should normally be in infinitive rather than -ing forms.

At the level of lexis, the correct usage of many verbs and nouns depends on their usual collocational pattern. Owing to a lack of exposure and practice, English-L2 learners tend to have problems acquiring the collocational patterns associated with target verbs and nouns. This is arguably the case with one subset of transitive verbs like discuss, emphasize and blame (9a, 10a, 11a), which do not take a preposition, as opposed to their corresponding nominalization supported by a ‘delexical verb’ (*‘have … discussion about NP’, 9b; ‘place … emphasis on NP’, 10b; and ‘put … blame on NP’, 11b). Non-standard structures as in 9c, 10c and 11c are arguably the result of the English-L2 learner/user confusing the collocational patterns of the (transitive) verb and the associated nominalization (Li in press).
9 (a) They discussed the project for two hours.
(b) They had a long discussion about the project.
(c) ? They discussed about the project for two hours.

10 (a) We should emphasize this more.
(b) We should place more emphasis on this.
(c) ? We should emphasize on this more.

11 (a) Don’t blame her so much!
(b) Don’t put so much blame on her.
(c) ? Don’t blame on her so much.

Plenty of non-standard usage patterns may be accounted for by a similar misanalysis, as shown in the spread of the non-standard complementation pattern of recommend to English-L1 speakers (e.g. ‘He recommended institutions to conduct … ’, see examples 3–4 above). Likewise, in class is such a high-frequency prepositional phrase that English-L2 learners might take a long time to realize that in classroom is inadmissible without the definite article the. Other examples in my data include the use of behind as a post-nominal modifier as in the reason behind (12), the redundant use of about in concerning about X (13), and the plural form of room in the idiom room for improvement (14):

12 After finished my associate degree, I chose English as major in my degree. There were several reasons behind. Firstly …

13 May I refer to the following email to Head and Research Degrees Coordinator dated 22 November 2007 concerning about the Research Students’ Research Output …

14 Despite the fact that there are still rooms for improvement in my English, especially the writing skills, I have never forgotten my own identity as a Chinese even I am able to acquaint myself well with English.

Some of these apparent anomalies are arguably due to idiosyncrasies in Standard English. For example, ‘the reason behind’, in analogy to ‘the day before/after’ or ‘the point above/below’, seems quite reasonable. And it is only relatively recently that concerning and regarding are formally recognized as prepositions in some dictionaries (see, e.g., Collins Cobuild Dictionary), thanks in part to insights obtained in corpus linguistics. This fine detail has yet to trickle down to the English-L2 classroom. There is some evidence that the usage patterns of the verb concern and its derivatives are complex and learner-unfriendly. For instance, many English-L2 learners would say/write father concerns you or father concerns about you (meaning ‘father is concerned about you’), partly because they overlook the syntactic constraint of the verb concern, partly due to incomplete learning of the periphrastic expression be concerned about (e.g. father is concerned about you) and the prepositional use of concerning (e.g. concerning your safety; Li and Chan 2001):

’something concerns someone’
’someone is concerned about someone/something’

Another group of learner-unfriendly words are adjectives with a meaning related to the degree of difficulty and probability, for example, difficult, easy, common, convenient, compulsory, necessary, unnecessary, possible, impossible, etc. One syntactic constraint associated with these adjectives is that in general, the clause should start with the dummy subject it rather than a ‘human’ subject. For example:
UNCONVENTIONAL FORM OR INNOVATION?

15 (a) *I am difficult/not easy to learn English well.
   (b) It is difficult/not easy for me to learn English well.
16 (a) *We are inconvenient to see you now.
   (b) It is inconvenient for us to see you now.

For Chinese EFL learners, however, the normative use of this structure (known as ‘postponed carrier’ in functional grammar, as in 15b and 16b; see Lock 1995) is learner-unfriendly for two main reasons: the non-existence of a functional equivalent of ‘it’ in their native language (unlike many European languages in this regard), and the fact that, in Chinese, sentences with such meanings tend to begin with a human subject. This is probably why even highly educated Chinese bilingual users of English are sometimes prone to produce this non-standard structure known as ‘pseudo-tough movement’ (see Li and Chan 2001). In one seminar given by a Chinese Singaporean lecturer on the impact of the spread of the Chinese language in the world, he said, ‘you are difficult to buy non-Chinese products’. (This syntactic constraint is neutralized when the covert object of the verb in the embedded clause is the same as the subject in the matrix clause. Compare: John is easy to please; Liu Xiang is difficult to beat.)

Learner-unfriendliness is also attested in another salient Standard English structure which is known as ‘reduced relative clause’ (RRC). When a post-nominal modifier consists of a relative clause in the passive voice (e.g. I bought that book which was published yesterday), Standard English allows for a stylistic variant whereby the relative pronoun and the finite auxiliary may be ellipted (e.g. I bought that book published yesterday). The RRC structure, however, is blocked if the verb is intransitive (e.g. I saw the accident which happened yesterday, but not *I saw the accident happened yesterday). Such a lexico-syntactic constraint is often overlooked by even advanced English-L2 users. In one quarter-page public notice in a leading English daily in Hong Kong, for example, the verb appeared was used in the same RRC structure as in published:

17 We note from the reports/articles appeared at the front page and page 3 of the South China Sunday Morning Post published on 27th August 2000 … that a toy company called ‘City Toys Ltd.’ … has employed underage workers.

(South China Morning Post, 1 September 2000: 3)

Where the verb in a post-nominal modifier is intransitive (e.g. appear), it should either be ‘introduced’ by a relative pronoun (i.e. … which appeared …) or in -ing form (i.e. … appearing …).

Previous accounts of learner errors in second-language acquisition (SLA) tended to focus on the source of errors, with the primary factor being either cross-linguistic influence from the learners’ L1, or incomplete learning of L2, or some combination of these (for a critique of this analytical stance, see Jenkins 2006). While there is some truth in such explanatory accounts, they are incomplete without appreciating the fact that the target language, Standard English, is a system that leaks and, as we have seen, in extreme cases to the extent of logical inconsistency. Another source of difficulty is instability, as shown in various stylistic variants at practically all linguistic levels. Following the emergence of English as a global language, with the result that learners from different L1 backgrounds often have to learn one or more standard varieties of English, a troubling question arises: should English-L2 users’ non-standard performance and usage patterns be necessarily dismissed as ‘errors’? After more than two decades of
research in World Englishes and other related paradigms, few would dispute that at least some of the non-standard features produced by English-L2 users should be regarded as legitimate and recognized as innovations rather than errors. The question is where to draw the line.

Deviations from Standard English: errors or innovations?

Standard varieties of English are products of successive stages of standardization as a direct result of decades (e.g. AusE) or even centuries (e.g. BE and AmE) of codification and/or language planning (Kirkpatrick 2007). To some extent, what standards do is to impose some order on a state of unsystematic variation. For a long time, standards of English were modelled prescriptively on the lexicogrammar in Latin, regardless of how English was actually used by its speakers (Milroy and Milroy 1985). Over time, the prescriptive approach gave way to a descriptive stance among contemporary linguists and grammarians; in the process dogmatic usage patterns (e.g. it’s I) modelled on Latin gradually succumbed to the collective forces of popular usage and choice (e.g. it’s me). Before English emerged as the world’s de facto global language, such collective forces naturally referred to those exerted by the everyday language use patterns of its English-L1 users. Now that English is a required additional language in most non-English-L1 countries in the world, especially in view of the fact that English-L1 users are increasingly outnumbered by English-L2 users, the question arises whether such forces of language change should be attributed to English-L1 users alone. To cite one classic example: why should prepone, a well-motivated verb – an antonym for that matter – coined in analogy to postpone, be dismissed as a non-English word, even though it has been widely attested among speakers of English on the Indian subcontinent (Widdowson 1994; cf. discuss about NP, emphasize on NP and blame on NP, see examples 9–11)? A Wikipedia entry reads: “Prepone” is not an English word. It’s commonly used in Indian subcontinent to mean the opposite of “post-pone”, but the rest of the world is largely unaware of it (http://wiki.answers.com/Q/Why_the_word_'prepone'_is_not_in_any_dictionary).

Public awareness of a new coinage, however, is far from being the reason why that coinage is not accepted as an innovation in World Englishes (151,000 hits were returned in a Google search in mid-April 2009). Clearly other more potent factors are at stake here. First and foremost, the status of prepone is low because its active use to date tends to be limited to the popular parlance of users who are labelled as non-native speakers. Second, more importantly, innovation – including the power to label new coinage as such – was traditionally thought to be the exclusive right of native speakers, notably those residing in UK, USA, Canada, Australia and New Zealand. So what needs to be done before such an ingenious coinage as prepone is accepted as part of the lexicon in Standard English?

To my knowledge, Bamgbose (1998) is the most elaborate treatise on the theoretical distinction between English-L2 errors and innovations. Coming from a World Englishes perspective, he asks, ‘why should a native-variety-based standard continue to license the norms of non-native Englishes?’ (p. 3). As he explains, the current state of affairs favours standard varieties of English, partly because all existing standards are upheld to be correct until otherwise replaced with alternative standards or complemented by stylistic variants, but also because they are the most elaborately codified to date: ‘By
default, the only codified norms available (which are based on native varieties) will continue to license what is acceptable and what is not, even when there is a desire to encourage and institutionalize non-native English norms’ (Bamgbose 1998: 5).

Owing to this prestige factor, English-L2 speakers tend to admire native accents, even though their own pronunciation does not sound native-like, reflecting thereby a kind of ‘love–hate relationship’ (p. 7). This point has received empirical support in a recent study of Chinese speakers’ perceptions of English accents (Li 2009, cf. Jenkins 2007).

To calibrate the status of a local usage as either an error or innovation, Bamgbose (1998) indicates that there are five interrelated internal factors or measures:

- **Demographic:** how many acrolectal speakers use it? Since the language use patterns of basilectal and mesolectal speakers tend to be socially stigmatized, the prospect of the usage being favourably received in the local community is dim if it is not used by acrolectal speakers.
- **Geographical:** how widely has it spread? In principle, the farther it spreads, the higher its acceptance rate.
- **Authoritative:** what is the social status of those who use it? In general, people who are knowledgeable are vested with authority. Thus ‘writers, teachers, media practitioners, examination bodies, publishing houses, and influential opinion leaders’ (p. 4) tend to be viewed favourably as credible sources of linguistic innovations, for ‘the use of unconventional forms may become hallowed, simply because such use has become associated with respected authorities or writers’ (p. 4).
- **Codification:** where is the usage sanctioned? One sure way to legitimate a local usage is to have it included in all kinds of written ‘authorities’, such as dictionaries, course books and reference manuals for teachers.
- **Acceptability:** what are the attitudes of users and non-users towards this usage? In general, compared with linguistic innovations, cultural and pragmatic innovations tend to get accepted more easily and are more likely to be tolerated and nativized.

Of these five internal measures, Bamgbose points out rightly that codification and acceptability are the most important. Beyond any doubt, the key to language change is codification, a point which ‘is too important to be belabored’ (p. 4). Once a local usage is enshrined in the dictionary or even in a course book, the legitimation process is complete (Dolezal 2006; Butler 2007). This in turn will help tilt the balance, if gradually, in favour of accepting that local usage although, as Bamgbose has observed, English-L2 users, including decision-makers in the education domain, tend to resist making this move (1998).

**Internet as catalyst of acceptance: web-enabled innovations in cyberspace**

In the decade since Bamgbose’s article, the question of grammaticality and acceptability has become considerably more complex following significant breakthroughs in ICT and global advances in bi- or multilingual e-literacy, which invariably includes some English. In the first decade of the new millennium, in some real sense the ‘global village’ has rendered the world smaller following dramatic improvements in telecommunication mediated by the internet. Physical barriers marked by political and geographical boundaries, real or imagined, are increasingly rendered obsolete relative to
people’s desire to access information or communicate with others in cyberspace, wherever their internet workstation is located. For about two decades, information on the internet has been and continues to expand at an exponential rate, in more languages than ever, but search engines like Google, Yahoo and Baidu have made this task increasingly manageable for web-surfers (cf. Graddol 2006). Today, whatever the information in the public domain, be it language- or image-dominant, it is rarely more than just a few clicks away. As a result of this development, ‘geography’ and ‘demography’ as measures of English users’ perception of the correctness of a local usage have become comparatively less significant. Much more pervasive today is what may be termed ‘virtual vitality’: whatever query about normative English usage one has, a quick check through Google or Yahoo (or any other search engine) can instantiate as many glocal examples as there are in various ‘cyber communities’, be they English-L2 or English-L1 users (Gupta 2005, 2006, 2007; cf. Pakir 1999).

Gupta (2007), for example, examines the extent of anglophony in official websites of the ten ASEAN nations and found that with few exceptions (notably Myanmar), English is widely used in the key domains of government and education. She also found a ‘hierarchy of Anglophony’ (p. 366), with English being more commonly used for internal purposes in some ASEAN nations (notably former British colonies) than in others. In terms of the extent of variation, despite minor divergence in spelling and usage patterns, which Gupta regards as ‘differences of preference rather than categorical’ (p. 357), the formal features of English across ASEAN websites are remarkably similar. This high degree of unity of Standard English is attributed to a loose consensus of elite users, suggesting that ‘codification of English follows practice, rather than determining it’ (p. 357).

