The Effect of Furigana on Lexical Inferencing of Unknown Kanji Words

Joy A. Palmer
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Other Languages, Societies, and Cultures Commons

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/etd/3235

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
The Effect of Furigana on Lexical Inferencing of Unknown Kanji Words

Joy Palmer

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

J Paul Warnick, Chair
Masakazu Watabe
Dan P. Dewey

Center for Language Studies
Brigham Young University

June 2012

Copyright © 2012 Joy Palmer
All Rights Reserved
ABSTRACT

The Effect of Furigana on Lexical Inferencing of Unknown Kanji Words

Joy Palmer
Center for Language Studies, BYU
Master of Arts

The present study investigates the effect of furigana on lexical inferencing. After completing a pretest to determine their knowledge of the target words, participants read a passage and completed a think-aloud protocol and questionnaire. The experimental group read a passage with furigana over all kanji words while the control group read a passage without furigana. The protocols were evaluated to determine the rate and quality of lexical inferences of 16 target kanji words. The results of the questionnaire were evaluated to determine participant perception of passage and kanji difficulty, self-assessed percentage of the story that was understood, and the degree to which they liked the story. It was found that the group with furigana made more correct inferences than the control group. It was also found that the furigana group perceived the kanji in the passage to be easier than the control group did. Furigana did not seem to affect the degree to which the participants liked the story, their perception of the difficulty of the story or the percentage of the story that they understood. Implications for theory and pedagogy are discussed.

Keywords: Furigana, Inferencing, Kanji, Vocabulary, Japanese
ACKNOWLEDGEMENTS

I wish to express appreciation to the members of my thesis committee, Dr. Warnick, Dr. Watabe, and Dr. Dewey, for their support and guidance throughout the process of writing this thesis. I would also like to thank Dr. Clifford, Dr. Bush, Agnes Welch and the staff of the Center for Language Studies. Their assistance and advice has been invaluable as I have navigated through my graduate program. I am grateful also to my family and friends, who have been endlessly supportive of my endeavors. In particular, I would like to thank Shinsuke Tsuchiya and Yu Chinen for their assistance at several vital stages of the research and writing process.
# Table of Contents

Chapter 1: Introduction ....................................................................................................... 1  
  Statement of Purpose ...................................................................................................... 3  
  Research Questions ......................................................................................................... 4  
  Significance of the Study ................................................................................................. 4  

Chapter 2: Review of the literature ..................................................................................... 5  
  Reading Models .............................................................................................................. 5  
  Working Memory in Language Comprehension ............................................................ 6  
  The Importance of Word Recognition for Reading ....................................................... 7  
  Word Recognition and Comprehension in L2 ............................................................... 9  
  Phonological Information and Word Recognition ......................................................... 11  
  Logography or Phonography? Framing the Debate ......................................................... 12  
  Logography or Phonography? Evidence from Psychology and Linguistics ................. 15  
  Phonology and Learning to Read Kanji ........................................................................... 21  
  Differences in Learners Based on Orthographic Background ........................................ 22  
  Phonology, Orthography and Japanese Language Teaching ........................................ 24  
  The significance of the phonography/logography debate. ............................................. 24  
  Phonology and Comprehension Processes .................................................................... 25  
  Lexical Inferencing: A Key Strategy ............................................................................. 26  
  Perceptions of Text Difficulty ....................................................................................... 31  

Chapter 3: Method ............................................................................................................ 33  
  Variables ....................................................................................................................... 33  
  Participants ..................................................................................................................... 34  
  Treatment ....................................................................................................................... 37  
  Instruments ..................................................................................................................... 37  
  Decontextualized kanji pretest .................................................................................... 37  
  Think-aloud protocol. ................................................................................................. 38  
  Questionnaire. ............................................................................................................. 41  
  Data Analysis ............................................................................................................... 41  

Chapter 4: Results and Discussion .................................................................................... 42  
  Research Question 1 ..................................................................................................... 42  
  Research Question 2 ..................................................................................................... 48  
  Research Question 3 ..................................................................................................... 55
## Table of Contents

