Thai Undergraduate Student Awareness of Regular and Irregular Grapheme–Phoneme Correspondence

David D. Perrodin

Mahidol University, Thailand

Contact: David D. Perrodin daviddperrodin@gmail.com

ABSTRACT
Recent events in Thailand in reference to the teaching of phonics for better comprehension of English vocabulary have highlighted the overuse of identifying letter-sound relationships in English by utilizing the familiar Thai orthography to assist developing Thai EFL learners. This paper investigated the long-term effects of using such pedagogy on recognizing regular and irregular Grapheme–Phoneme Correspondences (GPC) in English by Thai undergraduate students. To address this matter, the study used a convenience sampling of 373 first-year university students from 11 general education English classes at a mid-sized private university near Bangkok, Thailand. The familiar English poem I Take It You Already Know was employed for data collection, for it consists of a practical ratio of frequent and infrequent English grapheme-phoneme correspondences. Extensive lists of recognized grapheme-phoneme correspondences were used to identify the frequent or regular and the infrequent or irregular main phoneme present in each of the 60 most frequently queried content words of the poem. Point-Biserial Correlation was employed to measure the strength of association between the frequency occurrence of the most queried content words from the data set and the examined weighted word frequency data. The findings suggest that, in general, the Thai undergraduate students demonstrated an overall lack of recognition of regular and irregular Grapheme-Phoneme Correspondences of English.

KEYWORDS: Grapheme-Phoneme Correspondence; Orthography; Phonics; Pronunciation;

ARTICLE HISTORY
Received dd mm yyyy
Accepted dd mm yyyy

Introduction
A growing body of literature recognizes that regardless of the categorization of English as a Foreign Language (EFL) learners, comprehension of English vocabulary at all levels has been found to originate from English reading proficiency ((Frost, 2005); (Henderson, 2017)). Although there are still limitations to the current empirical work in this area, evidence that has received increased attention over the past few
years in Thailand suggests that Thai EFL learners have gained recognition of Grapheme-Phoneme Correspondence of English as a result of extensive English reading ((Vibulpatanavong & Evans, 2019); (Winskel & Iemwanthong, 2009)).

Alternatively, recent events in Thailand in reference to the teaching of phonics for better comprehension of English vocabulary have highlighted the overuse of identifying letter-sound relationships in English by utilizing the familiar Thai orthography to assist young Thai EFL learners in 'correctly pronouncing written English words' ((Ladkert, 2009); (Nensiri & Sukavatee, 2018); (Wixey & Eamoraphan, 2017)). Unfortunately, the lasting effects of such practices continuing to be utilized at the undergraduate level have attracted very little attention from Thailand’s scholarly community. For that reason, this investigation will attempt to provide a snapshot of Thai undergraduate students’ recognition of regular and irregular Grapheme-Phoneme Correspondence of English.

The following pages will illustrate that although in the initial stages of learning English vocabulary, it may be necessary to manipulate some English phonemes (Yavas, 2016) with the conventional Thai spelling system to make English vocabulary comprehensible, there must be limits as to how much long-term exposure a learner should have to this pedagogy ((Nensiri & Sukavatee, 2018); (Sayeski, Earle, Eslinger, & Whitenton, 2016); (Thep-Ackrapong, 2005); (Vibulpatanavong & Evans, 2019))

**Recognition of Written Word Forms and Their Phonemic Representations**

To become skillful readers of English, EFL language learners must acquire the ability to recognize the written word forms as intimately interconnected and explicable only by reference to the whole word (orthographical decoding), combined with the ability to interpret printed English words letter by letter into phonemic representations (phonological recoding) (Knoepke, Richter, Isberner, Naumann, & Neeb, 2014).
While a variety of explanations of the hypothesis of orthographical decoding have been suggested, this paper will use one adapted from Pollatsek & Treiman (2015), who saw orthographical decoding as an ability to apply background knowledge of grapheme-phoneme correspondence, including knowledge of letter patterns of English, to recognize words correctly, and likewise, to decipher unfamiliar words within English. As well, orthographical decoding of English requires some background knowledge and skills of English such as the knowledge of grapheme-phoneme relations, the ability to blend or merge the separate phonemes into a coarticulated whole, and the knowledge of displayed vocabulary to recognize the meanings of the words which are to be pronounced. As a result, beginner EFL learners must receive extensive exposure to, instruction in, and practice of English to gain a rudimentary knowledge of English in order to be able to convert graphemes into phonemes and to hold the sounds in their memory long enough to correspondingly blend the consonants and vowels in the correct order for proper pronunciation ((Christensen & Bowey, 2005); (Knoepke et al., 2014); (S. D. Krashen, 2009); (Oney & Goldman, 1984)).