Recent developments on the internet are thus exerting considerable impact on our perceptions of what counts as an error (i.e. the form is an unintended violation of some Standard English norm), as opposed to a linguistic innovation (i.e. the form is intended as a carrier of a new, probably culture-specific meaning with a local or glocal character). We have seen that more and more users of English turn to the internet as an act of licensing or means of legitimation (see examples 5–7): if an English usage is attested by a large enough number of users on the internet, especially if glocal and English-L1 users are included, it is difficult to insist that it is an error. One instructive example is the status of the collocation advanced booking, which appears to be non-standard but which is found on a large number of websites, including those of international hotel-booking agencies (see, e.g., www.epoquehotels.com/specials-promotions/promo-info.php/hotel/unahotelcatania/promo/1977) and a journal article on travel research (see Chen and Schwartz 2008). Or consider the spelling of irresistible which, while non-standard according to dictionaries in standard varieties of English and Microsoft Word, is no less popular than the normative counterpart irresistible, probably because the suffix -able is semantically and orthographically more transparent (compare the increasingly popular trend of writing everyday to mean ‘every day’, can not [VERB] instead of cannot [VERB]. These examples show that the spread of a new usage has the potential to catch on and command a mass following, especially if it is well motivated. When it later spreads to formal communication among educated English users on the internet, the legitimation process is half complete. When that happens, it is the duty of the lexicographer and/or grammarian to have its legitimate status – as an acceptable variant – formally recognized. In short, advances in ICT help explain why our attitudes towards the perceived legitimacy of a new English usage are less bound today by
Why acrolectal, educated English users? This is related to Bamgbose’s third factor or measure: authority. Just as renowned literary figures, writing in any language, enjoy the unquestioned prerogative or poetic licence to deviate from existing lexico-grammatical norms of the language, so educated speakers and writers have the unparalleled privilege to ‘bend’ the language at times to suit their context-specific needs. Such a move from an ‘authority’ would rarely raise any eyebrows, for it is generally perceived as a novel way of meaning-making, whatever the communicative purpose (e.g. new concept, imagery or metaphor). The same expression, produced by learners in the classroom or in some language-learning context (e.g. students’ assessed class- or homework), would tend to be dismissed as ‘interlanguage’ in need of correction. For instance, a student of English who feels inspired by the Chinese Premier Wen Jiabao’s rendition of weiji (危機, ‘crisis’) in Mandarin as a disyllabic word composed of the morphemes ‘crisis’ and ‘opportunity’ (Wen’s official visit to London, February 2009), and who is tempted to capture both morphemes by the coinage crisistunity, may be praised by the teacher as ‘a good attempt’, but it would nonetheless be dismissed as ‘non-standard’ – along with other ‘interlanguage’ errors. Yet when this coinage appears in a feature article of an English daily, as it does (Gao 2009), complete with sound justification and supportive illustrations, no reader will question its status as a well-conceived innovation. A Google search of crisistunity in mid-February 2009 failed to yield any hits. Another Google search two months later (13 April 2009), however, returned over 330 hits, including translations of the original English article into foreign languages such as Italian and Russian, but also in an e-newsletter update of Broome County Peace Action, New York (March 2009, http://bcpeaceaction.org/update.pdf), and a website entitled ‘Jump Ultimate Star’ featuring air travel, leisure activities and other links (http://crisistunity.com/justp/). Interestingly, the 330 plus hits also include a few other websites containing a similar word ‘crisitunity’ (with only one ‘s’), which was apparently coined by Homer Simpson:

Crisitunity: A Chinese word referred [sic] to by Homer Simpson that means both crisis and opportunity [sic], just like Ercle.


Upon being told that the Chinese word for ‘crisis’ is the same as their word for ‘opportunity,’ Homer Simpson gave the word ‘crisitude’ to the English-speaking world.


Crisistunity (coined by a Chinese-L1, English-L2 speaker) and crisitude (coined or adapted by an English-L1 speaker) may sound clumsy to the ear phonologically, but they appear to be catching on. This is a clear example of lexical innovations inspired by Chinese ‘equivalents’ which are similar in meaning, albeit with subtle semantic nuances. Of further interest is that, after lexical innovations were coined (apparently) independently by an English-L2 and an English-L1 user, the English-L2 coinage (crisistunity) appears to be ‘crossing’ into English-L1 territories, as shown in the above-mentioned New York Broome County Peace Action e-update:
President Obama is a centrist. We don’t need a Goldilocks President (not too hot, not ‘crisistunity’ say [sic] Kevin.

(http://bcpeaceaction.org/update.pdf, p. 2)

In terms of process, the spread of crisistunity seems not so different in kind from the popularization of an English-L1 coinage like nonebrity, denoting a celebrity who is famous for nothing in particular. There is thus some indication that hybrids and bilingual creativity (Kachru 1995) by English-L2 users have good potential to be appropriated by English-L2 and English-L1 users alike – thanks to forces of globalization mediated and facilitated by the internet.

A second example comes from Phan Le Ha’s (2008) book where, in the section ‘Ha and English’, she writes:

[My parents] did not have the right to choose the language they liked [to study] at that time, Russian or Chinese or French. For historical and political reasons, these languages had high status in Vietnam in those days. It also meant that learning and teaching English then would lead people to an ‘insecure’ future with almost no chance for further study overseas. And going overseas in the 1970s, 1980s and early 1990s did not just bring about new knowledge but also meant ‘changing one’s material life’ to ‘wealthi-ness’ or at least ‘well-furnituredness’.

(Phan 2008: 15)

The author is unmistakable about her Vietnamese-L1 and English-L2 background. The use of scare quotes in ‘well-furnituredness’ (and ‘wealthi-ness’) is a sign of its potentially objectionable status. This is partly confirmed by the result of a Google search (mid-February 2009), which returns no other entry than Phan’s (2008) book page itself, suggesting that this usage is idiosyncratic. Be that as it may, the fact that it has survived the copy-editing stage of the book-production process is suggestive of a high level of tolerance of non-standard English usage in works written by acrolectal, educated English-L2 writers.

To sum up, Bamgbose’s five internal factors or measures of innovation discussed above should be complemented by a sixth, namely, the popular choice of acrolectal English-L2 users in cyberspace.

**Conclusion**

One consequence of the emergence of English as the world’s de facto global language is that, whatever a person’s first-language background, he or she will be disadvantaged without learning at least some English. The variety of English which has the greatest currency is Standard English (Li 2007). Despite being standardized and codified for decades (e.g. AusE) or centuries (e.g. BrE and AmE), a standard variety of English is a system that leaks. For millions of English-L2 users, this is one source of learner-unfriendliness. Another source is considerable variation within a standard variety of English. These two sources of learner-unfriendliness, coupled with cross-linguistic influence from the previously acquired language(s), help account for English-L2 learners’ propensity to produce all kinds of non-standard features at all stages of the learning process.
For a long time, deviations from Standard English norms were characterized as unsuccessful attempts at imitating the ways native speakers use English, or ‘errors’ in short. Research in World Englishes and other related paradigms for over two decades, however, has made a very strong case for the legitimacy of non-standard features found in the Englishes of ESL users who use English for intra-ethnic communication. The fine line between errors and innovations has been challenged. It has been shown that many of the seemingly non-standard ESL usage patterns are in fact well-motivated innovations, subject to five factors or measures (Bamgbose 1998): ‘demographic’ (i.e. percentage of acrolectal users vis-à-vis mesolectal and basilectal users), ‘geographical’, ‘authoritative’, ‘codification’ and ‘acceptability’ (i.e. attitudes).

A decade after Bamgbose’s (1998) landmark article, ‘authoritative’, ‘codification’ and ‘acceptability’ remain important measures of innovation, but ‘demographic’ and ‘geographical’ are arguably declining in relative significance following dramatic advances world-wide in ICT (information and communication technologies) – internet communication in particular. Increasingly, English-L2 and English-L1 users alike may turn to the internet to ascertain the ‘virtual vitality’ of a given coinage or usage pattern with the help of a search engine like Google, Yahoo or Baidu. This practice has significant impact on the degree of its perceived legitimacy and acceptability. Therefore, Bamgbose’s five internal factors or measures need to be complemented with a sixth, namely the popular choice of acrolectal, educated users of English on the internet, whatever their first language may be (cf. Gupta 2005, 2006, 2007).

Suggestions for further reading

Bamgbose (1998) is a seminal article covering the key issues in the debate concerning the slippery distinction between errors and innovations.

Breiteneder’s (2005) paper provides empirical evidence how the ‘third person -s’ is systematically flouted by speakers of English as a European lingua franca (cf. Breiteneder 2009).

For a theoretically informed discussion of identity-driven ‘user English’ as opposed to acquisition-based ‘learner English’, see Kirkpatrick (2007).

References


DAVID C.S. LI


LIDIAN DAVID CHENG-SHIU LI


UNCONVENTIONAL FORM OR INNOVATION?


Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism

Qin Xie*
Hong Kong Institute of Education, Hong Kong

A R T I C L E   I N F O
Article history:
Received 18 January 2014
Received in revised form 7 March 2015
Accepted 10 March 2015
Available online

Keywords:
Washback
Test design
Test preparation
Test-taker perceptions
Structural equation modeling

A B S T R A C T
This study utilized Structural Equation Modeling to investigate the washback mechanism, focusing on two design aspects of an English language proficiency test: component weighting (weight assigned to different test papers) and testing methods (item format), and their washback on test preparation. Two months before taking the test, a large sample of test-takers (N = 1000) were surveyed regarding their perceptions of the two design aspects and their test preparation activities. Their official test scores were collected when they were available. Data was analyzed to estimate the washback effects of perceived changes on test-taker time management and approaches to test preparation, and their test performance. The study found that test-takers spent more time on the papers with higher weight and less on those with lower weight. Reporting component scores seemed unable to adjust this tendency. Meanwhile, favorable perception of test validity was associated with a higher level of engagement in both desirable language learning activities and focused test preparation (drilling and cramming). This suggests that favorable perceptions may not be able to reduce negative washback, but may be able to promote positive ones.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Washback, or backwash, refers to the influence of high-stakes external testing on teaching and learning within the classroom (Alderson & Wall, 1993; Cheng & Curtis, 2012; Green, 2013; Qi, 2011). Existing studies on high-stakes English language proficiency tests (e.g., Alderson & Wall, 1993; Cheng, 2005) provide insights into the complex nature of washback. However, the exact mechanism of washback remains unclear (Cheng, Watanabe, & Curtis, 2004). This is certainly because of the complex nature of washback, where multiple factors interact and shape the extent and nature of the washback getting into the classroom. It may also be partly due to the methodological limitations of existing washback studies. Most existing studies adopt qualitative methods (Cheng et al., 2004; Green, 2007; Matoush & Fu, 2012; Qi, 2005; Zhan & Andrews, 2013); very few studies employ quantitative methods. While qualitative methods can effectively identify various factors affecting the washback getting into classrooms, they are not as effective for unraveling the network of relationships among these factors. Quantitative methods, on the other hand, can answer questions such as in what ways and to what extent do different factors associated with testing affect teaching and learning in classrooms? Such questions are essential to understand the

* B4-2/F-12, Hong Kong Institute of Education, NO. 10 Lo Ping Road, Tai Po Market, New Territories, Hong Kong. Tel.: +852 2948 8368.
E-mail address: qxie@ied.edu.hk.

http://dx.doi.org/10.1016/j.system.2015.03.002
0346-251X © 2015 Elsevier Ltd. All rights reserved.
mechanism of washback. This study surveyed a large sample of student test-takers and utilized Structural Equation Modeling to explore the mechanism underlying the washback of high-stakes testing. The study was situated within the context of China’s reform on the College English Test, focusing on two aspects of test design and their washback effects on test-taker time management, approaches to preparation, and test performance. The following section reviews relevant literature on washback, component weighting, and Structural Equation Modeling.

2. Literature review

Many studies on test washback were conducted following the introduction of new language tests (or new test features) to investigate whether changes made to the tests were followed by desirable changes in teaching and learning (e.g., Cheng, 2005; Hung, 2012; Wall, 2000; Wall & Alderson, 1993). These studies found washback to be highly complex, because multiple factors and multiple stakeholders coexist, and their complex interactions largely determine the extent and nature of washback getting into the classroom (Matoush & Fu, 2012; Shih, 2010; Spratt, 2005; Watanabe, 2004b; Zhan & Andrews, 2013).

For instance, the multiple-choice (MC) test format associated with modern psychometric testing is believed by many to be one source of detrimental washback (Messick, 1996), whereas open-response testing methods, which put less restriction on test-takers, are considered capable of promoting beneficial washback and enhancing test validity. The examples of open-response testing are fill-in the blanks, short-answer questions, and essay writing. Performance testing, which adopts more open-ended testing methods, is generally perceived to have higher potential for generating beneficial washback. Because tasks in performance tests are closer to real-life tasks (more authentic), they tend to receive positive responses from test users, including teachers and students, in regards to their validity in assessing language proficiency (Struyven, Dochy, & Janssens, 2005). However, studies on performance testing found its washback to be merely superficial (Andrews, Fullilove, & Wong, 2002; Cheng et al., 2004; Wall, 2000). Changes made to these tests will in turn produce changes in the content of teaching and learning, but not in the methods. They may change the manner of test preparation but not its nature. Test washback is described as a “blunt instrument” (Andrews, 1995), which often falls short of the test designers’ intentions; that is, good intentions often do not materialize in classrooms.

Compared with the number of washback studies on testing methods associated with performance testing, there are considerably fewer studies on component weighting and its effects on teaching and learning. It is a shared view that if one component is not assessed or is taken away from a test set, this component is likely to be ignored by teachers and test-takers (Kane & Case, 2004; Watanabe, 2004a). However, little has been done regarding the exact weighting of test components and their differential impact on teaching and learning. If test-takers are driven by what is viewed as valuable (Spolsky, 1995; Xie, 2013), it is reasonable to infer that they will spend more time and resources on the components with higher weighting and vice versa. However, such reasoning remains intuitive and may be simplistic. The effects of differential component weighting may be more complex than that. Test preparation is likely to be affected by forces other than rational considerations. For instance, influences from previous test experience and consideration of the cost-effectiveness of resources may affect test preparation. It has been observed elsewhere (e.g., Powers, 1987) that test-takers invest more effort in the sections considered to be “coachable” (i.e., where performance can be improved within a relatively short timeframe) in order to achieve a better return from their investment.

To discourage test-takers’ tendency to ignore less weighted components, many test developers chose to report component scores along with a global score. However, this measure alone may not be sufficient for the purpose. The test also has to adopt a non-complementary or conjunctive scoring system. In a non-complementary scoring system, a minimal requirement or a threshold for each component is set so that it is necessary to pass each component in order to pass the whole test. The driving test is an example of the non-complementary scoring system (Kane & Case, 2004). To get a driver’s license, test-takers must pass both a written test of traffic rules and a road test of driving skills. A high score on the written test cannot compensate for a low score on the road test, and vice versa. By contrast, in a complementary scoring system, decisions are made based on composite total scores, which are computed as a weighted sum or an average of component scores. Thus, high scores on one component can compensate for low scores on the other components (Hambleton & Slater, 1997; Kane & Case, 2004).

Most literature on component weighting is within the area of educational measurement, where the primary concern is the impact of different weighting schemes on the reliability and validity of composite scores (Rudner, 2001; Sawaki, 2007; Wainer & Thissen, 1993). In this area of studies, consequences of component weighting are mentioned in association with the concept of nominal weighting, which refers to the stated weighting-scheme of a testing program. Nominal weighting reflects policy intention and value judgment regarding the relative importance of different components within a test set, which contrasts with effective weighting, the actual weighting in the composite scores owing to differences in component-score reliability and variance. Effective weighting is a psychometric concern, whereas nominal weighting is pertinent to educational policy and curriculum innovation. Although there are many psychometric studies on effective weighting, little research has been conducted on nominal weighting and its impact on teaching and learning. Since nominal weighting of high-stakes testing conveys a clear, explicit value-message to educational practitioners and test-takers, investigations of nominal weighting schemes, especially in terms of their potential and effectiveness in directing teaching and learning, are clearly warranted. Such studies can provide useful information to facilitate curriculum innovation.

