<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Properties of preschool Chinese: implications for learning to read</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other</strong></td>
<td><strong>Contributor(s)</strong></td>
</tr>
<tr>
<td></td>
<td>University of Hong Kong; 陳倩嬌</td>
</tr>
<tr>
<td><strong>Author(s)</strong></td>
<td>Chan, Sin-kiu</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2003</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/56193">http://hdl.handle.net/10722/56193</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; The author retains all proprietary rights, such as patent rights and the right to use in future works.</td>
</tr>
</tbody>
</table>
Properties of Preschool Chinese: Implications for Learning to Read

CHAN, Sin Kiu

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, May 2, 2003.
ABSTRACT

Chinese characters extracted from teaching materials from six sources were entered into a database for linguistic analysis. Teaching materials are standard teaching materials that published by three different publishers, three self-compiled teaching materials compiled preschools of different socioeconomic status. Each Chinese character was coded according to their features in the database. Characteristics of preschool Chinese characters suggest a slow development of explicit representations on how phonetics and signifies symbolize phonological and semantic information.
INTRODUCTION

The Chinese writing system has often been described as a logographic system. This is because Chinese words are represented by distinct symbols (Shu & Anderson, 2000). These symbols, the basic units of written Chinese, are called characters. It has been presumed that children learn Chinese characters by rote memory (e.g., Tunmer and Hoover, 1992). If characters are learnt only by rote memory, two problems will arise (Gough, Juels, & Griffith, 1992). Firstly, visually similar characters will become difficult to memorize and differentiate as the number of characters increase. Secondly, if characters are all independent individuals with arbitrary configurations and phonological correspondence, new characters cannot be read. Can Chinese characters only be learnt by rote memory? To understand whether this is the case, we should first look at the characteristics of Chinese characters.

Chinese Characters

Chinese characters can be divided into two broad categories: simple and complex (Li, 1993). Simple characters are of holistic visual units that cannot be further divided into meaningful sublexical units (e.g., 田 /tin4/ “field” and 水 /soey2/ “water”). Complex characters are composed of sub-character units (e.g., 焚 /fan4/ “to burn”, with sub-character unit 林 /lam4/ “woods” and 火 /fo2/ “fire”), where sub-character units can be free (independent characters in their own rights) (e.g., 林 and 火 in 焚 /fan4/ “to burn”) or bound (orthographic units that cannot exist on their own) (e.g., 且 and 焰 in 捷/jiu4/ “far”).

According to Li (1993), 95% of all Modern Chinese characters are complex characters and about 94% of the complex characters are phonetic compounds. Each phonetic compound is composed of two components, a semantic radical (significs thereafter) and a phonetic radical (phonetics thereafter). Significs carry semantic information (e.g., significs 火 /fo2/ meaning “fire” in 焚 /jin4/ meaning “light up”) while phonetics carry phonetic information (e.g., phonetics 然 /jin4/ in 焚 /jin4/). As phonetic compounds contribute to a large proportion of Chinese characters, strategies that can help to memorize phonetic compounds would relieve the load of learning Chinese characters. Hoosain (1991) has reported that
approximately 7000 modern Chinese characters are mainly composed of 200 significs and 800 phonetics. It can be inferred that if children can memorize characters by chunking them into their phonetics and significs, instead of learning 7000 characters, learning only 1000 sub-character components will be sufficient. The memory load can then be lessened.

**Can Chinese Characters Only be Learnt by Rote Memory?**

Simple characters, which constitute 5% of total modern Chinese characters, can only be learnt by rote memory. This is because the relationship of the characters and their meaning as well as their pronunciation is arbitrary in modern Chinese. On the contrary, phonetic compounds can be learnt by analogy. For example, when a child encounters a novel character 鰍 (/lei5/, “carp”), if the child can recognize the phonetic (i.e., 里 /lei5/) and he knows the pronunciation of it, the child can then pronounce the novel phonetic compound 鰍 (/lei5/) as 里 (/lei5/). If the child can recognize which is the signific (i.e. 魚, “fish”) and he knows the meaning of it, the child can then infer the meaning of the character that 鰍 (“carp”) is a fish.

Learning Chinese characters is not only by rote memory. Phonetic compounds can be learnt by inferring pronunciation and meaning from their phonetics and significs.

Studies show that children do make use of phonetics and significs when reading Chinese characters. In Ho and Bryant (1997a)'s study, children pronounced unfamiliar Chinese characters as the pronunciation of their phonetics. This shows that children can recognize the phonetics and that it provides phonetic information. In the study conducted by Shu and Anderson (1997), children could choose the target character with appropriate significs when the meaning of the target character is given. This shows that children know significs provide semantic information.

**Explicit Representation Aids Learning to Read**

As mentioned, to make inferences on meaning and pronunciation of phonetic compounds, children have to first recognize phonetic compounds, then recognize the corresponding phonetics and significs. They finally have to be consciously aware of the role of phonetics and
significs and apply the awareness when making inferences. This conscious awareness of the existence of sub-character components and how each of these components symbolize semantic and phonetic information (explicit representation thereafter) aid their learning to read.