Relatedly, Garnham (1985), in their earlier work, introduced the hypothesis of phonological recoding as being a solution to the comprehending of words during the process of early reading. Therefore, the recognition of grapheme-phoneme correspondences is primarily acquired by EFL learners early on during the process of reading and speaking English, or in other words, the promotion of the phonological recoding of letter patterns into the sound patterns by the auditory analysis of the spoken words (Vandervelden & Siegel, 1995). Thus, EFL learners apply the principles of orthographical decoding in their process of using grapheme-phoneme correspondences to recognize words and phonological recoding in identifying the systematic relationships which correspond between the graphemes and phonemes of English to retrieve the pronunciation of
unknown vocabulary ((Gates, 2017); (Gerlach, 2016); (Oney & Goldman, 1984); (Tokuda, 2016); (Vandervelden & Siegel, 1995)).

Grapheme–Phoneme Correspondence

According to the definition provided by Yavas (2016), Grapheme-Phoneme Correspondence (GPC) is the process of matching together the graphemes and phonemes, the conventional spelling and letter pattern system of a language, and vice versa. A phoneme is the smallest unit of sound within a word in a particular language where each unit of sound distinguishes one word from another in that language and is represented by or corresponds with a grapheme (Pollatsek & Treiman, 2015). At the same time, a grapheme symbolizes phonemes within words and represents consonants and vowels, respectively, as well as letters or combinations of letters that distinguish words from each other (Reid & Elbeheri, 2009). Hence, it is the connections between graphemes and phonemes, which provide foundational elements that enable EFL learners to recognize the sounds of words and acquire essential reading skills (Pollatsek & Treiman, 2015).

The classifications of most writing systems are usually based on the notion of a phoneme being represented by a grapheme as being the smallest functional characteristic unit of any writing system (Henderson, 2017). However, even though the English language, for instance, is made up of twenty-six letters and forty-four phonemes that are composed of as few as one letter and as many as four letters, and approximately 280 graphemes which are involved in over 540 grapheme-phoneme correspondences, the spelling system of English is variable and consists of alternative ways to represent phonemes and alternative sounds for graphemes (Yavas, 2016).

Similarly, as mentioned previously, phonics is a teaching method to read by correlating sounds (phonemes) with letters or groups of letters (graphemes) in the English alphabetic writing system. Phonics involves the development of phonemic awareness and the knowledge of grapheme-
phoneme correspondences, and often the recognition of spelling patterns within the English language (Yavas, 2016). Interestingly, the grapheme-phoneme correspondence of some English words is considered frequent when graphemes represent only one phoneme each, and other frequent phonemes, and irregular or infrequent when the pronunciations of these words do not conform to the general grapheme-phoneme system. Consequently, this inconsistency and irregularity in the English orthography tend to reduce the transparency of the English grapheme-phoneme sound system, thereby regularly delaying phonemic awareness, and similarly, delaying reading acquisition of the English language for Thai EFL learners (Liao, Kuo, Deenang, & Mok, 2015); (Vibulpatanavong & Evans, 2019); (Winskel & Iemwanthong, 2009).

**Grapheme-Phoneme Frequency Count**

In their research, Fry (2004) summarized and simplified the extensive data collected in a previous grapheme-phoneme frequency count by Hanna et al. (1966). Fry reanalyzed the considerable amount of data in order to streamline the original findings and make the conclusions more practical for reading and spelling instruction and EFL teachers and curriculum designers alike. In their classification of both vowel and consonant grapheme-phoneme correspondences, Fry (2004) assigned classifications, although somewhat arbitrary, to assist in making instructional decisions and provide an empirical frequency summary of grapheme-phoneme correspondences. The frequency of grapheme-phoneme correspondences are typically considered ‘Frequent’ or ‘Regular’ when having a high frequency of graphemes that represent only one phoneme each, and other frequent phonemes, and ‘Infrequent’ or ‘Irregular’ when having a low frequency of otherwise less or uncommon phonemes ((Brooks, 2015); (Ziegler, Stone, & Jacobs, 1997)). Although the study by Fry (2004) was intended to provide content for phonemic awareness instruction, a potential source of bias for
this study is the researcher was focused on the impression that phonics and phonemic awareness are the answers to successful literacy acquisition.

In contrast, other studies related to phonics and phonemic awareness have concluded that this area’s overall research does not provide a suitable basis for concluding the necessity of phonemic awareness training for EFL learners ((S. D. Krashen, 2009); (2001); (Smith, 2004)). Likewise, other studies have found that language learning and reading skills crucially depend on acquiring grapheme-phoneme correspondences (See Ziegler et al.(2002) Appendixes B & C for one of the most comprehensive mappings of grapheme-phoneme correspondence).