Another gap identified in the washback literature concerns the research method adopted. Most existing washback studies are qualitative; there is a general lack of quantitative studies. Although qualitative studies can identify salient factors affecting
the nature and extent of washback (Watanabe, 2004a), they are not capable of evaluating the relative importance of these factors or estimating the precise relationships among these factors (Xie & Andrews, 2013). To date, the relationships between aspects of test design and types of washback remain under-investigated, and the mechanism of washback remains unclear (Cheng & Curtis, 2012). While this is certainly because of the complex nature of washback, it may also be partly due to the limitation of qualitative research methods. This study adopts Structural Equation Modeling (SEM), a sophisticated quantitative methodology, to study the washback mechanism.

SEM is widely applied in social science and psychology to investigate networks of complex relationships among a range of variables. It is particularly suitable for investigating complex social—educational phenomena. In a SEM study, a conceptual model will be specified in advance based on a theory, previous research, or the researchers’ hypotheses. This conceptual model stipulates a range of complex relationships among the variables; these relationships will be tested by fitting the model with empirical data. The closer a model fits the data, the better it explains the underlying relationships among the variables, and the more plausible the model is (see Kunnan, 1998 for a very accessible introduction of SEM). SEM can answer research questions such as the following which are investigated in the present study:

1. To what extent do test-taker perceptions of component weighting and score profiling affect their time management for test preparation and their test performance?
2. To what extent do test-taker perceptions of test design affect their approaches to test preparation and their test performance?

3. Context of this study

This study was conducted within the context of China’s reform on a large scale English language proficiency test, i.e., the College English Test (CET). This section provides international readers some background information regarding this educational context, the test reform, and the two aspects of test design under study.

In China, debates and controversies associated with high-stakes examinations and their detrimental effects on education have a long history. For more than one thousand years, the Imperial Examination for civil servants had been used as the most important mechanism to select elite personnel for important government offices (Cheng & Curtis, 2010). Because of its detrimental effects on education, the Imperial Examination was considered to be the primary obstacle that must be removed in order for the country to move toward modernization. In 1905, the Imperial Examination was formally abolished by the last emperor. One hundred years later, however, the education system of 21st-century China is still considered examination oriented (Cheng & Curtis, 2010; Qi, 2007). The first decade of the new millennium witnessed several large-scale breaches of test security in high-stakes examinations, which triggered massive outburst of public discontent toward the examination-oriented education system. Subsequently, large-scale educational reforms were launched. High-stakes examinations were reformed in order to drive educational innovations for classroom teaching and learning. In the area of English language education, the College English Test series (CET) is one of those under innovation (Jin, 2005a).

CET is a paper-based, standardized test of Academic English proficiency for tertiary students in China. It has two test batteries targeting test-takers at two levels, CET4 and 6. CET4 is the more influential of the two as it is mandatory for all non-English major undergraduates across most universities in China. CET4 is administered twice a year, each time involving over two million test-takers (Jin, 2005b; Yang & Weir, 1998). The test was originally designed to assess students’ English proficiency at the end of their two-year College English program. However, test scores of CET4 have been widely used as a university graduation requirement and for job applications, making it high-stakes.

In recent decades, CET4 has received much criticism of its test design and uses of test results. Amidst other criticisms, the test is alleged to have a negative impact on teaching and learning (Yang, 2003; Zhang, 2003). As a partial response to such criticism, a large-scale testing reform was launched in 2004. The test reform aimed at shifting the pedagogic focus from teaching knowledge of language to developing the ability to use the language, particularly that of listening and speaking, and to promote positive washback on teaching and learning (Jin, 2005a, 2005b). Two major sets of changes have been made to improve the design of CET4.

The first set of changes was made in the way that test scores are computed and reported in the revised CET4 (R-CET4) (see Table 1). The second set of changes introduced more open-response test items to enhance test validity (see Table 2). The scoring of test components was adjusted in terms of how their weighting contributes to the overall composite score. More weight was assigned to the listening (from 20% to 35% of total score) and the writing component (from 15% to 20%); less weight was given to the reading component (from 40% to 35%). Meanwhile, a more detailed score profile replaced a pass

<table>
<thead>
<tr>
<th>CET4 components</th>
<th>Weighting</th>
<th>R-CET4 components</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Listening</td>
<td>20%</td>
<td>1 Listening</td>
<td>35%</td>
</tr>
<tr>
<td>II Reading</td>
<td>40%</td>
<td>2 Reading</td>
<td>35%</td>
</tr>
<tr>
<td>III Vocabulary</td>
<td>15%</td>
<td>3 Integrative skills (Cloze)</td>
<td>10%</td>
</tr>
<tr>
<td>IV Cloze</td>
<td>10%</td>
<td>4 Writing</td>
<td>20%</td>
</tr>
<tr>
<td>V Writing</td>
<td>15%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Changes in component weighting.
4. Method

4.1. Participants and data collection procedure

Participants initially comprised 1000 test-takers from one university in South China. This university is ranked in the lower tier among all universities in the national matriculation system. It recruits students nationwide, but the majority of its certificate. Besides a total composite score, the new score profile also provided separate scores for the four components, i.e., the listening, the reading, the integrative skills, and the writing components.

The two sets of changes were introduced to serve two purposes, i.e., to direct pedagogic attention to the components with more weighting (e.g., the listening component), and to ensure that the components with less weighting (the writing and the integrative skills components) are not neglected. For the two changes to achieve their purposes, however, they must satisfy two assumptions. The first assumption is that differential component weighting indeed affects test-takers’ evaluation of the test component concerned so that they actually pay more attention to the weighted component. The second assumption is that a more detailed score profile will force test-takers to strive for a balanced profile and not neglect the less weighted components.

However, these two assumptions may not be warranted. Although R-CET4 reports component scores alongside a global composite score, users of CET4 scores tend to make decisions based on the composite scores. Specifically, an overall composite score of 425 is commonly used as the pass-and-fail cut-off point for R-CET4. Few test users set thresholds for individual test components. It is, therefore, possible for test-takers to “pass” the test, if they achieve high scores on the reading and the listening components but low scores on the writing and the integrative skills components. That said, CET4 is not the only test encountering such a problem. Many high-stakes testing programs rely on multiple-component tests (e.g., Scholastic Aptitude Test, Graduate Record Exam, and TOEFL), but their users tend to use composite scores in decision-making. Hitherto, scarcely any studies have investigated the impact of component weighting on test-takers and their teachers.

The second set of major changes is introducing less-closed testing methods into the R-CET4. Altogether five new test methods are incorporated into the R-CET4 (Table 2), including long conversations and compound dictation in the listening component, speed reading and a banked cloze test in the reading component and a translation task in the writing component. The response format for the five new test methods is fill-in-the-blanks. According to the test developer (Jin, 2005b), the purposes of introducing less-closed test methods into CET4 are to enhance test validity and to induce positive washback. Critics of CET (Cai, 2006; Han, Dai, & Yang, 2004), however, consider the changes made to CET test design inadequate, conservative, and limited, and hence lacking the power to redirect teaching and learning of English within the classroom to focus on speaking and listening. Both the critics and the test developers, however, share one common assumption; that is, if a test is more valid, it is more likely to induce beneficial washback and to discourage negative washback. Such an assumption may not be warranted either.

Traditionally, test validity refers to the extent to which the test has measured what it intends to measure. White (1995) describes test validity as the honesty of a test in that it does what it claims. Contemporary test validity theory (Bachman & Palmer, 2010; Messick, 1988) considers validity to be contingent on the purpose(s) that a test serves, and the interpretation and use of the test results. A test may or may not be valid for this or that purpose, or this or that decision. Moreover, existing studies on washback have found that the mechanism of washback is complex, involving socio-cultural and educational factors beyond the technical considerations of test validity and reliability. There is no deterministic relationship between test validity and washback (Alderson & Wall, 1993; Messick, 1996).

To sum up, although both sets of changes have the potential to induce positive washback, it remains unclear whether they actually do, and, if they do, to what extent and in what ways. The present study collected data from a large sample of test-takers to investigate the washback of these two features of test design, and to verify whether the previously mentioned assumptions are warranted.

4. Method

4.1. Participants and data collection procedure

Table 2
Introduction of new testing methods.

<table>
<thead>
<tr>
<th>Components</th>
<th>Sub-tests</th>
<th>Testing methods</th>
<th>Input</th>
<th>Response</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>Short conversations</td>
<td>audio conversations</td>
<td>MC</td>
<td>Fill in blanks</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>Long conversations*</td>
<td>audio speech</td>
<td>MC</td>
<td>Fill in blanks</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Compound dictation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>Careful reading</td>
<td>passages</td>
<td>MC</td>
<td>Fill in blanks</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Ranked cloze*</td>
<td>a passage with blanks &amp; a word list</td>
<td>Fill in blanks</td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Speed reading*</td>
<td>a long passage &amp; uncompleted sentences</td>
<td>Yes/No/NG</td>
<td>Fill in blanks</td>
<td>7%</td>
</tr>
<tr>
<td>Integrative skills</td>
<td>Cloze</td>
<td>a passage with blanks</td>
<td>MC</td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>Writing</td>
<td>Translation from Chinese to English*</td>
<td>uncompleted sentences with Chinese prompts</td>
<td>Fill in blanks</td>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>Essay writing</td>
<td>writing prompt</td>
<td>free</td>
<td></td>
<td>15%</td>
</tr>
</tbody>
</table>

Note: the asterisk indicates the new testing methods added in the CET4.
students come from the Pearl River Delta. All participants were Year 2 students from this university, who had registered to take the R-CET4 in about two months’ time. Ten weeks before taking the R-CET4, they answered a perception questionnaire. Two weeks before taking the R-CET4, they answered a second questionnaire regarding the activities they had engaged in to prepare for the test in the previous eight weeks. Their official R-CET4 scores were collected from the university registrar’s office when they became available. Test-takers’ student ID was used to link individual test scores with the questionnaire data. After the invalid cases were deleted, 887 cases were kept for the perception questionnaire; 872 cases for the test-preparation questionnaire and 933 cases for test scores.

4.2. Instrument

Two questionnaires were used in this study: the Test-taker Perception of Assessment Demand Questionnaire and the Test Preparation Practice Questionnaire. Reports on the development and validation of the two questionnaires are published elsewhere (Xie, 2011, 2013). This study used 24 items from the perception questionnaire, 22 of which focus on perceived target language skills, and two on perception of component weighting and score profile. It used 41 items from the Test Preparation Practice Questionnaire, including 38 items focusing on language learning activities, and three items on their time-management during the test-preparation period.

The two questionnaires were developed in a qualitative study, in which an open-ended questionnaire was given to 30 test-takers, and six focus-group interviews (N = 5 in each group) were conducted subsequently. Data from the interviews and the open-ended questionnaires were transcribed, coded, categorized, and quantified. The most frequently expressed views were identified and rephrased into statements, which were subsequently used as questionnaire items. For perception questionnaire, the list of statements about skills perceived to be necessary for the CET4 were compared with the target skills listed on the test syllabus. If a target skill on the test syllabus did not have a corresponding statement on the list, a statement describing that skill was added to the list. On the other hand, if a statement describes a skill perceived by test-takers as necessary, but the skill was not listed on the syllabus, the statement was still kept in the questionnaire. For the test preparation questionnaire, the list of statements identified from the qualitative data was compared with established questionnaires on learning strategies, such as the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, Garcia, & McKeachie, 1991), the Strategy Inventory for Language Learning (Oxford, 1990), and the Cognitive and Metacognitive Strategy Questionnaire (Purpura, 1999). As with the development of the perception questionnaire, the statements generated from interviews were used as the primary list to feed into the questionnaire.

Afterward, the questionnaires went through a rigorous validation process. Firstly, each questionnaire was pretested with six target test-takers in six sessions. Each session lasted one hour, involving one researcher and one test-taker. Test-takers were asked to think aloud while answering each item. The purpose of pretesting was to identify any ambiguity or confusion in the items to improve the clarity, accuracy, and appropriateness of the wording for each item. Items were revised after each session, and the revised version was used for the next pretesting session. Six sessions of pretesting were conducted; participants considered the items sufficiently clear in the last two sessions. The questionnaires were then piloted with 157 students, and the pilot data were analyzed via exploratory factor analysis. Items were examined one after another to check their psychometric qualities. Only the items with high loadings on the intended factors and low loadings on the unintended factors were selected.

The perception questionnaire asks test-takers the extent to which they consider each English language skill as necessary for answering questions on the test correctly. Twenty-two items in four subscales are used to assess test-takers’ perceived needs of target listening skills (Plist), reading skills (Pread), writing skills (Pwrit), and linguistic knowledge (Pling). Each item represents one aspect of the target language skills or linguistic knowledge, and together, the 22 items cover all subskills prescribed in the test syllabus. For instance, Plist consists of seven items (α = .808), such as In order to answer questions in the listening test correctly, I must grasp the gist of what is being said. Pread has four items (α = .683), such as In order to answer questions correctly in the reading test, I must understand the key points of the passages. Similarly, Pling has six items (α = .848), such as In order to … . I must have enough knowledge of phrases and collocations. Pwrit consists of five items (α = .852), such as In order to get a good score on the essay writing, I must write in smooth and fluent English. In addition, the perception questionnaire uses two items to assess test-takers’ perceived impact of component weighting (P-weight) and score profiling (P-profile); these two items can be found in the Appendix.

The preparation questionnaire asks test-takers about their focused test-preparation practice in five subscales comprising: Test-preparation management (α = .871/10 items), e.g., During the test-preparation period in the past eight weeks, I analyzed test papers to identify frequently assessed areas; Rehearsing test-taking skills (α = .888/7 items), e.g., I trained myself to choose options through logical elimination; Socio-affective strategies (α = .836/6 items), e.g., I tried to learn from others about ways to improve test scores; Drilling (α = .782/4 items), e.g., I focused on improving the target reading skill to understand key sentences in a reading passage; Memorizing (α = .835/5 items), e.g., I recited sentence patterns. Finally, the questionnaire assesses test-takers’ engagement with six language skill development activities (α = .883/6 items) and their time management. The six items are in the Appendix.