Ho and Bryant (1997a) and Shu and Anderson (1997)’s studies show a positive relationship between explicit representations and reading proficiency. In Ho and Bryant (1997a) and Shu and Anderson (1997)’s studies, children who had better performance when doing a task which requires the application of conscious phonological and semantic awareness, had better reading proficiency. The development of the explicit representations would allow one to make inferences on the pronunciation and the meaning of novel characters. It can be deduced that the direction of relationship is: developed explicit representations causes a better reading proficiency. Development of explicit representation aids learning to read.

These explicit representations are not innate. They develop in normal children who learn Chinese characters. Though the function of sub-character components is not taught explicitly, normal children can derive and develop explicit representation by exposure to various characters (Ho & Bryant, 1997; Shu & Anderson, 1997). Development of explicit representations would be governed by the orthography of a language. For example, explicit representations of letters representing phonemes can only be developed in an alphabetic writing system such as English. This is because English orthography encodes phonemes in letters. However, Chinese orthography does not have a letter to phoneme correspondence. This explicit representation on phoneme can hardly be developed among children who learn Chinese only (Huang & Hanley, 1994). It can be deduced that characteristics of Chinese characters that children are exposed to would affect children’s development of explicit representations.

Shu, Anderson and Wu (2000) suggested that explicit representations would be progressively refined with an increasing exposure to various characters. For example, a child
may first develop the representation that phonetics are on the right and they give the pronunciation of the characters. When the child is exposed to characters with phonetics in other positions, his representations refined that sub-character components that provide information on pronunciation (the phonetics) is always on the right. The development of explicit representations will then be determined by the characteristics of characters that children are exposed to. Thus, studying the characteristics of characters that children are exposed to, at different ages, allows us to have an insight on the possible development of explicit representations and how children learn to read Chinese.

**Characteristics of Chinese characters and the development of explicit representation**

As phonetic compounds constitute about 90% of modern Chinese characters, understanding how children learn to read these characters is important. To understand the possible reading development in children, an investigation on the characteristics of Chinese characters that children are exposed to at different ages should be done. A study on the Chinese orthography of primary schools in Beijing on simplified characters (Shu, Chen, Anderson, Wu & Xuan, 2000) was done to investigate the possible acquisition of school characters. However, a study on the characteristics of preschool Chinese characters is missing. Studies have shown that explicit representations should have begun to develop when children are in their preschool level (Ho & Bryant, 1997a). Ho and Bryant (1997a) found that children in primary one read unfamiliar phonetic compounds by producing the pronunciation of their phonetics. It shows that children at preschool level should have developed the explicit representation, that phonetics gives pronunciation of its phonetic compounds. To gain a better understanding of how children learn to read, the characteristics of preschool characters is thus investigated in the present study. As most children of age three to six in Hong Kong study in preschool, children will be exposed to the characters in preschool teaching materials. In this study, the characters of preschool teaching materials were then investigated. It was of particular interest to investigate the characteristics listed below:
Proportion of simple characters, phonetic compounds and other compound characters across grades, and the number of simple figures can be phonetics or significs across grades.

Phonetics and significs of phonetic compounds can either be free or bound. Free phonetics and significs might themselves be simple characters on their own. When children are exposed to more simple characters that act as sub-character units in compound characters, they can notice meaningful sub-character units and then be able to distinguish whether a character is a compound figure. Continuous exposure to the phonetic compounds, and knowing the meaning and pronunciation of free sub-character units, children may figure out that some sub-character units usually appear as phonetics and some usually appear as significs. A large number of simple characters that act as phonetics or significs in preschool favors identification of phonetic compounds and development of explicit representation.

Proportion of simple characters, phonetic compounds and other compound characters across different grades can give an insight on whether children can only use rote memory when recognizing characters and the relative importance of the development of explicit representation. A high proportion of phonetic compounds suggest a high relative importance of the development of explicit representation.

Proportion of regular, semi-regular and irregular characters across grades. Phonetic regularity measures the degree of sound correspondence between the phonetic compound and its free phonetics.

In high regularity phonetic compounds, the pronunciation of the phonetics is exactly the same as its whole character (e.g. 清/tsing1/ and phonetics 青/tsing1/). Children can easily read the phonetic compound by reading its phonetics. When various regular phonetic compounds are encountered, children can easily recognize the relationship between phonetics and its whole characters. A high proportion of regular characters would be favorable to the development of explicit representation, that phonetics provides phonetic information. However, in low regularity characters, the relationship between the phonetic information
provided by phonetics and the pronunciation of the whole characters is not direct (e.g., 猜 /tsaa1/ and phonetics 青/tsing1/). If the proportion of low regularity characters is high, it is likely that the children will need to be exposed to a large number of characters to derive the relation that phonetics carry phonetic information.

Proportion of high, mid and low consistency characters across grades. Irregularities exist between the phonological information provided by phonetics and the pronunciation of the whole character in some characters. Besides the measure of regularity mentioned in the previous section, consistency can also reflect how the phonetics gives phonetic information.