**English Pronunciation and Thai Orthography**

Studies have shown that the teaching of English pronunciation is often neglected in primary and secondary level EFL classrooms in Thailand because Thai teachers of English often lack proficient English pronunciation ability ((Ladkert, 2009); (Nensiri & Sukavatee, 2018); (Wixey & Eamoraphan, 2017)). Equally, the general differences of grapheme-phoneme correspondences in English pronunciation may also pose a problem for even highly proficient Thai EFL learners. As mentioned earlier, the English language comprises twenty-six letters, consisting of twenty-one consonant graphemes and five vowel graphemes, representing forty-four phonemes ((Reid & Elbeheri, 2009); (Yavas, 2016)). Comparably, the Thai language consists of forty-four consonant graphemes, twenty-one phonemes as initial consonants and eight phonemes as final consonants, and twenty-eight vowel graphemes, representing a very complex set of Thai language phonemic rules (Wilairatana, Mizutani, Kuntonbutr, & Tsutomu, 2019).

Research has shown that the probable cause of English pronunciation difficulty by Thai EFL learners is the interference of different phonetic
representations of corresponding phonemes in English and Thai languages ((Ladkert, 2009); (Nensiri & Sukavatee, 2018), (Wixey & Eamoraphan, 2017)). These studies that have attempted different strategies for improving English pronunciation with Thai EFL learners found significant positive associations between phonemic awareness, verbal short-term memory, and working memory with vocabulary spelling success. These researchers recommended that phonics be taught as a 'preparation' for English learners' further practice in early spelling features, especially for final consonants and short vowels. This letter-sound knowledge should be measured to improve ongoing spelling achievement. However, these studies failed to consider the broader implications of the lack of vocabulary comprehension the learners would face with long-term exposure to this type of instruction and practice ((Algethami, 2016); (Perry et al., 2002)).

This research, therefore, aims to show that it may be necessary to abandon the practice of more advanced EFL learners correlating English phonemes with the Thai spelling system once learners become more proficient in English and in their level of confidence regarding their ability to pronounce English vocabulary (S. Krashen, 2001); (William, 2016).

**Method**

**Research Questions**

What are the most queried content words of the general English poem I Take It You Already Know in which learners correlated English phonemes with the Thai spelling system?

**Target Population**

This project used a convenience sampling of 454 first-year students from 11 general education English classes at a mid-sized private university on the outskirts of Bangkok, Thailand. A final participation rate of 373 (N=373) students (82%) of the initial 454 students recruited from the 11 groups to participate in this study.
**Participant Profile**

The total participants were 373 (N=373) first-year students from different faculties who took part in the first year first-semester 'English I' general education classes. The students in the sampling were mostly 18 or 19 years old and had an average of 12 years of exposure to English as a Foreign Language education.

**Data Collection**

The general English poem I Take It You Already Know (Watt, 1954) was employed for data collection. This instrument was chosen for it contains a practical ratio of frequent/regular and infrequent/irregular English grapheme-phoneme correspondences. The students were given Research Participant Consent Forms written in Thai and the reason for data collection by their Thai lecturers. It was expressed in Thai that the data will be collected to investigate the most queried content words of the poem in which the participants correlated English phonemes with the Thai spelling system.

**Measures**

The poems were then collected from the 11 groups and analyzed. Two independent markers meticulously reviewed each of the poems. Since content words such as adjectives, adverbs, nouns, and verbs were the focus of this analysis, function words such as auxiliary verbs, conjunctions, determiners, prepositions, and pronouns were disregarded.

The markers noted that there were 60 content words in which the participants most often queried the letter-sound relationships. These 60 content words were tallied, and frequency was used to generate a list of the most queried vocabulary words. Then, comprehensive lists of ‘weighted word frequency’ data (Beale, 2019), a scaled indication of the relationships between sound, spelling, and word frequency in a major corpus frequency list, of frequent/regular and infrequent/irregular English grapheme-
phoneme correspondences identification ((Beale, 2019); (Brooks, 2015); (Fry, 2004); (Ziegler et al., 1997)) was used to identify the frequency of the main phoneme present in each word (Larsen, Kohnen, Nickels, & McArthur, 2015).