1 Cases were considered to be invalid (1) if the response patterns on the questionnaire clearly indicated that the respondents were not serious in answering the questionnaire items, or (2) if answers to more than 30% of the items were missing.
4.3. Data analysis

SPSS 16.0 was used for cleaning data, examining item distributions, and computing composite variables. Mplus 5.0 (Muthén & Muthén, 2010) was used for statistical modeling. Mplus was selected among other SEM packages primarily because it provides multiple robust estimation methods, which can handle data sets with non-normal distributions and missing values. These two problems are regularly encountered by empirical researchers when collecting longitudinal data at multiple time points. Examination of variable distributions found some items with minor to moderate skewness and kurtosis statistics (Table 4, LD1-6), which indicated that these variables did not have normal distributions. Moreover, because data were collected on three occasions over an extended period, there were missing values in the dataset. Items with minor to moderate distribution problems were carefully examined individually to ensure that their distributions were not due to incorrect data entry. MLR, a robust maximal likelihood estimation method that accommodates skewed distributions and missing values, was adopted to ensure the accuracy of statistical modeling (Byrne, 2011; Muthén & Muthén, 2010).

The use of a 6-point scale for the perception questionnaire was also related to concerns of item distributions. Test-taker perceptions were scored on a 5-point scale in the pilot study, but the results were found to be skewed (with test-takers tending to agree strongly). Because a finer scale could detect a wider variance for each item, thereby attenuating the skewness problem in the pilot study, a 6-point scale was used in the present study for the perception questionnaire. On the other hand, because items of the TPP did not encounter such a problem, a 5-point scale was still used for this questionnaire in the present study. This difference in scaling across the two questionnaires did not affect the results of Structural Equation Modeling, where results were standardized to be unit free. However, when interpreting descriptive statistics such as means and standard deviations, readers need to note the differences in the scale ranges.

Nine subscale-level composite variables, including four indicators of perceived language skills (of linguistic knowledge, listening, reading, and writing) and five indicators of test preparation, were computed by averaging item scores within each subscale. Thus, the score ranges of the nine composite variables are the same as those of the items. These composite variables were used as observed indicators for the two latent factors in the structural model. Five popular model fit indices were used to evaluate statistical models. Chi-square statistic indicates the absolute difference between the hypothetical model and the data set. Besides Chi-square statistic, four other comparative model fit indices were used: Root mean square error approximate (RMSEA), standardized root mean square error (SRMR), comparative fit index (CFI), and Tucker–Lewis index (TLI). RMSEA values lower than .05 and SRMR values lower than .08 are considered to be close fits. CFI and TLI values larger than .90 are considered to be good, and larger than .95 are considered to be close model fits (Hu & Bentler, 1999).

Table 3
Descriptive statistics for variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Mini</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CET4 listening</td>
<td>934</td>
<td>86</td>
<td>200</td>
<td>131.45</td>
<td>18.81</td>
<td>.312</td>
<td>.250</td>
</tr>
<tr>
<td>CET4 reading</td>
<td>935</td>
<td>86</td>
<td>215</td>
<td>145.15</td>
<td>21.79</td>
<td>.087</td>
<td>.115</td>
</tr>
<tr>
<td>CET4 writing</td>
<td>933</td>
<td>49</td>
<td>119</td>
<td>79.16</td>
<td>11.04</td>
<td>.260</td>
<td>.810</td>
</tr>
<tr>
<td>CET4 integrative</td>
<td>929</td>
<td>26</td>
<td>66</td>
<td>43.46</td>
<td>6.16</td>
<td>.200</td>
<td>.578</td>
</tr>
<tr>
<td>Plng</td>
<td>887</td>
<td>3.00</td>
<td>6.00</td>
<td>4.75</td>
<td>.55</td>
<td>-.024</td>
<td>.072</td>
</tr>
<tr>
<td>Plist</td>
<td>887</td>
<td>2.86</td>
<td>6.00</td>
<td>4.60</td>
<td>.57</td>
<td>.159</td>
<td>-.091</td>
</tr>
<tr>
<td>Pred</td>
<td>887</td>
<td>3.00</td>
<td>6.00</td>
<td>4.95</td>
<td>.59</td>
<td>-.123</td>
<td>-.404</td>
</tr>
<tr>
<td>Pwrit</td>
<td>887</td>
<td>3.50</td>
<td>6.00</td>
<td>5.04</td>
<td>.55</td>
<td>-.134</td>
<td>-.342</td>
</tr>
<tr>
<td>LD 3_1</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.69</td>
<td>.86</td>
<td>1.328</td>
<td>1.830</td>
</tr>
<tr>
<td>LD 3_2</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.68</td>
<td>.87</td>
<td>1.403</td>
<td>2.032</td>
</tr>
<tr>
<td>LD 3_3</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.66</td>
<td>.87</td>
<td>1.277</td>
<td>1.236</td>
</tr>
<tr>
<td>LD 3_5</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.42</td>
<td>.75</td>
<td>2.088</td>
<td>4.754</td>
</tr>
<tr>
<td>LD 3_6</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.50</td>
<td>.87</td>
<td>1.945</td>
<td>3.577</td>
</tr>
<tr>
<td>LD 3_7</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.55</td>
<td>.89</td>
<td>1.864</td>
<td>3.354</td>
</tr>
<tr>
<td>TTS</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>3.43</td>
<td>.77</td>
<td>-.571</td>
<td>.358</td>
</tr>
<tr>
<td>TPM</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>3.18</td>
<td>.70</td>
<td>-.235</td>
<td>.262</td>
</tr>
<tr>
<td>MEM</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>2.79</td>
<td>.88</td>
<td>-.109</td>
<td>-.425</td>
</tr>
<tr>
<td>SOAF</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>2.45</td>
<td>.82</td>
<td>.301</td>
<td>-.229</td>
</tr>
<tr>
<td>DRILL</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>2.85</td>
<td>.80</td>
<td>-.112</td>
<td>-.377</td>
</tr>
</tbody>
</table>

Note: N = 887 for P3.1–3.2, N = 872 for TP2.1–2.4.

Table 4
Descriptive statistics for observed variables in structural models.

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Mini</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CET4 listening</td>
<td>934</td>
<td>86</td>
<td>200</td>
<td>131.45</td>
<td>18.81</td>
<td>.312</td>
<td>.250</td>
</tr>
<tr>
<td>CET4 reading</td>
<td>935</td>
<td>86</td>
<td>215</td>
<td>145.15</td>
<td>21.79</td>
<td>.087</td>
<td>.115</td>
</tr>
<tr>
<td>CET4 writing</td>
<td>933</td>
<td>49</td>
<td>119</td>
<td>79.16</td>
<td>11.04</td>
<td>.260</td>
<td>.810</td>
</tr>
<tr>
<td>CET4 integrative</td>
<td>929</td>
<td>26</td>
<td>66</td>
<td>43.46</td>
<td>6.16</td>
<td>.200</td>
<td>.578</td>
</tr>
<tr>
<td>Plng</td>
<td>887</td>
<td>3.00</td>
<td>6.00</td>
<td>4.75</td>
<td>.55</td>
<td>-.024</td>
<td>.072</td>
</tr>
<tr>
<td>Plist</td>
<td>887</td>
<td>2.86</td>
<td>6.00</td>
<td>4.60</td>
<td>.57</td>
<td>.159</td>
<td>-.091</td>
</tr>
<tr>
<td>Pred</td>
<td>887</td>
<td>3.00</td>
<td>6.00</td>
<td>4.95</td>
<td>.59</td>
<td>-.123</td>
<td>-.404</td>
</tr>
<tr>
<td>Pwrit</td>
<td>887</td>
<td>3.50</td>
<td>6.00</td>
<td>5.04</td>
<td>.55</td>
<td>-.134</td>
<td>-.342</td>
</tr>
<tr>
<td>LD 3_1</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.69</td>
<td>.86</td>
<td>1.328</td>
<td>1.830</td>
</tr>
<tr>
<td>LD 3_2</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.68</td>
<td>.87</td>
<td>1.403</td>
<td>2.032</td>
</tr>
<tr>
<td>LD 3_3</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.66</td>
<td>.87</td>
<td>1.277</td>
<td>1.236</td>
</tr>
<tr>
<td>LD 3_5</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.42</td>
<td>.75</td>
<td>2.088</td>
<td>4.754</td>
</tr>
<tr>
<td>LD 3_6</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.50</td>
<td>.87</td>
<td>1.945</td>
<td>3.577</td>
</tr>
<tr>
<td>LD 3_7</td>
<td>872</td>
<td>1</td>
<td>5</td>
<td>1.55</td>
<td>.89</td>
<td>1.864</td>
<td>3.354</td>
</tr>
<tr>
<td>TTS</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>3.43</td>
<td>.77</td>
<td>-.571</td>
<td>.358</td>
</tr>
<tr>
<td>TPM</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>3.18</td>
<td>.70</td>
<td>-.235</td>
<td>.262</td>
</tr>
<tr>
<td>MEM</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>2.79</td>
<td>.88</td>
<td>-.109</td>
<td>-.425</td>
</tr>
<tr>
<td>SOAF</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>2.45</td>
<td>.82</td>
<td>.301</td>
<td>-.229</td>
</tr>
<tr>
<td>DRILL</td>
<td>872</td>
<td>1.00</td>
<td>5.00</td>
<td>2.85</td>
<td>.80</td>
<td>-.112</td>
<td>-.377</td>
</tr>
</tbody>
</table>
5. Results and findings

5.1. Impact of component weighting and score profile

The descriptive statistics of all variables in model 1 (Fig. 1) are presented in Table 3. P-weight and P-sprofile ask about the perceived impact of component weighting and score profiling. They have similar means (mean = 3.90 and 3.80) and similar standard deviations (SD = 1.182 and 1.192). On a 6-point scale where 3 = some, 4 = medium, 5 = strong, the two means suggest the perceived impact of the two features is moderate. Roughly, 34% of test-takers perceived some but not much impact (2.7 < 3.9) and another 34% perceived much stronger impact (3.9 < 5.1). Variables 3–5 ask test-takers to estimate the time they spent practicing listening (T-listening), reading (T-reading), and writing (T-writing) on a 5-point scale. On average, test-takers spent about half an hour per day practicing listening (mean = 1.84) and reading (mean = 2.09), but less on writing (mean = 1.51). They spent more time practicing reading than listening (t = 8.90, p < .001, Cohen’s d = .297, small effect), and much less on writing (t = 20.17, p < .001, Cohen’s d = .686, medium effect).

Fig. 1 presents the results of path analysis, which estimates the relations between perceived impact and test-taker’s time management during preparation. Specifically, the model estimated whether test-takers who perceived a larger impact of component weighting and score profiling also spent more time on practicing listening, reading, and writing respectively. The error terms of the three variables relating to time were fixed to be correlated for model identification (Byrne, 2011). There are two reasons for these error terms to be correlated. The first reason is method effect, because the three items share similar wording and are listed together for the test-takers; the second is because the three items asked test-takers about their time spent listening, reading, and writing, test-taker responses may be positively correlated to each other because they are all indicators of test-taker effort devoted to preparation and motivation. Forcing these three error terms to be uncorrelated may thus seem unwarranted, whereas freeing the error terms to be correlated is more appropriate.

Fig. 1. Path model with standardized parameters.

---

2 Cohen’s d is calculated using equation 8 for the effect size of within-subject designs in Morris and DeShon (2002).
The path model achieves excellent model fitness \( \chi^2 = 13.495, df = 8, \text{RMSEA} = .026, \text{SRMR} = .020, \text{CFI} = .993, \text{TLI} = .975 \), which indicates that the model is a close representation of the variance and covariance matrix among the observed variables in the model. Standardized coefficients suggest that perceptions positively contribute to test-takers' time spent listening, reading, and writing, but only the path to T-listening is statistically significant. This is probably because the adjustments made to writing and reading are relatively minor, i.e., 5% (from 40% to 35% for reading and from 15% to 20% for writing). Such minor adjustments may not be considered by test-takers as sufficiently significant to adjust their time distribution. Specifically, the minor reduction of weight from the reading section did not seem to discourage them from spending less time on reading. Similarly, the minor amount of weight increase on writing did not seem to encourage test-takers to spend more time on writing. Only the path from perceptions to T-listening reaches statistical significance \( (\beta = .094, p < .05) \), probably because the listening component receives a much greater weight adjustment, i.e., 15% (from 20% to 35%).

Time spent listening (T-listening), in turn, positively contributes to test scores in the listening component \( (\beta = .117, p < .01) \) and the reading component \( (\beta = .089, p < .05) \), though the contributions are fairly small. On the other hand, time spent reading (T-reading) is not associated with better test performance on listening, reading or writing. The finding that T-listening has a positive impact on listening and reading but T-reading does not is interesting and clearly warrants further investigation. One plausible explanation is related to a popular strategy test-takers utilized in preparing for listening tests, that of studying the listening transcripts and consulting them recursively whenever they encountered specific difficulties in their listening. As this strategy engages both listening and reading, it may help to enhance both skills. It is not clear however, why time spent on reading did not improve the performance on the reading component. Finally, time spent writing (T-writing) is negatively associated with scores on listening \( (\beta = -.094, p < .05) \) and reading \( (\beta = -.093, p < .05) \). This pattern is surprising and is different from the previous studies. Reasons for this surprising finding are discussed later in association with washback.

5.2. Impact of perceived test validity

Table 4 reports the descriptive statistics of the variables used in this round of modeling. Except for the six variables (LD1 to 6) that have minor to moderate non-normal distributions, all other variables have normal distributions. As reported earlier, LD1 to 6 assess test-takers' engagement with language learning activities during the test preparation period. Because many test-takers did not engage in language learning, the distributions of LD1 to 6 are skewed. These skewed distributions, however, are accurate depiction of students' disengagement with language learning activities during test preparation. As the robust MLR method was used to estimate the structural model, the minor non-normality in the data set was of little consequence.

The four variables measuring perceptions of test validity (Ping to Pwrit) have rather high means. On a 6-point scale, all means are close to 5.0, suggesting that test-takers agreed that in order to answer the questions in the R-CET4 correctly they must possess adequate linguistic knowledge, and listening, reading and writing skills. The five test preparation variables (TTS to DRILL) have means ranging from 2.45 to 3.43 on a 5-point scale. In comparison, the language skill development strategies have much lower measures (LD1 to 6 means lower than 1.7). This finding suggests test-takers used the five test-preparation strategies much more frequently than the language skill development strategies. In fact, the means suggest test-takers used the LD strategies rarely; some did not use them at all.