Phonetic consistency measures the degree of sound correspondence among phonetic compounds of the same phonetic family. A phonetic family is a family of phonetic compounds that share the same phonetics, including free or bound phonetics (e.g., for free phonetics, 精 /dzing1/ “essence”, 清 /tsing1/ “clear”, 猜 /tsaa1/ “guess” share the same phonetics 青/tsing1/, they are in the same phonetic family. For bound phonetics, 離 /lei4/ “leave”, 潢 /lei4/ “wet”, 璃 /lei4/ “glass” share the same phonetics 离 which is not an independent character, they are in the same phonetic family). The more the number of different pronunciations a phonetic family has, the less consistent the phonetic family is. Considering the example of characters with phonetics 青/tsing1/ and 離, phonetic compounds with phonetics 离 are consistent as they have the same pronunciation /lei4/. For characters with phonetic 青/tsing1/, there are three possible pronunciations of phonetic compounds. The phonetic consistency for phonetic compounds having phonetic 青 is low.

Phonetic consistency may change when one learns a new phonetic compound with different pronunciation as those previously learnt in a phonetic family. For example, children may first learn the characters 青/tsing1/ and 清/tsing1/. The phonetic consistency is high when children are only exposed to these two characters of its phonetic family. However, when character 猜/tsaa1/ is learnt, two different pronunciations exist in the same phonetic family. The phonetic consistency is then lowered. Thus, phonetic consistency changes when one
learns more phonetic compounds. A high proportion of high phonetic consistency characters would be favorable to the explicit phonetic representation development. Children can easily derive the relationship between phonetics and phonological information given in characters with high consistency.

Proportion of different transparency of phonetic compounds across grades. Similarly, an irregular relationship between the semantic information provided by significs and the meaning of the whole characters is also found.

Semantic transparency measures the degree of correspondence of meaning between the characters and its significs. The more regular the relationship between semantic information the significs provides and that of the character, the higher is the semantic transparency. For example, in the character 爸 (/ba1/, “father”), the semantic 父 (father) has the same meaning as its whole character. The semantic transparency of the phonetic compounds is high. However, in the character 蝦 (/ha1/, “prawn”), the significs 虾 (insect) is not related to the meaning of its whole character. The semantic transparency of 蝦 is low. A high proportion of characters with high semantic transparency suggests a favorable development of explicit representation, that significs carry semantic information. A low proportion of that implied that children will have to be exposed to a large number of phonetic compounds to derive this explicit semantic representations.

Position of phonetics and significs of phonetic compounds across grades. If the phonetics or significs are usually found at a particular position, children can recognize phonetics or significs easily by their positions. If there are no particular positions for phonetics and significs to be in, children can hardly make use of positional information in recognition of phonetics and significs.

Differences between Characteristics of Characters in Different Teaching Materials

There are 1036 preschools in Hong Kong. Two types of preschools exist in Hong
Kong, they are namely, kindergarten and nursery. Of those 1036 preschools, there are 788 kindergartens (Education Department, 2001) and 258 nurseries (Social Welfare Department, 2002). Preschools need to follow the “Guide to the Pre-primary Curriculum” (Curriculum Development Council, 1996) to design curriculum and teaching materials. Two broad categories of teaching materials are used in these preschools: standard teaching materials and self-compiled teaching materials. Standard teaching materials are sets of teaching materials designed by different publishers for different pre-school grades. Self-compiled teaching materials are teaching materials that are designed by individual preschool. Different categories of teaching materials may allow different explicit representation in preschools.

Children will be promoted to primary one after finishing their preschool years. Different preschools use different teaching materials which in turn affect the acquisition of different explicit representations. The standard teaching materials provided by different publishers and the self-complied teaching material may differ in their choice of characters for different preschool grade.

Selection of self-compiled teaching materials may be affected by the socioeconomic status of the district that the preschool belongs to. It is of particular interest to investigate the relationship of characteristics of characters chosen and the socioeconomic status of the preschool district.

Research Objectives

The primary aim of the present study is to investigate the change in characteristics of characters, taught in different pre-school grades which provides information on the development of explicit representation during preschool and how well children are prepared for primary school. The present study also aims to investigate the differences in the characteristics of characters between different standard teaching sets and self-compiled teaching sets from districts of different socioeconomic status.
METHOD

Sampling

The standard teaching materials chosen were the most popular three among all the preschools in Hong Kong. Telephone interviews were done with 50% of the preschools, sampled randomly from the full list of preschools in Hong Kong (1036 preschools, including 788 kindergartens and 258 nursery), to investigate i) the percentage of preschools using standard teaching materials versus self-compiled teaching materials, and ii) the percentage of preschools using different sets of standard teaching materials. Two hundreds and eighty three preschools were successfully interviewed. The response rate of the telephone interviews was 54%. Eighty-nine percent of the preschools interviewed were using standard teaching materials while 10% preschools compiled their own teaching materials. Among the preschools using standard teaching materials, the materials were published by five publishers. They were standard teaching materials of the Educational Publishing House (Education), the Crystal Education Publications (Crystal), Oxford University Press (Oxford), Modern Educational Research Society (Modern) and the Longman Hong Kong Education (Longman). Seventy-eight percent of the preschools used Education, 8.5% used Crystal, 7.8% used Oxford, 3.7% used Modern and 1.9% using Longman. Since the top three publishers accounted for 94.4% of preschools in HK, standard teaching materials in the academic year 2001-2002 by these three publishers were used for the present study.

The self-compiled teaching materials in the academic year 2001-2002 were chosen randomly from three preschools of three districts with controlled socioeconomic status. Theoretically, there are nine possible combinations of income and education levels. According to the information on income level and education level of District Council districts provided by Census and Statistic Department (2001), districts were categorized into seven combinations of income level and education level (Table 1).
Table 1.