For classification, grapheme-phoneme correspondences with a frequency higher than 100 were classified as Frequent/Regular and with a frequency of 100 or less as Infrequent/irregular (Beale, 2019). The Expected Value or Mean (Fairclough, 2010) of the frequency occurrence was then be calculated. A subjective assumption of 80 percent of words above the expected value should have an infrequent or irregular main phoneme present in the word. Finally, the vocabulary words were sorted in descending order with a higher degree of queried frequency near the top (Brooks, 2015); (Fry, 2004); (Ziegler et al., 1997). Point-Biserial Correlation was utilized to measure the strength of association between the occurrence of the frequency of queried vocabulary words from the poem and the weighted word frequency data (Beale, 2019); (Fry, 2004); (Ziegler et al., 1997).

Result
The frequency of queried vocabulary words within the pronunciation poem is displayed in Table 1 below. The most prominent finding to emerge from this study is that 35 (58%) of the 60 most queried words have a frequent or regular main phoneme present in that word, and 25 (42%), presented in bold, have an infrequent or irregular main phoneme present ((Beale, 2019); (Thep-Ackrapong, 2005); (William, 2016); (Ziegler et al., 1997)).

<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>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hiccough</td>
<td>200</td>
<td>16</td>
<td>debt</td>
<td>119</td>
<td>31</td>
<td>traps</td>
<td>75</td>
<td>46</td>
<td>rose</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>dough</td>
<td>181</td>
<td>17</td>
<td>pear</td>
<td>112</td>
<td>32</td>
<td>heard</td>
<td>75</td>
<td>47</td>
<td>bother</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>threat</td>
<td>172</td>
<td>18</td>
<td>dreadful</td>
<td>109</td>
<td>33</td>
<td>font</td>
<td>74</td>
<td>48</td>
<td>card</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>bough</td>
<td>169</td>
<td>19</td>
<td>sigh</td>
<td>106</td>
<td>34</td>
<td>deed</td>
<td>73</td>
<td>49</td>
<td>brother</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>thwart</td>
<td>169</td>
<td>20</td>
<td>beard</td>
<td>98</td>
<td>35</td>
<td>dead</td>
<td>70</td>
<td>50</td>
<td>match</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 1
The Occurrence of Frequency of Most Queried Vocabulary Words
The Expected Value (Fairclough, 2010) of the frequency from the table above was calculated. The Expected Value from the table showing frequencies of the 60 most queried words was calculated at 87 ($\bar{x} = 87$). Therefore, any vocabulary word with a frequency of 87 or higher should be considered noteworthy. Subsequently, of the most common queried vocabulary words above, the Mean was then defined according to the expected values in the developed table. Noticeably, of the 23 words present above the expected value ($\bar{x} = 87$), the distribution of words with 9 frequent and 14 infrequent grapheme-phoneme correspondences presents a slight dichotomy at 39 percent and 61 percent, respectively. This was far below the assumption that 80 percent of words above the expected value have an infrequent or irregular main phoneme present in the word.

Similarly, the 37 words present below the Mean have a more comprehensive distribution range with 26 frequent and 11 infrequent grapheme-phoneme correspondences. Thus, it was surprising that the number of words with an infrequent or irregular main phoneme present in that word did not appreciably differ from the words with a frequent or regular main phoneme present in that word to be found above the Mean 15 words, and below with 11 words. Furthermore, one could concur that the more challenging infrequent or irregular grapheme-phoneme correspondences would have been the most queried content words and therefore should have been located above the Mean or Expected Value ($\bar{x} = \ldots$).
87) in the frequency list of transliterated vocabulary words ((Perry et al., 2002); (Sayeski et al., 2016); (Ziegler et al., 1997)).

As shown in Table 2 below, Point-Biserial Correlation was utilized to measure the strength of association between the occurrence of the frequency of queried 60 (N=60) content words from the poem and the weighted word frequency data ((Beale, 2019); (Fry, 2004); (Ziegler et al., 1997); (Wheelan, 2014)). The findings show a negative relationship (r = - .137) between the two variables with a p-value (p = 0.296) higher than 0.05 (p = > 0.05) thus indicating strong evidence that the results are not statistically significant between the two variables.

Table 2
Strength of Association Between Queries and Weighted Word Frequency (N=60)

<table>
<thead>
<tr>
<th>Correlation (r)</th>
<th>Sig. (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.137</td>
<td>0.296</td>
</tr>
</tbody>
</table>

Another important finding was that this study produced results that corroborate with Ziegler et al. (1997) 's conclusions in their previous work regarding the various pronunciations for -ough. As shown in Table 1 above, -ough is found in over one-third of the 23 content words located above the Mean. Therefore, these results may be an accurate representation of the challenges of recognizing infrequent grapheme-phoneme correspondences ((Brooks, 2015); (Perry et al., 2002); (Sayeski et al., 2016); (Ziegler et al., 1997)).