To sum up, test-takers seemed to hold favorable perceptions regarding the validity of the R-CET4. They rehearsed test-taking skills, drilled and memorized, but scarcely used any language development strategies such as reading extensively in English, listening to authentic English broadcasts or using English to communicate.

Latent structural modeling was used to estimate the impact of perceived test validity on test-taker approaches to preparation. The results are presented in Table 5 and Fig. 2. Table 5 presents the model fitness for the measurement and the structural model. Both models achieve excellent fitness, with RMSEA and SRMR values close to .05, and CFI and TLI values well above .90. The measurement model involves all four latent factors and their 18 observed indicators. For the consideration of space, only the structural model is presented in Fig. 2, with standardized parameter estimations.

For a similar reason as reported earlier, an error correlation between Language Development activity (LD) and Test Preparation (TP) was fixed in this model. By fixing this error correlation, the structural model recognizes the possibility that test-takers' approaches to preparation may well be influenced by a factor not included in the model, which could have an effect on both LD and TP (Byrne, 2011). It is conceivable that highly motivated test-takers would use both LD and TP more frequently than test-takers with lower motivation. Factors like motivation, though not directly related to this study, should be recognized in model specification so as to enhance the accuracy in parameter estimation. Model comparison (Table 5) between the final model and a model without the correlated error terms (Model 2) also suggests that the final model is significantly better \( (\Delta \chi^2 = 118.103, df = 1, p < .001) \), thus lending support for correlating the two error terms.

<table>
<thead>
<tr>
<th>Models</th>
<th>( \chi^2 )</th>
<th>( df )</th>
<th>( \chi^2/df )</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>CFI</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Measurement model</td>
<td>558.198</td>
<td>146</td>
<td>3.823</td>
<td>.053</td>
<td>.049</td>
<td>.935</td>
<td>.924</td>
</tr>
<tr>
<td>2. Structural model (Fig. 2)</td>
<td>560.747</td>
<td>147</td>
<td>3.814</td>
<td>.053</td>
<td>.050</td>
<td>.934</td>
<td>.924</td>
</tr>
<tr>
<td>3. Structural model without correlating error terms</td>
<td>678.850</td>
<td>148</td>
<td>4.586</td>
<td>.060</td>
<td>.101</td>
<td>.916</td>
<td>.903</td>
</tr>
</tbody>
</table>
Standardized path coefficients suggest that favorable perceptions of test validity significantly contribute to test preparation (TP) \( (\beta = .389, p < .01) \), which, in turn, contribute to test performance \( (\beta = .230, p < .01) \). Favorable perceptions also positively contribute to LD though the contribution does not reach statistical significance \( (\beta = .084, \text{n.s.}) \). However, test-takers with a higher level of LD tended to perform less well \( (\beta = -.111, p < .05) \). Finally, the direct effect of perceptions (P-test validity) on test performance is not significantly different from zero. This finding suggests that after partitioning the indirect effects, test-taker perception does not have a direct effect on test performance.

6. Discussion

To bring about positive washback on teaching and learning, a series of changes were made to the CET4, with two major ones related to its design. The first set of changes was to adjust component weighting and provide component scores in a score profile. The second set of changes was to add less-closed testing methods and reduce the use of MC items. This study investigated test-taker perceptions of the two sets of changes and their effects on test-taker time management, approaches to test preparation and test performance. The purpose of this study was to investigate the mechanism of washback underlying large-scale testing. CET4 itself was not the focus, though its reform provided an opportunity for such an investigation. The study found test-taker perceptions had small but significant effects on their time spent practicing listening, but not that on reading and writing. Furthermore, test-takers tended to perceive the R-CET4 favorably in terms of its ability to assess genuine English language skills and knowledge. Such favorable perceptions, however, only had marginal effects on the learning approach (positive washback). Meanwhile, favorable perceptions were also associated with a significantly higher level of focused test preparation (negative washback).

The first finding suggests that applying component weighting to leverage curriculum innovation exerts a certain, albeit small, influence on test-takers. Firstly, the more test-takers endorsed the importance of the adjusted weighting scheme, the more time they would spend listening \( (\beta = .094, p < .05) \). Test-takers who endorsed the adjusted weighting scheme spent more time listening than those who did not. Secondly, although test-takers still spent more time reading than listening, the difference between the two was small \( (\text{Cohen’s } d = .297, \text{small effect}) \). Although the listening component has the same weight \( (35\%) \) as the reading, the listening component was considered by test-takers to be less coachable and more difficult to improve over a short period of time than the reading \( (\text{Author 1}) \). The consideration of coachability \( (\text{Powers, 1987}) \) may be one of the reasons for test-takers to spend more time reading than listening. However, contrary to the belief about the

![Fig. 2. The structural model with standardized parameter estimations.](image-url)
ineffectiveness of practicing listening, spending time listening was found to contribute significantly to both listening ($\beta = .117$, $p < .01$) and reading performance ($\beta = .089$, $p < .05$), whereas time spent reading alone did not seem to help. Though the exact reasons warrant further investigation, test-takers’ engagement in integrated listening and reading activities is likely to be one reason. Teachers usually use integrated reading and listening activities to help students develop skills to use the target language for authentic life tasks. This finding suggests these activities can also be used during test preparation to enhance test performance on individual skill components. On the other hand, since time spent reading did not have an effect on reading performance, teachers and students should spend less time drilling the reading component.

Although a score profile may have the potential to expose and penalize selective neglect of the components carrying low weights, this measure was not as effective as would be desired. There is little difference in the time spent writing between test-takers who endorsed the importance of this measure and those who did not ($\beta = .064$, n.s.). Knowing that the score profile would expose their writing scores did not push test takers to spend more time writing. Compared with time spent on reading and listening, much less time was spent writing (Cohen’s $d = .686$, medium effect). This is a cause of concern since the writing component is also rated as the most valid form for assessing language skills by both test-takers and test developers (Yang & Weir, 1998), i.e., the most valid component received the least attention. This general lack of attention to the writing component may be related to its relatively low weighting (15% of the total) and the complementary scoring system that the R-CET4 is operating on. It may also because practicing writing is more laborious and less rewarding. While it is relatively easier for test-takers to check their answers to the reading and listening items and to learn from mistakes, in a typical test-preparation setting, test-takers do not get prompt feedback on the quality of their writing. Thus, they can hardly learn how to improve their writing. This in turn makes the writing practice frustrating if not futile. Further studies are needed to investigate what test developers and teachers can do to enhance test-taker engagement with the writing component. It is worth noting that on-line automated essay scoring engines are now available which can provide immediate feedback to student writers (Enright & Quinlan, 2010). Teachers may want to explore this technology to encourage students to practice writing and to establish a classroom environment of regular writing. Besides improving language proficiency, regular and productive engagement with the target language will cultivate a habit of writing and motivate students to learn. The latter is especially important for language learners living in an environment where the target language is a foreign language which is seldom used for real-life communication. In such a context, creating opportunities for using the target language (in writing and speaking) can be as important as providing high-quality instructions.

Relating to component weighting, this study observed relatively low correlations among the four test components, which range from the lowest (.263) between the writing and the integrative skills component, to the highest (.412) between the reading and the listening components. This suggests that the four components are distinct, and are not compensatory to each other in terms of measurement (Kane & Case, 2004). Thus, the use of a gross composite score based on a complementary scoring approach may not be appropriate, as it can hide important skill deficiency. It seems necessary to set minimal requirements for individual test components, especially those with lower weighting such as the writing component in this study, so that test-takers have to give all components due attention.

The second major finding regarding the effects of perceived test validity on test preparation is interesting. Following the assumption about the deterministic relationship between test validity and positive washback, we would expect favourable perceptions to be associated with an increase of positive washback (more engagement in language skill development activities) and a decrease in negative washback (such as drilling and cramming based on past test papers). In other words, there should be a positive (+) path between perceptions and language development activities and a negative (−) path between perceptions and focused test preparation. The first path is indeed positive (albeit not statistically significant, $\beta = .084$). However, the second path is also found positive (and statistically significant, $\beta = .389$, $p < .05$). Favorable perceptions seemed to encourage positive washback in that test-takers reported slightly higher level of engagement with language learning activities. However, rather than discouraging cramming and drilling, favorable perceptions seemed to intensify such practices. The assumption that enhanced test validity would discourage cramming and drilling is clearly unwarranted. Washback is indeed a “blunt instrument” (Andrews, 1995). Favourable perceptions of test validity are associated with both positive and negative washback, which take place simultaneously and, interestingly, are not contradictory to each other.

Counter-intuitive at first sight, this finding is understandable with further thought. Test preparation involves primarily drilling and cramming based on previous test papers. If test-takers endorsed the validity of a test and hence, the credibility of the test papers, practicing test papers could be a logical way forward, because practicing (valid) test papers would enhance test performance as well as language ability. Regardless of test design, the alignment of content in testing, teaching, and learning seems inevitable for the purpose of test preparation. High-stakes testing acts like a curriculum magnet, attracting all learning resources to focus on a narrow range of skills and knowledge that will appear in the test. Instead of discouraging test-takers from cramming and drilling, favorable perceptions of test validity may provide additional incentive for doing so. Although changes in test design seemed to direct test-taker attention toward learning activities with broader focuses, the influence was rather marginal. The fact that negative and positive washback took place simultaneously suggests that the mechanism for engineering positive washback is likely to be distinct from the one that prevents negative washback. Consistent with previous studies (Cheng, 2005; Qi, 2011), this study found that cramming and drilling are likely to continue regardless of test design. In fact, they may be inevitable as long as they indeed boost test scores.

This study found that cramming and drilling were indeed effective in boosting test scores, echoing previous findings on test preparation (Farnsworth, 2013; Green, 2007; Koretz, 2009; Messick, 1982; Qi, 2011). This finding is not surprising. Like other measurement instruments, language proficiency tests rely on sampling where test-taker performance on a given
sample of language tasks is used to infer their performance over a much broader domain of tasks involving the target language. The inferences from test score to real-life performance only holds when the tasks in a test are sampled randomly and unpredictably to adequately represent the domain. In practice, however, the language tasks in standardized tests are neither random nor unpredictable but are limited and restricted to certain skills and test formats (and necessarily so for the sake of reliability). Thus, to a certain extent, standardized tests are always predictable. Because of this inherent limitation of standardized tests, cramming and drilling can boost test scores, creating inflated test scores (Koretz, McCaffrey, & Hamilton, 2001) and leading to the widely acknowledged phenomenon of high-scorer but of low ability (see more about this in Qi, 2011).

Furthermore, this study found that test-takers who reported a higher level of engagement with language development activities (LD) and those who spent more time on writing performed less well. Precise reasons for this counter-intuitive finding clearly warrant further investigation. One speculation is related to test-takers’ strategic competence (see Phakiti, 2008), that is, the ability to use strategies appropriately according to perceived task requirement. Test performance may be affected by test-takers’ (lack of) strategic competence, which may have confounded the effects of LD on performance. Compared with test preparation, LD is better considered as a long-term strategy, which may not be appropriate for short-term test preparation. Test-takers who engaged in LD during test preparation may be those who were not sufficiently strategic and hence unable to adjust their learning strategies. Their lack of strategic competence may negatively affect their test performance, thus confounding the effects of LD. Although scholars in language testing have long argued that strategic competence is one essential component of communicative competence (Bachman & Palmer, 1996; Phakiti, 2008), exactly how it should be assessed and factored into language tests remains unclear. Even if strategic competence can be reliably measured, it will be controversial whether to control it or to measure and report it as a legitimate aspect of the language proficiency construct. Further studies may investigate characteristics of those test-takers who engage in extensive language-learning activities and spend time practicing writing during test preparation to better understand why they perform less well. To help these students so that they will not be disadvantaged in test taking, teachers can provide them tutoring on effective test-preparation and test-taking strategies and advise them how to make appropriate use of these strategies according to different task requirements.

7. Conclusion

In conclusion, this study found adjusting component weighting can affect test-taker time management, butos effects may be small. As long as composite scores are used for making decisions, reporting component scores may not be able to adjust the test-taker tendency to neglect the components with lower weighting. Favorable perception of test validity may bring about some desirable washback (albeit small), but it can also intensify negative washback. The mechanism for engineering positive washback seems to be distinct from the one that discourages negative washback. Given the paucity of quantitative studies on specific aspects of test design and their washback, this study is considered a worthwhile step forward. However, the study has several limitations. First, only test-takers preparing for R-CET4 were studied. The study would be strengthened, if it could have involved test-takers who took both CET4 and R-CET4 and compared their perceptions and preparation practices before and after the reform. Second, because all test-takers are from one lower tier university, findings of this study regarding the impact of R-CET4 cannot be reliably extended to the universities with more able students. It is conceivable that test-takers in lower-tier universities may have different motivation towards CET4 which may affect their test preparation. Thus, the limited influence of the writing component may be specific to lower-tier universities similar to the one under study. Finally, since only the test-preparation period was under study, findings regarding the mechanism of washback cannot be extended to non-test-preparation periods. It is possible that when the test is not an immediate concern, test-takers may exhibit different approaches to learning, and their time management may also be different. Further studies could look beyond one single setting and beyond the test-preparation period for a more comprehensive understanding and modeling of the washback mechanism.

Acknowledgment

The author wants to thank the teacher and student participants of this study, especially Ms. Li Anhong, Ms. Cen Rentao, Ms. Jiang Tongling, Ms. Nei Xiaoying, and Miss. Chen Zhiwei. In preparing the manuscript for publication, the author has received partial support from the Department of Linguistics and Modern Language Studies and the Research and Development Office of the Hong Kong Institute of Education (project #R6451 and #R3565).

Appendix

P-weighting: In R-CET4, Listening takes 35% of the total score; Reading takes 35%, Essay 15%, Cloze 10%, and Translation 5%. Will this score distribution influence your preparation for CET4?

P-score profile: The R-CET4 provides a score profile, where test scores for the four components, listening, reading, writing, and integrative skills, are reported separately. Will this influence your preparation for CET4?

During test preparation period in the past eight weeks,
LD1. I kept on practicing my spoken English;
LD2. I kept on reading extensively e.g., English newspapers/websites.
LD3. I kept on listening to radio broadcasting in English.
LD4. I kept on writing diaries/blogs in English.
LD5. I kept on communicating with English native speakers whenever possible.
LD6. I kept on using English whenever possible, e.g., writing emails.