<table>
<thead>
<tr>
<th>Education level (Edu)</th>
<th>Income level (Inc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>High</td>
<td>Wan Chai</td>
</tr>
<tr>
<td>Central and West</td>
<td></td>
</tr>
<tr>
<td>Eastern</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>Tsuen Wan</td>
</tr>
<tr>
<td>Sai Kung</td>
<td>Shatin</td>
</tr>
<tr>
<td></td>
<td>Tai Po</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Yau Tsim Mong</td>
</tr>
<tr>
<td></td>
<td>Sham Shui Po</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kwai Ching</td>
</tr>
<tr>
<td></td>
<td>Wong Tai Sin</td>
</tr>
<tr>
<td></td>
<td>Kung Tong</td>
</tr>
<tr>
<td></td>
<td>Yuen Long</td>
</tr>
</tbody>
</table>

Three sets of self-compiled teaching materials, one from each of the following districts: high(Inc)-high(Edu) (HH), high(Inc)-mid(Edu) (HM) and mid(Inc)-mid(Edu) (MM), were chosen for the present study.

Characters of primary one data based on the textbook published by New Asia were selected from the Hong Kong Corpus of Primary School Chinese (Leung & Lee, 2002) because it is the most popular publisher used in Primary schools in Hong Kong.

Procedure

Chinese characters of teaching materials in these six preschools were entered into a software program FileMakerPro4.0 Chinese version. Hong Kong Corpus of Preschool Chinese characters was then formed. Chinese characters of Primary one were selected from the Hong Kong Corpus of Primary School. They are from the textbooks published by New Asia Publishing House (New Asia).

Each character from the text was entered as a separate record and each character was encoded for:

1) The character’s pronunciation in the form of International Phonetic Alphabet (IPA)
2) Its type: i) simple characters, ii) phonetic compounds, or iii) other compound figures
3) The phonetics it carries, including free and bound phonetics. Bound phonetics were coded with numbers
4) Its regularity: the characters were encoded into one of the following eight categories, i) same onset-rime and tone with its phonetic, ii) same onset-rime (but not tone) with its phonetic, iii) same onset and tone (but not rime) with its phonetic, iv) same rime and tone (but not onset) with its phonetic, v) same onset (but not rime and tone) with its phonetic, vi) same rime (but not onset and tone) with its phonetic, vii) same tone (but not onset and tone) with its phonetic, viii) different onset-rime and tone with its phonetic.

5) The position of its phonetics, they were categorized according to the system used by Shu et al. (2000). It included seven positions: i) left, ii) right, iii) top, iv) bottom, v) partly surrounded by a significs, vi) completely surrounded by significs, and vii) partly or entirely surrounding a significs.

6) The boundedness of the phonetic it carries

7) The significs it carries

8) Transparency: the characters were categorized into one of the following seven categories: i) same meaning as its significs, ii) same semantic category that the significs represents, iii) meaning directly related to that of its significs, iv) meaning indirectly related to that of its significs, v) extended meaning directly or indirectly related to that of its significs, vi) meaning unrelated to that of its significs, and vii) cannot be classified into the previous categories (adapted from the classification used by Shu et al. (2000)).

9) The position of its significs: they were categorized into one of the following six positions relative to the phonetics: left, right, top, bottom, completely surrounded a phonetics, partly surrounding a phonetics, and partly or completely surrounded by a phonetics (adapted from Shu et al. (2000))

10) Grade from which the character comes

11) Publisher of teaching materials from which the character comes

As phonetic consistency changes across grades, it is not coded in the database and the consistency value was calculated separately for each grade.
Measurement

The possible reading development in preschool children are under investigation. The possible reading development is measured by the following parameters:

i) Number of different new characters across grades. It represents the load of learning.

ii) Proportion of simple characters, phonetic compounds, and other compound characters across grades from K1 to P1. A high proportion of simple characters at earlier grades and a high proportion of phonetic compounds in later grades favor the recognition of phonetic compounds.

iii) Number of simple characters which will become phonetics and signifies in primary school across grades. A large number these simple characters favor the development of explicit representations.

iv) Proportion of different phonetic regularity of phonetic compounds across grades. A high proportion of regular phonetic compounds during early grades facilitate children to derive the explicit representation, that free phonetics provides phonological information.

v) Proportion of different phonetic consistency of phonetic compounds across grades. A high proportion of consistent phonetic compounds favor an early development of explicit representation, that free and bound phonetics provide phonological information. Later introduction of inconsistent phonetic compounds will allow children to refine their explicit representation, that characters of the same phonetic family may not have the same pronunciations.

vi) Proportion of different position of phonetics within phonetic compounds across grades. A high proportion of phonetics at a particular position facilitates the recognition of phonetics in a phonetic compound.

vii) Proportion of different transparency of phonetic compounds across grades. A high proportion of high transparency characters during early grades facilitate children to
derive the explicit representations that signifies provide semantic information.

viii) Proportion of different position of signifies within phonetic compounds across grades. A high proportion of signifies at a particular position facilitates the recognition of signifies in a phonetic compound

RESULTS AND DISCUSSION

All the parameters below are i) compared across grades, ii) compared against different teaching materials, and iii) compared against different publishers or different socioeconomic status.