Research Question 4 ................................................................. 56  
Chapter 5: Discussion and Conclusions ....................................... 58  
  Summary of Results ................................................................. 58  
  Limitations of This Research .................................................. 59  
  Implications of the Results for Theory and Pedagogy .................. 63  
  Suggestions for Future Research ............................................. 64  
  Conclusion ............................................................................. 67  
References .................................................................................. 68  
Appendix A: Kanji Pretest .......................................................... 80  
Appendix B: Story Used in the Protocol ....................................... 82  
Appendix C: Background Questionnaire ..................................... 85  
Appendix D: Selected Samples of Protocol Scoring ....................... 86
List of Tables

Table 1: Ages of Participants in Years ................................................................. 34
Table 2: Mean Length of Time Spent in Japan (in Months).................................... 36
Table 3: Number of Correct Answers on the Pretest............................................. 43
Table 4: Number of Inference Opportunities ....................................................... 43
Table 5: Number of Inference Attempts ............................................................ 44
Table 6: Inference Attempts (as a Percentage of Unknown Target Words) .......... 45
Table 7: Pearson Correlations ............................................................................ 46
Table 8: Group and Pretest Score as Predictors of Percentage of Inference Attempts..... 47
Table 9: Percentage of Inference Attempts by Class .......................................... 47
Table 10: Percentage of Inference Attempts by Length of Time in Japan ............. 48
Table 11: Number of Correct Inferences ........................................................... 49
Table 12: Percentage of Correct Inferences (Out of Total Inference Attempts) ....... 49
Table 13: Group and Pretest Score as Predictors of Percentage of Correct Inferences .... 50
Table 14: Number of Partially Correct Inferences .............................................. 51
Table 15: Percentage of Partially Correct Inferences .......................................... 51
Table 16: Number of Incorrect Inferences ......................................................... 51
Table 17: Percentage of Incorrect Inferences ..................................................... 52
Table 18: Analysis of Control Group Responses (All Target Words) ................... 53
Table 19: Analysis of Control Group Responses (Unknown Words Only) .......... 54
Table 20: Perception of Difficulty ..................................................................... 56
Table 21: "Did you like the story?" Likert Scale ................................................. 57
List of Figures

Figure 1: Gender by Group ............................................................................................... 34
Figure 2: Time Spent in Japan by Group .......................................................................... 35
Figure 3: Highest Level Japanese Class Taken by Group ................................................ 37
Figure 4: Number of Correct Answers on the Pretest ....................................................... 42
Figure 5: Summary of Mean Percentages ......................................................................... 44
Figure 6: Summary of Mean Percentages of Correct, Partially Correct and Incorrect ..... 49
Figure 7: Analysis of Control Group Responses (Unknown Words Only) ...................... 54
Chapter 1: Introduction

In 1980, it was noted that reading is one of the most-researched and least-understood fields of education (Clarke, 1980). Though considerable progress has been made in the more than three decades that have passed since that observation, there are still large gaps in our understanding of reading, second language (L2) reading in particular.

Part of the challenge of gaining a full understanding of reading is the fact that comprehension processes, such as those involved in reading, take place within the brain where they cannot be observed directly. Thus, reading research necessarily must be grounded in a theory of comprehension processes that attempts to explain what cannot be observed directly. The present study is grounded in the theory of language comprehension and working memory as put forth by Baddeley and Hitch (1974), Gathercole and Baddeley (1993), and Just and Carpenter (1987). This theory was developed based on data from first language experiments and observation.

Much of the existent L2 reading research draws directly on L1 reading research. While there is evidence of similarities between L1 and L2 reading, it is essential that L2 reading be investigated in its own right in order to determine the extent to which L1 research is applicable to L2 reading (Koda, 2005). Furthermore, reading research, in both L1 and L2, has focused largely on English (Share, 2008). A more complete understanding of the processes of L2 reading can be gained only through research into L2 reading in a variety of languages.