References
Andrews, S. (1995). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berry, & V. Berry (Eds.), Bringing about change in language education (pp. 67–82). Hong Kong: Department of Curriculum Studies, the University of Hong Kong.
Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling

Qin Xie and Stephen Andrews

Language Testing 2013 30: 49 originally published online 9 July 2012
DOI: 10.1177/0265532212442634

The online version of this article can be found at:
http://ltj.sagepub.com/content/30/1/49

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Language Testing can be found at:

Email Alerts: http://ltj.sagepub.com/cgi/alerts

Subscriptions: http://ltj.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://ltj.sagepub.com/content/30/1/49.refs.html

>> Version of Record - Jan 13, 2013

OnlineFirst Version of Record - Jul 10, 2012

OnlineFirst Version of Record - Jul 9, 2012

What is This?
Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling

Qin Xie
Hong Kong Institute of Education, Hong Kong

Stephen Andrews
University of Hong Kong, Hong Kong

Abstract
This study introduces Expectancy-value motivation theory to explain the paths of influences from perceptions of test design and uses to test preparation as a special case of washback on learning. Based on this theory, two conceptual models were proposed and tested via Structural Equation Modeling. Data collection involved over 870 test takers of College English Test Band 4 in China. A perception of assessment questionnaire was given at the beginning of a 10-week preparation period; a test preparation questionnaire was given eight weeks later. Test takers who endorsed high-stakes, instrumental test uses as the primary purpose for taking the test tended to value test taking; test takers who perceived test design positively tended to attach high importance to test taking and appeared more confident. Furthermore, higher endorsed task value and higher expectation of test success jointly contributed to greater engagement in test preparation. Knowledge of the test was also related to increased self-regulation in test preparation and more practice of test-taking skills.

Keywords
Expectancy-value theory, perceptions of assessment, Structural Equation Modeling, test use, washback

Corresponding author:
Qin Xie, English department, Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong SAR.
Email: qxie@ied.edu.hk
Introduction

Washback or backwash, a term widely used in language testing and applied linguistics, refers to the influence of testing on teaching and learning. Although the number of studies on washback have increased considerably since the seminal work of Alderson and Wall (1993), it is still not clear precisely how testing influences teaching and learning. Many studies on washback (e.g. Cheng, 2005; Qi, 2005) are triggered by concerns about the negative influence of testing on teaching and learning, such as teaching and learning to the test and excessive practice of test papers, which lead to narrowing of the curriculum and downplaying the goal of improving language ability.

Some educators and critics believe the source of negative washback is standardized testing, especially the Multiple Choice test format. Some testers, however, argue that test design does not affect the nature of washback; it is the misuse or abuse of test results that triggers negative washback. Despite heated debates in language testing, we are not aware of any study that has actually examined the effects of test design and test use in order to better understand influences of testing on teaching and learning. Although multiple factors have been identified that affect the nature and scope of washback, there is little agreement regarding the differential effects of test design and test uses. It is still not clear how these factors work together, which ones are more salient, or in what ways test design and test use influence teaching and learning.

Many reasons account for the current state of research on washback. One is methodological. Most existing washback studies are primarily qualitative and exploratory (e.g. Alderson & Hamp-Lyons, 1996; Cheng, 2005). These studies have identified multiple factors contributing to the washback phenomenon, yet they do not assess the relations among these factors statistically. Thus, there is a need for quantitative studies of these mediating factors to examine the relationships between them. Another reason is the lack of an appropriate theory to guide statistical examination of the relationships among the numerous factors identified. Without a sound theory, statistical modeling is data driven and post hoc, and therefore prone to error.

In light of the state of research on washback, the present study sets out to examine this phenomenon quantitatively via the guidance of an established theory. Expectancy-value motivation theory is taken from learning psychology as a way to interpret the mechanism of washback on learning during test preparation. A model of washback on learning is conceptualized on the basis of this theory and verified statistically via Structural Equation Modeling.

Test design vs. test use

A central issue in the debate regarding the influence of test design and use on teaching and learning concerns the source of adverse consequences and the parties to be held accountable. Testers argue that tests are neutral measurement tools and therefore harmless. Adverse consequences (i.e. negative washback in different forms) are primarily a function of test misuse and abuse; therefore, the (mis)users should be held responsible. Shohamy (2001), for instance, argues that a test is powerful not because of its technical strength but because of the social and/or political functions that tests are used to perform.

Educators and critics, however, believe there is a connection between the way a test is designed, and the way teachers teach and students learn. They criticize standardized testing for promoting exam-driven, superficial learning approaches, and demand reform of the assessment system (e.g. Entwistle & Entwistle, 1992; Fredericksen & Collins, 1989; Frederiksen, 1984). One central criticism of standardized testing focuses on its dominant use of Multiple Choice Questions (MCQ). Educators believe MCQ tend to assess knowledge-based memorization and factual recall, but not complex, high-level thinking skills (e.g. Fulcher, 2000; Scouller, 1998). Because deep learners may perform more poorly on MCQ-based examinations than surface, rote learners (e.g. Scouller, 1998), opponents argue that standardized testing discourages the teaching and learning of high-level cognitive skills. Based on these educational considerations, Fredericksen and Collins (1989) argue that test validation should incorporate an examination of its consequences. Accordingly, if a test fails to induce positive washback, it is not systematically valid.

Despite these debates, no consensus has yet been reached regarding the differential effects of test design and use on learning. Although each side has evidence to support its stand, no empirical studies have been conducted to tease out the effects of test design from those of test use by examining both simultaneously within one model. If test design and test uses do affect the teaching and learning that precede testing, it is likely that their influences follow different paths.

**Washback mechanism: The black box**

Most existing washback studies are primarily qualitative and exploratory (e.g. Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993; Andrews, Fullilove, & Wong, 2002; Cheng, 2005; Gosa, 2004; Wall & Alderson, 1993; Watanabe, 1996). Many focus on teachers and teaching; few focus on learners and learning. Existing studies have found washback to be highly complex, with contextual factors and personal factors all playing a role in shaping its nature and scope (e.g. Wall & Alderson, 1993). Within classrooms, teacher factors are found to be pivotal (Alderson & Hamp-Lyons, 1996; Qi, 2005). In a given educational setting, contextual factors and individual factors interact and largely determine both the kind and amount of washback that occurs. However, it is still not clear in what ways these factors combine and interact with each other.

Compared with washback on teachers and teaching, studies on learners and learning are scarce. Only a few empirical studies of washback have focused on learning and learners. For example, in Gosa’s (2004) study in Romania, she found students’ expectation of assessment was the single most important factor in explaining students’ perspectives on teaching and learning activities. The students’ expectation of assessment not only influenced their attitudes towards teaching, but also their own learning. Nine out of the 10 diarists in the study did not prepare for the test at all because of their expectation that the test was easy. Similarly, Green (2007) compared an IELTS preparation course with university language courses. He concluded that individual learners’ goals and their
'understanding of test demand' (p. 93) influenced their learning outcomes to a greater extent than their choice of course and its content. Although the precise nature and process of washback on learning are still unclear, studies from learning psychology (cf. Struyven, Dochy, & Janssens, 2005) suggest that students coordinate their strategy use with their expectations of assessment demands.

**Expectancy-value theory**

In the area of learning psychology, effects of test expectation on learning are studied within the Expectancy-value motivation framework. Expectancy-value theory posits that 'Individuals’ choice, persistence, and performance can be explained by their beliefs about how well they will do on the activity and the extent to which they value the activity' (Wigfield & Eccles, 2000, p. 68). The full model for Expectancy-value theory is presented in Figure 1. This model (Jacobs & Eccles, 2000) is developed for the academic context on the basis of Atkinson’s (1964) Expectancy-value theory. Elaboration of the full model is beyond the scope of this paper. Of interest to this study are the four cells on the right-hand side of the full model. *Expectation of success* and *Subjective task value* are believed to be two crucial motivational factors jointly mediating the effects of a child’s goals and general self-schemata on achievement-related choices and performances. Personal short-term and long-term goals and perceptions of task demand are three of the six factors under *Child’s*
Goals and General Self-schemata, which are believed to influence expectation of success and task evaluation, and which, in turn, affect Achievement-Related Choices and Performance.

Expectancy-value theory posits that when facing a task and making related choices, students ask themselves two questions: ‘Do I want to do it?’ and ‘Can I do it?’ (Eccles & Wigfield, 2002). If they answer ‘yes’ to both, then they are more likely to engage in the task and perform well on it. Students’ answers to these two questions are believed to be affected by their personal goals and perceptions of task demands.

Expectation of success refers to individuals’ beliefs about how well they will do on upcoming tasks. In an academic context, this refers to students’ beliefs about their capability to conduct a given academic task successfully (Schunk, 1991). From the student perspective, this is analogous to answering the question ‘Can I do it?’ Expectation of success is closely related to the construct of self-efficacy (Bandura, 1982). Conceptually, self-efficacy stresses individual confidence towards the task, while the expectation construct in Expectancy-value theory stresses the expectation of outcomes. Empirically, these two constructs are closely related. In its operational form, expectation of success is often measured analogously to self-efficacy. Subjective task value is the extent to which individuals value the upcoming task as desirable. From the student perspective, the construct of task value is analogous to answering the question ‘Do I want to do it?’ Task value is operationalized in terms of attainment value, intrinsic interest, utility value, and costs (Jacobs & Eccles, 2000).

Numerous empirical studies have been conducted in connection with Expectancy-value theory and the key related constructs (e.g. Bandura, 1982; Bong, 2001; Pajares, 1996; Pintrich, 1999; Wigfield, 1994). Most of these have focused on the effects of goals and self-efficacy on academic achievement. Fewer studies examine their effects on learning. Pintrich and his colleagues are the most noted group focusing on learning and strategy use. Their studies (e.g. Pintrich, 1999, 2000) found that students’ self-efficacy and values consistently related positively to strategy use.

Conceptual models of washback on learning

The analogy between students facing an academic task and facing a test-taking challenge is straightforward. Preparation for test taking can be regarded as a special instance of learning, which is likely to be affected by test takers’ expectancy of success and the values attached to success. Expectation of success and test value can be two useful factors for explaining the paths of influence from testing to preparation. From this perspective, preparation for test taking, as a special case of washback on learning, can be understood as test takers’ learning behaviors directed by personal goals and in response to perceived cognitive and situational demands. It is conceivable that test takers’ perceptions of test demands may affect their judgment and expectation of success. Similarly, perceived test uses are likely to serve as short-term goals, which are likely to affect test takers’ evaluation of test importance.

The Jacobs and Eccles model (2000), however, is ambiguous as to the exact paths of influence from goals and perceptions to the two motivation factors: Value and Expectancy. It is not clear whether the former affect both or just one of the latter. Elsewhere (e.g. Eccles et al., 1983; Wigfield, 1994), goals are discussed in association with task evaluation, and perception of task demand is often associated with expectation of success or efficacy.
beliefs. Because of the theoretical ambiguity, two competing models are specified for statistical modeling, explaining the paths from testing to preparation with slight differences.

Model 1 (Figure 2 without the two dashed lines) hypothesizes that the two perception factors influence preparation entirely via the two motivation factors – Expectancy and Value – as posited by the theory. Model 2 (Figure 2) hypothesizes the same relationships between perceptions and preparation, but two additional direct paths (the dashed lines in Figure 2) are added. Meanwhile, both Model 1 and Model 2 recognize a positive correlation between the two factors (the double arrowed line), as noted by Eccles and Wigfield (2002).

Specifically, test takers’ perceptions of test design are defined as the language skills they perceive as necessary for test taking (labeled as P-test design). Test takers’ perceptions of test uses are defined as the degree to which they endorse high-stakes, instrumental test uses as their primary motive to take the test (labeled as P-test uses). Perception of language skills is considered a useful indicator of test design because the more test takers agree that they need to use language skills in order to answer test items correctly, the more likely it is that they will endorse the test design (Xie, 2011). Perceived test uses are essentially test takers’ short-term goals in relation to CET4 test taking and the potential uses of test results. Only high-stakes, instrumental test uses were assessed and reported in the present study because they are the focus of interest in the debate regarding test washback.

The present study assessed the above conceptual models to address the following two research questions:

1. Do test takers’ perceptions of test design and high-stakes test uses influence preparation simultaneously?
2. If so, in what ways and to what extent do perceived test design and test uses influence preparation?

Figure 2. Conceptual model 1 & 2
This section has reviewed the literature relating to washback and Expectancy-value theory, thereby presenting the conceptual models and two research questions. The next section outlines the research method.

Method

General methodological considerations

This study focuses on washback on learning during a special test preparation period prior to test taking. According to Prodromou (1995), washback proceeds on a continuum from covert to overt influence as the date of the test approaches, particularly when the test-taking dates are externally determined and the stakes are high. While washback prior to the preparation period is less observable and hence ‘covert,’ during the preparation period, washback is intensive, observable, and ‘overt.’ Examples of covert washback are the impact of the test syllabus on textbooks, the impact of test results on teacher appraisal, and the impact on students’ views of achievement. Overt washback takes the form of purposeful preparation to maximize performance (e.g. teaching and learning to the test, using test materials, practicing test papers, and drilling test items). Overt washback is the main interest of this study, because it seems to be the most closely related to test design and validity concerns, over which test designers can exert control.

Methodologically, overt washback during test preparation may be the most intense form of washback accessible to observation and measurement. Given that there is little theory within language testing regarding washback mechanisms, overt and intensive washback facilitates the initial exploration of the mechanisms.

At the practical level, the high-stakes nature of the CET4 and the controversies regarding its negative washback made it an excellent case for exploring the theoretical interests of this study. The study was conducted in China with test takers preparing for CET4, a standardized, norm-referenced English proficiency test taken at the end of a two-year College English program by all non-English-major undergraduates. It is a high-stakes test because its results are widely used for university graduation and job application. In recent years, CET4 has received wide criticism for its negative washback on teaching and learning. These criticisms, however, are more often emotionally charged and politically motivated than empirically grounded. There are few empirical studies investigating CET4 washback (e.g. Gu, 2005).

Participants

Participants of this study are from a university in Guangzhou, a large city in southern China. From April to June 2009, over 800 second-year university students who registered to take CET4 took part in this study. This university was selected due to its ranking among the lower tier. The washback from CET4 on teaching and learning is more obvious and intense in lower tier universities because students at these universities tend to have lower levels of English ability and are more likely to fail the test than those at higher tier universities. Because of the high-stakes nature of CET4 and relatively large failure rates, lower tier universities tend to spend more resources on test preparation. It is
understandable that CET4 affects these universities more than higher tier ones. The university’s lower tier status made it a highly appropriate choice for observing and studying the mechanisms of washback.