Number of Different New Characters Across Grades

The number of different new characters introduced in each grade in preschool and in primary one was shown in Appendix A. The load of learning increases with grades in standard teaching materials and the load decreases with grades in self-compiled teaching materials. For standard teaching materials, the number of different new characters increases across grades, except in Education. In Education, most new characters are introduced in K3 and the load of learning new characters is the heaviest in K3. In self-compiled teaching materials, on the contrary, the number of different new characters decreases across grades generally. The load of learning is highest at K1. Except in high(Inc)-mid(Edu) condition that the load of learning new characters is heaviest at K2. Although the trend of decreasing loading is the same in high(Inc)-high(Edu) and mid(Inc)-mid(Edu) conditions; however, the numbers of new characters introduced each year are of great difference. For example, many more new characters were introduced in mid(Inc)-mid(Edu) than high(Inc)-high(Edu) conditions (639 characters in MM and 273 in HH) in K1. The load of learning is exceptionally high in mid(Inc)-mid(Edu) conditions in K1.
Proportion of simple characters, phonetic compounds, and other compound characters across grades

In self-compiled teaching materials, around 60% of different characters are phonetic compounds, nearly 20% are other compound characters and around 20% are simple characters in all grades and teaching materials (including HH, HM and MM). The proportion of phonetic compounds, other compound characters and simple characters in preschools is similar to that in primary one (see P1 of Figure 1).

The proportion of phonetic compounds is smaller in standard teaching materials (around 40% on average). The proportion of phonetic compound is smaller than that in primary one (above 50%). Unlike self-compiled teaching materials, the proportion of simple characters, phonetic compounds and other compound characters varies across grades and across teaching materials (See Figure 1).

Figure 1. The proportion of phonetic compounds, other compound characters and simple characters across grades and different standard teaching materials.

The proportion of simple characters is higher than 30% in most of the standard teaching materials across grades (except in K3 of Education). These simple characters, can be signifies or phonetics and they can act as building blocks for compound characters learnt later in
primary school (see below).

**Number of Simple Characters can be Phonetics or Significs in Primary School Across Grades**

Under 40% of simple characters learnt in preschool can be phonetics from preschool to primary one. The number of simple figures which act as phonetics across grades in different teaching materials is shown in figure 2. The figure shows both the grade that the simple characters appear (K1 to K3) and the grade at which the simple characters act as phonetics (K1 to P1). A greater number of simple figures can act as phonetics in self-compiled teaching materials, compared with the standard teaching materials.

![Bar chart](image)

**Figure 2.** The number of simple characters which become phonetics in preschool to primary 1 across different teaching materials.

Under 30% of simple characters learnt in preschool can be significs from preschool to primary one. The number of simple figures which act as significs across grades in different teaching materials is shown in figure 3. The figure shows both the grade that the simple characters appear (K1 to K3) and the grade at which the simple characters act as significs (K1 to P1). A greater number of simple figures can act as significs in self-compiled teaching materials, compared with the standard teaching materials.
Figure 3. The number of simple characters which become significs in preschool to primary 1 across different teaching materials.

Proportion of Different Phonetic Regularity of Phonetic Compounds Across Grades

The characters are then grouped into i) regular (same onset-rime and tone), ii) semi-regular (same onset-rime but not tone), and iii) irregular (different onset or rime or both). In self-compiled teaching materials, about 70% of the characters are irregular, around 13% of the characters are semi-regular, and around 17% of the characters are regular in all grades and teaching materials (HH, HM and HM). These proportions are similar to that in primary one.

Irregular characters are predominant in different standard teaching materials (over 65% in all grades) and the percentage of irregular characters is similar to that in primary one. The proportions of regular, semi-regular and irregular characters in different standard teaching materials are shown in figure 4.
Figure 4. The proportion of regular, semi-regular and irregular characters across grades in different standard teaching materials.

Proportion of Different Phonetic Consistency of Phonetic Compounds Across Grades

Consistency was only calculated for phonetic compounds with a family size equal or over two. Family size is the number of characters that share the same phonetics. For example, family size of 青, 清, 精, 情, 猜 that share the same phonetics 青 is five in this case. Phonetic consistency is categorized into two categories: i) high (with only one pronunciation within a phonetic family) and ii) low (with two or more than two pronunciation within a phonetic family).

When no characters have a family size over two, 0% is marked for all degree of consistency. In standard teaching materials, no characters have a family size over two in K1 of all publishers and K2 of Crystal standard teaching materials. In general, only 15% of the phonetic compounds in K2 and 30% of the phonetic compound in K3 have a family size over 2. In teaching materials of Education and Oxford, the percentage of high consistency characters decreases with grade.

In self-compiled teaching materials, at least 20% of the phonetic compounds have a
family size over 2 in each grade of all teaching materials. The percentage of high consistency characters is highest at K2 and decrease at K3. This implies that more high consistency characters are introduced in K2 and characters with different pronunciation as those in their phonetic family were introduced in K3.

As more characters are introduced in later grades (K3), the number of pronunciations allowed for each phonetic family increased as the number of characters in a phonetic family increased. With an increase in the number of pronunciations, consistency decreases. Thus, consistency decreases with grade when more characters are introduced (see Appendix B).