One way to approach L2 reading research is to consider the challenges presented to a reader when reading in their L2. This type of research can lead to a better understanding of how learners deal with those challenges so that improved instructional methods or materials can be developed. In this regard, the investigation of inferencing (at both the text and lexical level) and
the relationship inferencing has to reading comprehension and incidental learning has been a fruitful strain of research in recent years. The types of strategies used by students in lexical inferencing and the relationship between lexical inferencing and the text environment have been of particular interest (for example see Bensoussan & Laufer, 1984; Nassaji, 2003a; Qian, 2004). It is hoped that by determining what learners do when they encounter an unknown word, we can use that knowledge to show them how to take better advantage of the context to aid comprehension and learning when they come across new vocabulary.

Among the many studies dealing with lexical inferencing, Kondo-Brown’s study (2006a) is of particular interest because it expands the research field by focusing on L2 Japanese, a language very different from English. Japanese poses an interesting challenge to English-speaking learners because of its unique orthographic system. As such, it is an excellent candidate for L2 reading research to expand our understanding of L2 reading in languages other than English. In her study, Kondo-Brown identified the clues that learners use to guess meaning when they encounter an unknown word. She found that phonological accessibility, meaning whether or not a learner can pronounce a word, appears to be related to how often and how accurately learners guess new word meanings, with phonologically accessible words guessed more often and more accurately.

Phonological accessibility is an issue with Japanese because of its unique orthographic system. Japanese is written with a combination of two syllabaries (hiragana and katakana) and Chinese characters (kanji). Though the syllabaries are basically phonetic, kanji are generally referred to as logographic. There are certain conventions dictating when each writing system is used in standard written Japanese. Kanji is used for most content words, such as nouns and verb stems, while hiragana is used for function words, such as particles and verb inflections, as well
as some content words. Katakana is used for non-Chinese loan words, scientific terms and for emphasis (much like the use of italics in English). In addition to these uses, hiragana can also be used as a sort of phonetic notation for kanji characters. Small hiragana characters, called furigana or ruby, are sometimes used to tell a reader how a character should be read in context. They are printed to the right of kanji characters in vertical writing, or above kanji when the passage is written horizontally. They often accompany kanji for personal names, uncommon characters or unusual readings. They are also commonly used in Japanese children’s books for kanji characters that children have not yet studied in school.

While learning to read in any foreign language can be difficult and time-consuming, the combination of several different writing systems can make learning to read in Japanese a particularly challenging endeavor. Since L2 readers are often presented with texts that contain unknown vocabulary, the ability to guess the meaning of unknown words is an important skill and an often-studied subject of study in L2 reading research. With the added complication of phonologically opaque kanji, lexical inferencing in Japanese deserves special attention in order to understand its processes.

Statement of Purpose

The present study builds on Kondo-Brown’s (2006a) observations by investigating the role of phonological transparency in lexical inferencing. It is the first research of its kind to directly address the relationship between phonological transparency and lexical inferencing by using an experimental design with a treatment and control group. The study addresses the following questions.
Research Questions

1. Does the presence of furigana affect how often participants attempt to infer the meaning of unknown kanji words?
2. Is there a difference in the quality of inferences when furigana is present?
3. How does the presence of furigana affect a learner’s perception of text difficulty?
4. How does the presence of furigana affect the extent to which a learner likes a text?

Significance of the Study

This line of research seeks to have a direct impact on Japanese L2 reading pedagogy in two ways. First, an improved understanding of how the text environment, including the presence of furigana, affects lexical inferencing may lead to better training of readers in how to best utilize the text environment. Second, a better understanding of how phonological transparency and lexical inferencing are related may help future materials developers design Japanese reading materials for optimal incidental learning. It is also hoped that the present study will shed light on the cognitive mechanisms involved in reading processes.
Chapter 2: Review of the literature

The current study investigates the process of reading Japanese as a second language, specifically the effect of phonological transparency on lexical inferencing and perceptions of difficulty. As such, in order to put the research in the context of the field, this literature review will focus on the following areas: reading models, the role of working memory in language comprehension, phonology and word recognition, lexical inferencing and how orthography affects learner perceptions and attitudes.