Discussion
The present study directly correlated grapheme-phoneme correspondence in the most frequently queried content words in which the participants associated English phonemes with the Thai spelling system in the employed pronunciation poem. Comprehensive lists of frequent and infrequent English grapheme-phoneme correspondences identification
(Brooks, 2015); (Fry, 2004); (Ziegler et al., 1997) were used to examine whether first-year Thai university students were aware of and could successfully distinguish between regular and irregular phonology-orthography relationships after an average of 12 years of English as a Foreign Language education (Liao et al., 2015); (Perrodin & Thupatemee, 2018); (Thep-Ackrapong, 2005); (Vibulpatanavong & Evans, 2019); (Winskel & Iemwanthong, 2009).

One of the more significant findings to emerge from this study is that there are 9 content words, slightly less than half of the total words, with frequently occurring grapheme-phoneme correspondences out of the 23 words which occur above the Mean. As derived from the findings in the above tables, the trend among Thai undergraduate students is to rely on the correlation of English phonemes with the Thai spelling system to identify frequently and infrequently occurring grapheme-phoneme correspondences (Knoepke et al., 2014); (Ladkert, 2009); (Liao et al., 2015); (Thep-Ackrapong, 2005); (Wixey & Eamoraphan, 2017). Considering the above findings, Thai undergraduate students with a lifetime of English as a Foreign Language education still do not possess the capability to differentiate and identify English grapheme-phoneme correspondences appropriately.

These findings suggest that in general, while the use of correlating English phonemes with the known spelling system of Thai may legitimately aid in the beginning stages to assist EFL learners in their ability to recognize the pronunciation of English vocabulary, the overuse of this pedagogical strategy has led to Thai undergraduate students to conflate pronunciations of Thai and English mistakenly, and has lead to an overall lack of comprehension of English vocabulary (Algethami, 2016); (Sun, Zhu, Chen, & Zhou, 2015); (Vaisman & Kahn-Horwitz, 2019); (Vibulpatanavong & Evans, 2019); (Winskel & Iemwanthong, 2009).
Given the findings from this study, some limitations may become evident. First, the poem only consisted of a small sample size of 60 designated content words applied for this analysis. Second, even though the learners were not entirely aware nor able to successfully identify frequent grapheme-phoneme correspondences, this issue may stem from the fact that some of the question words may not frequently occur in either the recognized British National Corpus (BNC) or the Corpus of Contemporary American English (COCA) corpus. Although quite arduous of a task, verifying the frequency occurrence of each vocabulary word utilized in this study with the work conducted by Beale (2019) in their comparison of such word usage frequencies with a standard corpus as to generate a weighted frequency of grapheme-phoneme correspondence, was an exciting research undertaking (Gontijo, Gontijo, & Shillcock, 2003).

While the comprised word frequency lists of Beale (2019) were primarily used in this study to identify frequent and infrequent English grapheme-phoneme correspondences identification were fairly comprehensive, there was still a limited amount of sample words for each phonogram and sound relationship. Furthermore, other word frequency lists ((Brooks, 2015); (Fry, 2004); (Ziegler et al., 1997)) were employed for the identification of grapheme-phoneme correspondences within the few remaining vocabulary words, therefore the slightly different arrangement and manipulation, to a certain extent, of the additional lists may have presumably affected some frequency accuracy of English grapheme-phoneme correspondences identification. Finally, the present study was carried out with a relatively small number of participants. Most participants possessed a beginner to elementary English proficiency level, which was revealed in a similar previous study (Perrodin & Thupatemee, 2018). Therefore, the level of English proficiency of the learners may as well affected the results.
Conclusion
This study has raised important questions about the nature of teaching and learning English pronunciation and the recognition of Grapheme-Phoneme Correspondences in Thailand. To begin with, although the teaching of phonics and phonemic awareness for better comprehension of English vocabulary may have its place in the early stages of English acquisition, the overuse of identifying letter-sound relationships in English by the correlation of English words with the familiar Thai orthography to assist young Thai EFL learners in 'correctly pronouncing written English words' has been shown to interfere with the students’ awareness of Grapheme-Phoneme Correspondences (Nensiri & Sukavatee, 2018); (Wixey & Eamoraphan, 2017). This research has shown that it may be necessary to abandon the practice correlating English phonemes with the Thai spelling system for more advanced EFL learners in order for them to become more proficient in English and in their level of confidence regarding their ability to pronounce English vocabulary (S. Krashen, 2001); (William, 2016). Finally, this study strengthens the idea that the lasting adverse habits generated from such phonic teaching practices in the Thai education system should become a focal point for Thailand's scholarly community.

References