Proportion of Different Positions of Phonetics within Phonetic Compounds Across Grades

Phonetics on the right position is predominant in all grades and all teaching materials (around 60%). Proportions of phonetics in different positions within phonetic compounds across grades in different teaching materials are shown in Appendix C.

Proportions of different transparency of phonetic compounds across grades

Phonetic compounds can be grouped into four transparencies: transparent (category i, ii and iii, see Method), semi-transparent (category iv and v, see Method), opaque (category vi, see Method) and other (category vii, Method)

Similar proportions of transparent (around 40%), semi-transparent (around 20%) and opaque (around 40%) characters were found across grades in all self-compiled teaching materials. These proportions are similar to those in primary one. In standard teaching materials, proportions of different transparency appeared differently across grades in different teaching materials (Crystal, Oxford and Education) (see Appendix D).

Proportion of Different Position of Significs Within Phonetic Compounds Across Grades

Significs on the left position is predominant in all grades and all teaching materials (around 60%). Proportions of significs in different positions within phonetic compounds across grades in different teaching materials are shown in Appendix E.
GENERAL DISCUSSION

The objectives of this study are two folded: First, to investigate the change in characteristics of characters, taught in different preschool grades which provide information on the development of explicit representations during preschool and how well children are prepared for primary school, and secondly, to investigate the differences in the characteristics of characters between different standard teaching sets and self-compiled teaching sets from districts of different socioeconomic status.

Development of Explicit Representations During Preschool

To establish explicit representation on how phonetics and significs provide phonological and semantic information, one should be able to i) identify phonetic compounds, ii) identify phonetics and significs in phonetic compounds, iii) understand how phonetics and significs symbolize phonological and semantic information of the whole characters.

Identification of Phonetic Compounds

Results show that only a small number of simple characters taught in preschools can be phonetics or significs, especially those in standard teaching materials. It suggests that children might be slow to recognize whether a character is a phonetic compound. Provided that phonetics and significs of the characters are free: when children have not learnt the sub-character components as independent characters, they cannot recognize that the phonetic compounds consist of meaningful sub-character units. Then, children can hardly recognize phonetic compounds. Not knowing the pronunciation of sub-character units, phonetics can hardly be identified. The small number of simple characters taught suggests that the identification of phonetic compounds and its phonetics may be slow to develop in preschool children.

Identification of Phonetics in Phonetic Compounds

A high proportion (around 60%) of phonetics is on the right in all grades and teaching
materials. Children may make use of this positional information to identify phonetics. This is supported by Shu et al (2000), who suggested children might first develop a representation that the sub-character components on the right will provide pronunciation. The children might be able to make use of this information for recognizing phonetics in phonetic compounds. Shu and his colleagues suggested that the representation will be refined when they are exposed to more characters and find that phonetics are not always on the right.

**Development of Explicit Representation on How Phonetics Symbolize Phonological Information**

The results of family size, phonetic regularity and phonetic consistency suggest that explicit representation, that how the phonetics represent pronunciation, is not favorable to develop in children at preschool level. It has been suggested that children develop explicit representation on how phonetics represents sounds by analogy (Ho & Bryant, 1997a). In standard teaching materials, there are over 40% phonetic compounds. However, only a small proportion of them have family sizes equal to or over two (generally 0% in K1, 15% in K2 and 30% in K3). Thus, only a small proportion of characters that share the same phonetics can be learnt by analogy. Even if they can be learnt by analogy, most of the characters (above 65%) are of low consistency, which share different pronunciations. Data from regularities also reveal that 70% of characters are irregular. The phonetic information carried by phonetics is likely to be unreliable in predicting the pronunciation of the characters in standard teaching materials. This suggests that children need to be exposed to a number of characters to derive the relationship between phonetics and pronunciation of the whole characters. The explicit representation is then slow to develop.

Self-compiled teaching materials are composed of i) a higher proportion of phonetic compounds (around 60% in all grade and teaching materials), ii) a higher proportion of characters with family size over two (ranged from an average of 30% in K1 to over 50% in
K3) and iii) a higher average proportion of high consistency characters. They suggest a faster development of explicit phonetic representation. In addition, there are more characters in self-compiled teaching materials. It is likely that rote memory cannot support children to recognize a lot of characters. They are prone to develop explicit representation at an earlier age, compared with those children using standard teaching materials, as its development aids children’s reading. In addition, children using self-compiled teaching materials have more experience in Chinese characters. Their development of explicit representations would then predicted to be faster, compared to those children using standard teaching materials.

Identification of Significs in Phonetic Compounds

About 60% significs are posited on the left in all grades and teaching materials. Children may be able to make use of this positional information in early identification of significs.

Possible Explicit Representation on How Significs Symbolizes Semantic Information

To develop explicit representation on how significs symbolizes semantic information, one should be able to i) identify phonetic compounds, ii) identify significs in phonetic compounds, iii) understand how significs symbolize phonological information of the whole characters.

The proportion of transparent characters is around 40% in all grades of different self-compiled teaching materials. When semi-transparent characters are taken into account, sixty percent of characters provide semantic information in all grades of self-compiled teaching materials. The average proportion of transparent and semi-transparent characters is lower in standard teaching materials.