In addition, students’ test preparation for CET4 in this university was homogeneous and intense. Every summer, two months before the CET4 was administered, all regular second-year English classes at the university were suspended to provide additional time for test preparation. Because students were aware of this practice in advance, the majority would not start preparation for the CET4 until this special period commenced. This pedagogic arrangement provided a logistically and ethically convenient context for the present study.

Instrumentation

Two separate questionnaires were created to survey students’ perceptions of assessment and test preparation practices. The two questionnaires were developed from a qualitative study in 2007 and were verified in a pilot study with another sample (N = 157) in 2008. Items with sound psychometric properties were selected and used in the present study. The perception questionnaire measured four major constructs: perception of test design, perception of test uses, expectation of success, and test value.

*Perception of test design* was measured by asking students about the language skills that they perceived as necessary for answering CET4 (e.g. *In order to answer questions on the short-conversation section correctly, I must grasp the gist of the talk*). In the construction of this scale, care was taken to map the items back to the official documents of the test (CET4 test syllabus). This was to ensure that all of the intended language skills appeared on the questionnaire, so that the questionnaire would fully represent these skills.

*Perception of test uses* was measured by asking students to what extent high-stakes test uses were their primary purposes for taking CET4. This scale has three items (α = .673): *I take CET4 in order to get CET4 certificate for job seeking; I take CET4 in order to get my degree; I take CET4 in order to get my graduation certificate.* These three items refer to the most prevalent, high-stakes uses of CET4 results (Jin, 2006).

*Expectation* was measured using the self-efficacy scale from the Motivated Strategy Learning Questionnaire (Pintrich, 1991) modified to fit the CET4 context. It comprises five items (α = .884): *If I prepare for it in appropriate ways, I believe I will pass CET4; If I try hard enough, I will pass CET4; I think I will do well on the test; Taking consideration of its difficulty and my ability, I think I can pass CET4; I have confidence in doing well on CET4.*

The *test value* scale asks test takers to evaluate the importance of CET4; it has three items (α = .738) incorporating attainment importance, utility value, and intrinsic value of interest: *Doing well on CET4 is very important for me; Passing CET4 is useful for my future; Taking CET4 can help me to learn the English language.*

*Test preparation* was defined in terms of students’ self-reported practices during the 10-week test preparation period. It had five scales: test analysis (i.e. assessing test and evaluating oneself); rehearsing test-taking skills; drilling target skills; memorizing; and
socio-affective strategies. Appendix A presents examples of the items used to measure perception and test preparation.

**Data collection**

Data collection started 10 weeks before the June 2009 administration of CET4 with the perception questionnaire given at the beginning of the preparation period. The test preparation questionnaire was given eight weeks later, near the end of the preparation period and about two weeks before the participants took the test.

**Data analysis**

*Data preparation.* All items were rated on Likert scales. Items in the perception questionnaire were scored on a 6-point Likert scale of agreement. Items in the Test preparation questionnaire were scored on a 5-point Likert scale of frequency. A 6-point scale was used for the perception questionnaire to avoid a possible skewness problem, which was encountered when the perception questionnaire was piloted. Because a finer scale could detect a wider variance for each item, thereby attenuating the skewness problem, a 6-point scale was used in the present study.

All items were entered into SPSS for data cleaning and preliminary analysis. Missing value analyses were conducted to check missing rates and missing patterns. Expectation Maximization (EM) algorithm was used to impute missing values at the item level (Graham, 2009). Outliers were identified through examining the stem-and-leaf plot, box-plot, and Mahalanobis Distance of each item (Tabachnick & Fidell, 2007). The skewness and kurtosis statistics of all items were examined to check normality assumptions of their distributions. Multivariate normality assumption was checked via Mardia Coefficients of Kurtosis. Mardia coefficient values lower than 3.0 were considered acceptable. After data cleaning and imputation, the data met the normality assumption of multivariate analysis and were ready to be analysed. 888 cases were kept for the perception questionnaire; 872 cases were kept for the test preparation questionnaire.

Item-level exploratory factor analysis (EFA) was conducted to find out whether the items designed to measure one construct indeed loaded together. Based on the EFA results, composite variables were computed by averaging item scores within each factor. Composite variables were used as observable indicators in Confirmatory Factor Analysis. Finally, Structural Equation Modeling (SEM) was used to assess the structural relationships. AMOS7.0 (Arbuckle, 2006) was used for all latent factor modeling.

**SEM.** SEM is a comprehensive statistical approach to testing theoretical hypotheses about the relationships among observed and latent variables (Hoyle, 1995). The purpose of structural modeling is to verify whether hypothesized relationships among variables are supported by empirical data. Usually, a model is specified a priori according to a substantive theory, common sense, or a hypothesis to be tested. SEM then estimates the discrepancy between the correlation-covariance matrix as implied by the model and the observed correlation-covariance matrix of the empirical data. The discrepancy is
measured by Chi-square statistics. The smaller the Chi-square value, the closer the data fit the model. When the fitness is satisfactory, the model is considered to be an approximate representation of the relationships among the variables in the model. It represents one plausible explanation until further evidence falsifies this explanation.

Besides Chi-square statistics, four other indices are often used to measure model fitness: Goodness-of-fit Index (GFI; Jöreskog & Sörbom, 1981), Comparative Fit Index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), and Root Mean Square Error of Approximation (RMSEA; Steiger, 1990). For GFI, CFI, and TLI, values above .90 are considered a good model fit. RMSEA values of .05 or lower are considered to represent a close fit, while values up to .08 represent a reasonable fit.

**Mediation analysis.** Mediation analysis is one way to explain a chain of relations from a predictor to a dependent variable, or the mechanism by which one factor exerts influences on another. Mediation exists when an antecedent variable affects an outcome variable indirectly through at least one intervening variable, or mediator (MacKinnon, Fairchild, & Fritz, 2007). Mediation analysis is prominent in testing theories regarding process mechanisms. Mediation models may involve one mediator or multiple mediators. The mediation models tested in the present study involve two mediators (i.e. Expectancy and Value) used to explain the paths of influences from test-taker perceptions to test preparation.

For a mediation relation between X and Y where M is the mediator, the standard notation uses a to represent the raw correlation between X and M, b to represent the partial correlation of M to Y adjusted for X, and c’ the relation of X to Y adjusted for M. This study uses four common measures to suggest the overall magnitude of mediation effects. The first is to examine whether the direct path c’ is significantly different from zero (Baron & Kenny, 1986). When the path c’ is not significantly different from zero, it is considered complete or perfect mediation. Otherwise, it is considered partial mediation. Partial mediation indicates the potential to identify additional mediators. The other three commonly used measures of mediation effects are ab, ab/(ab+c’) (Alwin & Hauser, 1975), and βa*βb (Preacher & Kelley, 2011). While ab represents unstandardized measure of indirect effect, βa*βb represents standardized indirect effect. ab/(ab+c’) is also called mediation ratio, which suggests the proportion of mediated effects to the total effects.

**Results**

In this section, the results of the data analysis are presented. Because of space limitations, item-level EFA results are not presented; instead, descriptive statistics of subscales and correlation matrix are provided in Appendix B. Interested readers can request further details via the corresponding author’s email.

**Measurement model**

Confirmatory factor analysis suggests that the two questionnaires have achieved satisfactory reliability and validity. The full measurement model is schematized in Figure 3.
Model fit indices for this measurement model suggest that it fits the data satisfactorily ($\chi^2 = 260.46$, $df = 57$, RMSEA = .054, CFI = .958, NFI = .947, TLI = .933). Given a large sample size and a relatively large number of observed indicators ($N = 12$), the Chi-square statistic is relatively small. RMSEA is .054, and the three comparative model fit indices are all above .90.

This measurement model includes five major scales: perception of test design (p-test design), perception of test uses (p-test uses), expectation, test value, and test preparation (Test prep). The p-test design scale has four subscales: perceived test demand for linguistic knowledge, perceived test demand for reading skills, perceived test demand for listening skills, and perceived test demand for writing skills. In this measurement model, the latent factor of p-test design represents the degree of test takers’ positive endorsement of test design. The higher the endorsement of language skills as necessary for test taking, the more the test takers perceived test design as positive. Endorsement of test demand is also indicative of test takers’ knowledge of the test.

The test preparation scale includes five sub-scales representing five types of test preparation practices: using socio-affective strategies to seek external help and to motivate oneself, conducting test analysis via assessing test requirements and evaluating oneself, memorizing vocabulary and model essays, drilling target language skills, and rehearsing test-taking skills. Sample items of the subscales are provided in Appendix A. The latent factor of Test Prep represents the degree of engagement in preparation indicated by coherent uses of various strategies.

**Hypothesis testing results**

Both hypothesized models fit the data satisfactorily. RMSEA is lower than .05, NFI, CFI, and TLI are above .95. Compared with Model1 ($\chi^2 = 189.50$, $df = 49$, RMSEA = .049,
CFI = .969, NFI = .959, TLI = .950), Model 2 (with direct paths) fits the data even better ($\chi^2 = 155.44$, $df = 47$, RMSEA = .044, CFI = .976, NFI = .966, TLI = .960). Because these two models are nested within each other, the Chi-square difference test is appropriate to test which model fits the data better. The result of the Chi-square difference test suggests Model 2 is better than Model 1 ($\Delta \chi^2 = 34.06$, $df = 2$, $p < .001$), which further suggests that adding two direct effects from Perception variables to Test preparation significantly improved model fitness. That is, the two mediators can explain only part, but not all, of the influence from Perceptions to Test preparation. Specifically, there is a significant direct effect ($\beta = .26$, $p < .001$) from P-test design to Test Prep, which cannot be explained by the two mediators. Figure 4 presents Model 2 with standardized parameters.

Of the four mediation paths tested in Model 2, three paths are significant (see Table 1). Because the path from P-test use to Expectation is not significantly different from zero ($\beta = .055$, n.s.), Expectation is not considered to be a useful mediator of the effects of P-test use on Test Prep. Value is considered an important mediator between P-test use and Test Prep. The direct path is not significantly different from zero ($c_1' = -.034$, n.s.), suggesting a complete mediation. That said, both the unstandardized and the standardized indirect effect measures ($ab = .034$, $\beta_a \times \beta_b = .044$) suggest that the total mediated effect is small. Moreover, both Expectation and Value are useful mediators for the effects from P-design to TP, though the effects are only partial ($c_2' = .326$, $p < .001$). Partial mediation suggests there may be other additional mediators between perception of test design and test preparation. The three effect size measures suggest that both the ratios of mediated effects and the magnitude of the effect mediated are small. Compared with the value mediator, the expectation factor explains more effects from P-design to TP.

Figure 4. Model 2 with standardized path coefficients and R-squares
Table 1. Effect size of mediation effects

<table>
<thead>
<tr>
<th>Paths</th>
<th>Descriptors</th>
<th>ab</th>
<th>ab/(ab + c')</th>
<th>$\beta_1 \cdot \beta_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-use→Value→TP</td>
<td>Complete</td>
<td>.034</td>
<td>–</td>
<td>.044</td>
</tr>
<tr>
<td>(a_1 \cdot b_1)</td>
<td>(c_1' = −.03, n.s.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-use→Expectation→TP</td>
<td>No mediation</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(a_2 \cdot b_2)</td>
<td>(a_2 = .055, n.s.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-design→Expectation→TP</td>
<td>Partial</td>
<td>.060</td>
<td>.154</td>
<td>.088</td>
</tr>
<tr>
<td>(a_3 \cdot b_2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-design→Value→TP</td>
<td>Partial</td>
<td>.030</td>
<td>.085</td>
<td>.034</td>
</tr>
<tr>
<td>(a_4 \cdot b_1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Post-hoc analysis: Differential effects of perceptions on test preparation practices

Finally, a post-hoc analysis of specific effects was conducted to find out whether perceived test design and endorsement of high-stakes test uses influenced the types of preparation that test takers chose to engage in. All direct paths from P-test design and P-test uses to the five indicators of test preparation were assessed in turn. Assessments of these specific effects were post hoc and data driven.

Three significant direct paths were identified. For the sake of model parsimony, only these three were kept in the final solution ($\chi^2 = 122.06, \text{df} = 52, \text{RMSEA} = .033, \text{CFI} = .986, \text{NFI} = .975, \text{TLI} = .975$). This round of post-hoc assessments found that test takers with higher awareness of necessary skills tended to use more meta-cognitive strategies in preparation, that is, higher usage of assess & evaluate strategies ($\beta = .12, p < .001$). They also rehearsed test-taking skills more frequently ($\beta = .19, p < .001$). In addition, test takers who endorsed high-stakes test uses tended to use socio-affective strategies less frequently during test preparation ($\beta = −.07, p = .018$).

Discussion

Research question 1 asks: ‘Do test taker perceptions of test design and high-stakes uses influence preparation simultaneously?’ The findings of this study indicate that both perceptions of test design and high-stakes test uses influenced preparation simultaneously. Compared with perceptions of high-stakes test uses, perceptions of test design seemed to exert relatively greater impact on overall test preparation (Standardized total effect: .386 vs. .003). This finding is consistent with previous studies that conclude instrumental goals exert relatively little effect on strategy use (Pintrich, 1999). However, this finding should not be used to suggest that high-stakes test uses exert little impact on preparation. Because this study measures preparation in terms of the usages of various preparation strategies, it does not measure other aspects of test preparation. For instance, in allocating time and effort, test takers strategically prioritize test-related tasks and materials. This is an important aspect of test preparation, but was not included in the test preparation construct in this study. Prioritizing test-related tasks during test preparation can be
viewed as a coping strategy, which might be associated with test anxiety and the stakes attached to test uses. Further studies should include more comprehensive aspects of test preparation in the dependent variable.

Research question 2 asks: ‘In what ways and to what extent do perceived test design and test use influence preparation?’ Two mediators (Expectation, Test-value) were introduced to explain the path(s) of influences from testing to preparation. This round of modeling found these two mediators to be useful and important in understanding the mechanisms of the washback process.

**Paths of influence from perceived test uses to preparation**

In this study, evaluation of test importance (Test-value) explains the influence from perception of test uses to preparation (a complete mediation). Meanwhile, no significant, direct relation was found between endorsement of instrumental test uses and expectation of success. This finding suggests that the endorsed instrumental uses affected preparation primarily via the value aspect of motivation. Test takers who endorsed instrumental test uses as short-term goals assigned more importance to test taking; the value attached to test taking motivated them to engage more in preparation via usage of multiple preparation strategies. That said, the total effect of endorsed instrumental test uses on preparation is slim (standardized total effect = .003). A negative direct path from P-test uses to Test Preparation is observed. However, the path coefficient is not significant statistically, and we do not consider it meaningful. Nevertheless, similar negative path coefficients have been observed elsewhere. For example, Pintrich (1999) reports standardized regression coefficients from extrinsic goals to use of cognitive strategies (from −.03 to .11) and to self-regulation (−.03 to .06). However, he does not mention whether the negative values are statistically significant or meaningful.