A high proportion of transparent and semi-transparent characters should favour the development of explicit representation of how significs symbolize semantic information. According to Shu and Anderson (1997), Beijing primary one children who have been exposed to a higher proportion of transparent (50%) and semi-transparent characters (30%) have not
clearly shown their development of explicit representation. It is hypothesized that a higher proportion of transparent characters will facilitate the development of explicit representation on how significs carry semantic information. A smaller proportion of transparent characters in Hong Kong preschool teaching materials suggest that explicit representation on how significs symbolize semantic information is not likely to be developed at Hong Kong preschool levels.

How Well Preschool Children are Prepared for Primary School

How well preschool children are prepared for primary school varies with the type of teaching materials.

Standard teaching materials. A smaller number of characters are included in standard teaching materials. A greater number of characters are then need to be learnt in primary one. The load of learning is the highest in primary one, except for that in Education. The load of learning is highest at K3 in Education.

Development of explicit representations on how phonetics and significs symbolize phonological and semantic information is predicted to be slow in preschoolers using standard teaching materials. Developing explicit representation aids learning to read. A slow development of explicit representation and a great number of new characters need to be learnt, suggest an inadequate preparation for primary one. Among standard teaching materials, children using Education teaching materials need to learn fewer characters in primary one. Children using Education teaching materials have better preparation for primary one. In addition, the number of simple characters that can act as significs or phonetics is smaller in standard teaching materials. These simple characters act as building blocks for phonetic compounds learnt at school level. A smaller number suggest less favorable learning of phonetics compounds in primary one.

Self-compiled teaching materials. A large number of characters are included in self-compiled teaching materials. Only a small number of characters are then need to be learnt
in primary one, especially in preschool in high(Inc)-mid(Edu) and mid(Inc)-mid(Edu) districts. The load of learning is the lightest in primary one.

Development of explicit representation on how phonetics and significs symbolize phonological and semantic information should be comparatively faster in preschoolers using self-compiled teaching materials. A relatively faster development of explicit representations that aid learning to read and a small number of new characters need to be learnt, suggest a better preparation for primary one. Moreover, the number of simple characters that can act as significs or phonetics is larger in self-compiled teaching materials. These simple characters act as building blocks for phonetic compounds learnt at school level. A larger number suggest more favorable learning of phonetic compounds in primary one.

Differences Between Standard Teaching Materials and Self-compiled Teaching Materials

Standard teaching materials. As already mentioned, standard teaching materials include fewer characters. The load of learning is lighter in children using standard teaching materials. It is also reasonable for preschoolers to learn characters with an increasing load of learning with grades. However, the fewer number of characters make children less experienced with characters. They should then develop explicit representations at a slower pace. The fewer number of simple characters act as phonetics or significs, a lower proportion of phonetic compounds, phonetically regular and consistent characters, and, a lower proportion of semantically transparent characters also suggest a slower development of explicit representations.

Self-compiled teaching materials. The self-compiled teaching materials included a large number of characters. The load of learning is much heavier in children using standard teaching materials, especially at early grade. It is unreasonable for preschoolers to learn characters with a heavy load of learning at an early grade. However, the greater number of characters allow children to gain more experience with characters. They are likely to have a
faster development of explicit representation. The greater number of simple characters act as phonetics or significs, a higher proportion of phonetic compounds, phonetically regular and consistent characters, and a higher proportion of semantically transparent characters also infer a faster development of explicit representations.

Differences Among Standard Teaching Materials and Self-compiled Teaching Materials

Standard teaching materials. Comparing the number of new characters in teaching materials of Crystal, Oxford and Education, the load of learning varied across grade among these different publishers. The load of learning increases from K1 to primary one in Crystal. Compared with the other two publishers, the load of learning in K1 is highest in Crystal. The load of learning increases across grades in the teaching materials of Oxford, except for a slight decrease in K3. For Education, the load of learning is increasing from K1 to K3, being the heaviest in K3.

The sudden increase in the number of new characters learnt in K3 of Education may give rise to the memory problem in character recognition. It may facilitate the development of explicit phonetic or semantic representations for learning novel characters in K3. In Crystal and Oxford, the load of learning is heaviest in primary one. The need to develop explicit phonetic or semantic representations occurs in a later grade and so the development of representations may start at a later grade, primary one for example.

Self-compiled teaching materials. The load of learning generally decreases from K1 to primary one, except in preschool of high(Inc)-mid(Edu) district. The load of learning is highest at K2 in preschool of high(Inc)-mid(Edu). Though the load of learning decreases with grades in both preschool of high(Inc)-high(Edu) and mid(Inc)-mid(Edu) districts, the magnitudes of the decrease are different. A slight decrease with grades is noticed in preschool of high(Inc)-high(Edu) district while an abrupt decrease with grades is noticed in preschool of mid(Inc)-mid(Edu) district. The load of learning is exceptionally high in preschool of
mid(Inc)-mid(Edu) district. 639 different new characters are learnt in K1 of preschool of high(Inc)-high(Edu) district. On average the load of learning is heaviest in mid(Inc)-mid(Edu) situation and the load of learning in high(Inc)-mid(Edu) situation is heavier than that in high(Inc)-high(Edu) situation. The load of learning is generally too heavy for children using self-compiled teaching materials. The appropriateness of selection of teaching materials should be in descending order is high(Inc)-high(Edu), high(Inc)-mid(Edu) and mid(Inc)-mid(Edu). Both income and the education levels of a family may have an effect on the selection of teaching materials, in this sample of teaching materials. Higher socioeconomic status tends to have better selection of teaching materials.

Based on the above evidence, the development of explicit phonetic and semantic representations should be slow in a preschool reader. Self-compiled teaching materials should facilitate children to have a faster development of explicit representations. For children using standard teaching materials, those using materials of Education, the most popular publisher, should have a relatively faster development of explicit representation among those using standard teaching materials. Socioeconomic status of the districts that preschools are situated has effect on the appropriate selection of self-compiled teaching materials.

RECOMMENDATIONS AND FURTHER STUDY

As the development of explicit representations on how phonetics and significs symbolize phonetic and semantic information help children to learn to read, characters that facilitate the faster development of these representations should be taught first. Simple characters that can become phonetics or significs should be taught first so that children can recognize the phonetic compounds. These simple characters can also act as building blocks for later learning of phonetic compounds. Regular, consistent and transparent character should then be taught first. By analogy, children can then easily develop the explicit representations, that
phonetics or signifies carry phonetic and semantic information of the characters. In addition, teaching instructions that explicitly explain the use of phonetics and signifies are recommended.

This study suggests that the development of explicit representation of how phonetics and signifies symbolize phonological and semantic information is slow at preschool level. However, children can still recognize a number of characters. Further study on the possible strategies applied by a presholder in reading Chinese characters should be carried out. Visual skills are suggested to be important in the early recognition of characters (Ho & Bryant, 1997b). Detecting and recognizing the visual pattern of orthographic units are but one of the possible strategies applied. Whether children make use of sub-character orthographic units for character recognition need further investigation.

The present study makes inferences on the possible development of explicit representations. Further studies to confirm i) the age at which children can identify phonetic compounds, ii) the age at which children can identify phonetics and signifies, and iii) the age at which children recognize the role of phonetics and signifies, would be valuable to confirm the development of explicit representation in preschool. Children using different teaching materials should be compared for their explicit representation development. Recommendations on which type of teaching materials should be used to facilitate explicit representation development can then be made.
REFERENCE


Chinese with Implications for Theory and Practice. Champaign, IL: Center for the Study of Reading.


### APPENDIX A

**Number of Different New Characters and Different Learnt Characters Across Grades in Different Teaching Materials**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Type</th>
<th>Standard teaching materials</th>
<th>Self-compiled teaching materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Crystal</td>
<td>Oxford Education</td>
</tr>
<tr>
<td>K1</td>
<td>No. of different new characters</td>
<td>37</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>No. of different learnt characters</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>K2</td>
<td>No. of different new characters</td>
<td>97</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>No. of different learnt characters</td>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>K3</td>
<td>No. of different new characters</td>
<td>176</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>No. of different learnt characters</td>
<td>117</td>
<td>127</td>
</tr>
<tr>
<td>P1</td>
<td>No. of different new characters</td>
<td>332</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td>No. of different learnt characters</td>
<td>207</td>
<td>245</td>
</tr>
</tbody>
</table>
APPENDIX B

Proportion of Different Consistency Characters Across Grades in Different Teaching Materials

<table>
<thead>
<tr>
<th>Grade</th>
<th>Consistency</th>
<th>Standard teaching materials</th>
<th>Self-compiled teaching materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Crystal</td>
<td>Oxford</td>
</tr>
<tr>
<td>K1</td>
<td>High</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K2</td>
<td>High</td>
<td>0%</td>
<td>32%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0%</td>
<td>68%</td>
</tr>
<tr>
<td>K3</td>
<td>High</td>
<td>7%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>93%</td>
<td>84%</td>
</tr>
<tr>
<td>P1</td>
<td>High</td>
<td>20%</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>80%</td>
<td>73%</td>
</tr>
</tbody>
</table>
APPENDIX C

The proportion of phonetics at various positions within phonetic compounds across grades in various teaching materials
APPENDIX D

Proportions of Transparent, Semi-transparent and Opaque Characters in Different Teaching Materials Across Grades

<table>
<thead>
<tr>
<th>Grade</th>
<th>Transparency</th>
<th>Standard teaching materials</th>
<th>Self-compiled teaching materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Crystal</td>
<td>Oxford</td>
</tr>
<tr>
<td>K1</td>
<td>Transparent</td>
<td>7%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>Semi-transparent</td>
<td>20%</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Opaque</td>
<td>73%</td>
<td>25%</td>
</tr>
<tr>
<td>K2</td>
<td>Transparent</td>
<td>26%</td>
<td>34%</td>
</tr>
<tr>
<td></td>
<td>Semi-transparent</td>
<td>31%</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>Opaque</td>
<td>44%</td>
<td>43%</td>
</tr>
<tr>
<td>K3</td>
<td>Transparent</td>
<td>34%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td>Semi-transparent</td>
<td>18%</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>Opaque</td>
<td>49%</td>
<td>36%</td>
</tr>
</tbody>
</table>

P1  | Transparent: 36%  | Semi-transparent: 22% | Opaque: 42%
APPENDIX E

The proportion of signics at various positions within phonetic compounds across grades in various teaching materials.
ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr. Leung Man Tak for his unconditional support and guidance during the development of this research. I also thank principals and teachers of the participated preschools for their provision of teaching materials.