Furthermore, this study also found a negative relation between the endorsement of instrumental test uses and the use of socio-affective strategies for seeking help. Test takers who endorsed instrumental test uses reported using socio-affective strategies less frequently during test preparation (β = −.07, p = .018). Karabenick (2004) observed that perception of a competitive classroom environment discouraged students from seeking achievement-related help. High-stakes test uses might intensify competitive classroom environments, and thus be inversely related to the usage of socio-affective strategies for seeking help.

**Path of influence from perceived test design to preparation**

This study found both Expectation and Test-value are useful mediators to explain the effects from perception of test design to preparation. For the mediation effects of expectation, this study found that test takers who endorsed language skills as necessary tended to show more confidence towards test taking; higher self-efficacy, in turn, is associated with greater engagement in preparation. Bandura (1982) noted that self-appraisal of various information is a necessary process in the formation of positive self-efficacy. Gist and Mitchell (1992) noted that perceived task requirements as one determinant of self-efficacy were associated with reduced fear and uncertainty. Perhaps knowledge of test
demands helps to dispel fear towards test taking. Though test takers might not yet possess the language skills and knowledge perceived as necessary, knowing what is necessary could give them a sense of direction and control over the upcoming task. As they learn more about what to prepare for, the locus of control returns to their hands rather than being the possession of a mysterious, unknown body. Thus, they appear more confident towards test taking.

Higher self-efficacy was associated with more engagement in preparation. In connection with Gosa’s (2004) finding that high self-efficacy in conjunction with perception of an easy test led to little test preparation, it suggests that test takers tended to perceive CET4 as challenging but not easy. Therefore, high self-efficacy positively contributed to greater engagement in preparation. Bandura (1982) noted that some self-doubt may actually activate learning and preparatory behaviour, while supreme self-efficacy in conjunction with perception of easy tasks may discourage them. Contingent with perceived task difficulty or easiness, self-efficacy may exercise positive or negative effects on learning effort. That said, existing studies observe more positive relations from self-efficacy to effort and usage of strategies than negative ones (e.g. Pajares, 1996; Pintrich, 1999).

Pintrich (1999, p. 463) summarized many studies and generated a range of standardized estimations between self-efficacy and cognitive strategy use (from .009 to .36) and between self-efficacy and self-regulation (from .12 to .58). The lower end of the ranges represents standardized regression coefficients, while the higher end represents zero-order correlations. In this study, the path coefficient from self-efficacy to engagement in preparation (β = .26, p < .001) is located in the middle. This relatively stronger relationship is probably because self-efficacy was measured with careful specification of test taking and preparation in this study. Because the judgment of self-efficacy is task and domain specific, specification in self-efficacy assessment and task correspondence may have strengthened the predictive accuracy of self-efficacy (Pajares, 1996).

Regarding the mediation effects of test value between perception of test design and preparation, this study found positive endorsement of intended test demand was related to higher evaluation of test importance (β = .148, p < .001). Eccles et al. (1983) noted that the value attached to academic tasks is dependent on both personal and task characteristics. When the task is perceived as instrumental for achieving personal goals, it is associated with higher utility value. On the other hand, when the task is perceived as interesting and meaningful, it is associated with intrinsic value. In this study, those test takers who endorsed test design positively might have also considered the preparation process as a useful and meaningful learning experience related to the long-term goal of developing language ability. In fact, one item of the test value scale states: ‘Taking CET4 can help me to learn English language.’ The usefulness of the preparatory process for achieving long-term goals may attract additional value to test taking.

Consistent with previous studies (e.g. Pintrich 1999), this study found higher task value is associated with greater engagement in preparation (β = .14, p < .001). For the relationships between task value and strategy use, a range between .03 and .67 is identified (Pintrich, 1999) with the lower end of the range representing standardized regression weights (the unique contribution of task value) and the higher end representing zero-order correlations. The path coefficient observed in this study indicates the unique
contribution of test value to preparation; it is located towards the lower end of the range, indicating a small but significant effect.

Furthermore, test taker perceptions of test design also affect usage of preparation strategies. This study found that test takers who endorsed language skills as necessary for test taking tended to conduct more test analysis ($\beta = .12, p < .001$) and practice more test-taking skills ($\beta = .19, p < .001$). According to L2 strategy researchers, task knowledge is one type of meta-cognitive knowledge, which provides the knowledge base for effective planning, monitoring and evaluating. It ‘prompts learners to initiate a task analysis to realize what it needs to be done. It also dictates what must be done to complete the task’ (Wenden, 1998, p. 524). Test analysis (assessing the test and evaluating self) may be affected by knowledge of test demand. Test takers might use their knowledge of the test to manage the preparation and to identify appropriate test-taking skills for test taking.

**Implications for practice**

For test developers and users of test results, in this case, the university registrar who will check whether graduation requirements have been met or potential employers who may require the CET as an employment requirement, the findings of this study indicate that both test design and use may affect preparation for the test simultaneously, though their paths of influence may be different. Overall, endorsement of intended language skills as a test demand contributed to engagement in test preparation (.36, medium effect); endorsement of high-stakes test uses only marginally contributed to test preparation (.003, marginal effect). Perceptions of test design and perceptions of test uses explained 24% variances of the test preparation factor.

Although this study shows that perceived test uses have less impact on overall preparation than does perceived test design, this finding is by no means conclusive. This finding is partly related to the way test preparation was measured. Further studies need to incorporate other aspects of preparation (e.g. strategic task prioritizing) into the measure to generate a more comprehensive view of the impact of high-stakes test uses.

For educational reformers using high-stakes testing as a lever of change, the findings of this study might be of special interest. The study provides one tentative answer to the question: ‘Why do measurement-driven interventions aiming at producing positive washback often miss their target?’ More often than not, intended washback is not realized; that is, what is learnt does not correspond to what was intended. If nothing else, the chain of mediators between what is assessed and what test takers prepare for reveals the complexity of the process from testing to learning. First of all, intended skills need to be endorsed by test takers as necessary. Positive endorsement of intended skills is necessary, but not sufficient. Test takers also need to consider the test valuable and doable (i.e. ‘I want to do it’ and ‘I can do it’). Perception of test demand acts as the first filter for the intended skills to get through. Expectation and Value act as the second filter. The skills intended by test designers may not be endorsed by test takers; the endorsed skills may not be prepared and learnt if test takers consider them unimportant or unattainable. Because these filters...
operate along the path from testing to learning, what passes through them may be considerably less than what is intended. Perhaps only the aspects of intended skills that are perceived by test takers as necessary, important, and manageable are prepared and hopefully learnt. This observation is consistent with Qi’s study (2005), which found that teachers refused to endorse the intended communicative language skills as necessary for test taking. The mismatch between the intended and endorsed test demands undermined the test designers’ intentions to promote innovation in teaching.

**Limitations and future directions**

Although the results of this study add to our understanding of the paths from testing to preparatory behaviours, there are several limitations that must be addressed in future studies. One is to examine test preparation as a special instance of washback on learning. For any high-stakes external testing context, when the test time approaches, the most intensive washback is likely to be observed. Overt washback of a high-stakes examination is nevertheless a special instance. It is likely to exhibit different characteristics from covert washback on learning when test taking is not immediate. As the purpose of this study was to explore the mechanisms of washback, confining the study of washback to a specific period not only made the study logistically feasible, but also made methodological sense. Test preparation is the time when testing exerts the most direct and visible impact on learning. Confining washback on learning within the test preparation period enhanced the internal validity of this study, though, at the same time, it reduced its generalizability to other contexts. Given this limitation, a natural extension would be to investigate ‘covert washback’ during pre-preparation time periods.

Another limitation is related to the adoption of Expectancy-value motivation theory. By adopting this theory, the present study assumes test preparation consists of planned, motivated, rational, and achievement-related behaviours, and excludes other possibilities (e.g. habitual, irrational, or emotional behaviours). Although test preparation is more likely to be planned and rational, other influences may also affect it. Such influences, however, were not investigated in this study.

Moreover, although preparatory behaviours can be studied analogously to learning and situated within a framework of achievement motivation, test preparation within a high-stakes context is different from more general learning in that it is often intrinsically associated with human reactions to pressure, fear, and anxiety. It would therefore be beneficial for future studies to investigate test preparation within the framework of coping mechanisms. Specifically, affective aspects of motivation such as test anxiety could be introduced into the model as another useful mediator for understanding the paths of influence of high-stakes test uses on the strategies employed in preparation and test taking.

Introducing Expectancy-value theory into washback research represents an important step toward understand washback on learning. For this purpose, it was considered prudent to simplify this phenomenon conceptually and methodologically, especially at this initial stage. In terms of methodology, the present study used self-report
questionnaires to capture test takers’ perceptions and preparation behaviours. This may be considered another limitation. In future studies, it will be desirable to triangulate data from other sources as well.

This study serves to demonstrate the feasibility of borrowing an established theory from other areas to explore washback mechanisms, and the feasibility of employing SEM methodology to assess the conceptual model statistically. A larger model with more factors, although beyond a single researcher’s capacity and logistical affordance, is conceivable for future researchers. A logical extension of this study is to include more contextual and social factors into the model as informed by the full model of Expectancy-value theory. Teachers’ perceptions of test design and uses, test takers’ interpretations of past experience, and their test performance could be included in the extended model. Additionally, a similar model for the low-stakes context could be specified and tested, and then compared with that observed in a high-stakes context.

Finally, this study provides empirical evidence that both test design and high-stakes test uses can affect test takers, albeit at different magnitudes and following different paths. Utilizing Expectancy-value motivation theory, we have provided potentially useful insights to understand the complexity of washback on learning. However, more empirical studies are necessary in order to obtain a more comprehensive picture of this phenomenon.

Acknowledgements

This paper is based on the first author’s PhD thesis, for which the second author was primary supervisor. Both authors would like to thank the co-supervisor, Raymond Lam, for his contributions to the study. We would also like to acknowledge that this study received partial support from Educational Testing Service under Small Grants for Doctoral Research in Second or Foreign Language Assessment.

Note

1. Upon successful completion of their Bachelor degree programmes, university graduates will receive both a degree certificate and a graduation certificate.

References


### Table A1. Sample items for perception and test preparation

<table>
<thead>
<tr>
<th>Subscales</th>
<th>Example items</th>
</tr>
</thead>
<tbody>
<tr>
<td>In order to answer questions...correctly, I must...</td>
<td></td>
</tr>
<tr>
<td>1. P-linguistic knowledge</td>
<td>...have enough knowledge of phrases and collocation</td>
</tr>
<tr>
<td>2. P-reading skills</td>
<td>...understand the key points in reading</td>
</tr>
<tr>
<td>3. P-listening skills</td>
<td>...make inference about the speakers’ attitudes and views</td>
</tr>
<tr>
<td>4. P-writing skills</td>
<td>...write in smooth and fluent English</td>
</tr>
<tr>
<td>5. Socio-affective strategies</td>
<td>...tried to learn from others;</td>
</tr>
<tr>
<td>6. Test analysis (assess &amp; evaluate)</td>
<td>...analyzed question types to identify frequently assessed questions;</td>
</tr>
<tr>
<td>7. Memorizing</td>
<td>...recited sentence patterns for improving performance on writing;</td>
</tr>
<tr>
<td>8. Drilling</td>
<td>...drilled on my reading comprehension skills;</td>
</tr>
<tr>
<td>9. Rehearse test-taking skills</td>
<td>...trained my skills to choose options through logic elimination;</td>
</tr>
</tbody>
</table>

For CET4 essay, I...
### Appendix B

#### Table A2. Scale descriptive statistics and zero-order correlation matrix (EM method)

<table>
<thead>
<tr>
<th>Scales</th>
<th>subscales</th>
<th>N</th>
<th>Items</th>
<th>α</th>
<th>Mean</th>
<th>SD</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-test uses</td>
<td></td>
<td>888</td>
<td>3</td>
<td>.673</td>
<td>4.878</td>
<td>.882</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test value</td>
<td></td>
<td>888</td>
<td>3</td>
<td>.738</td>
<td>4.526</td>
<td>.897</td>
<td>.391</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-test design</td>
<td>P-linguistic knowledge</td>
<td>888</td>
<td>8</td>
<td>.841</td>
<td>4.748</td>
<td>.621</td>
<td>.280</td>
<td>.296</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-reading skills</td>
<td>888</td>
<td>4</td>
<td>.784</td>
<td>4.955</td>
<td>.615</td>
<td>.253</td>
<td>.250</td>
<td>.583</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-listening skills</td>
<td>888</td>
<td>7</td>
<td>.809</td>
<td>4.603</td>
<td>.597</td>
<td>.296</td>
<td>.274</td>
<td>.666</td>
<td>.594</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-writing skills</td>
<td>888</td>
<td>6</td>
<td>.852</td>
<td>5.046</td>
<td>.628</td>
<td>.237</td>
<td>.274</td>
<td>.679</td>
<td>.528</td>
<td>.599</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectation</td>
<td></td>
<td>888</td>
<td>5</td>
<td>.884</td>
<td>4.466</td>
<td>.851</td>
<td>.175</td>
<td>.375</td>
<td>.321</td>
<td>.203</td>
<td>.295</td>
<td>.305</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test prep</td>
<td>Socio-affective</td>
<td>872</td>
<td>6</td>
<td>.836</td>
<td>1.601</td>
<td>.655</td>
<td>.016</td>
<td>.246</td>
<td>.203</td>
<td>.106</td>
<td>.137</td>
<td>.198</td>
<td>.272</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memorizing</td>
<td>872</td>
<td>5</td>
<td>.835</td>
<td>2.785</td>
<td>.881</td>
<td>.070</td>
<td>.175</td>
<td>.156</td>
<td>.139</td>
<td>.131</td>
<td>.193</td>
<td>.239</td>
<td>.513</td>
<td>.543</td>
<td>/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drilling</td>
<td>872</td>
<td>4</td>
<td>.782</td>
<td>2.846</td>
<td>.796</td>
<td>.106</td>
<td>.214</td>
<td>.201</td>
<td>.181</td>
<td>.194</td>
<td>.225</td>
<td>.249</td>
<td>.504</td>
<td>.578</td>
<td>.565</td>
<td>/</td>
</tr>
</tbody>
</table>