Reading Models

Models of reading traditionally fall into three categories: bottom-up, top-down and interactive (Anderson, 1999; Grabe, 2009). An extreme depiction of a bottom-up model of reading would describe reading as a process of converting text into information by decoding as one moves from symbol to symbol along a line of text. In the bottom-up reading model, the locus of meaning is in the text and the reader’s job is to extract that meaning through linguistic decoding. Decoding, in this sense, refers to the ability to use orthographic information to identify the word being represented by the written form. This process is also called word identification. Little attention is paid to reader factors (such as background knowledge) other than decoding skills and general language comprehension ability (Flesch, 1955; Hoover & Gough, 1990). By contrast, an extreme top-down view would hold that comprehension is directed by the reader who has a set of expectations, based on his or her background knowledge, and samples enough information to confirm those expectations or reject them and form new ones (Carrell & Eisterhold, 1983; Goodman, 1967). In an interactive model, top-down and bottom-up processes interact to facilitate comprehension of a text. In such a view, word-level and text-level processes interact to augment and enhance each other in both L1 (Rumelhart, 2004) and L2
(Carrell, 1988) reading. Furthermore, readers deficient in either top-down or bottom-up skills often attempt to use the other set of skills to compensate for their reading weaknesses (Stanovich, 1980). The present study is grounded in an interactive model of reading comprehension, recognizing that in a task such as lexical inferencing, both bottom-up skills, such as word recognition, and top-down processing skills, such as the ability to infer from context, have an important role to play.

**Working Memory in Language Comprehension**

It is not enough to simply state what information one thinks is paramount in reading comprehension, as the reading models above do. It is also necessary to discuss the cognitive processes that make reading comprehension possible. The current study is based upon a theoretical model of these cognitive processes put forth by Baddeley and Hitch (1974), as described by Gathercole and Baddeley (1993) and further elaborated by Baddeley (2000). This model is based on a division of the memory into long-term memory where information is stored and a separate working memory (sometimes called short-term memory) where information is processed prior to potential transfer to long-term memory. While not all psychologists subscribe to the idea of a separate working memory, the concept is widely accepted and provides a plausible explanation for much of what can be observed about human memory (Logie, 1994).

In particular, working memory is a useful construct for understanding both listening and reading comprehension. At its core, reading comprehension is similar to listening comprehension in that a listener and a reader must both take a sequence of information (sounds on the one hand and visual symbols on the other) and transform it into a meaningful message. In order to do this, the listener or reader must temporarily store information so that it can be related to subsequent information. Working memory does just this. Thus, working memory plays an important role in
the comprehension and production of both written and spoken language (Carpenter & Just, 1989; Gathercole & Baddeley, 1993).

Baddeley and Hitch (1974) originally posited a working memory system with three distinct components: the central executive, the phonological loop, and the visuo-spatial sketchpad. The central executive has many functions including regulatory ones. It is responsible for managing the activity within working memory and retrieving information from long-term memory. It has a limited capacity. The phonological loop consists of a phonological store which stores information in a phonological code and articulatory rehearsal (sometimes called subvocal rehearsal). The articulatory rehearsal process has two functions. First, it can be used to recode non-phonological information (such as printed words). Second, it is used to maintain the memory of information in the phonological store which would otherwise decay over time (Gathercole & Baddeley, 1993; Logie, 1994). The visuo-spatial sketchpad processes and stores visual and spatial information as well as mental images that are created from verbal messages. However, the visuo-spatial sketchpad does not seem to be an important part of language comprehension (Gathercole & Baddeley, 1993). Baddeley (2000) later added the episodic buffer, which is seen as temporarily storing and integrating information from the sub-systems, to better account for the human ability to combine information from the phonological loop and visuo-spatial sketchpad to create a coherent perception of the world.

**The Importance of Word Recognition for Reading**

In order for working memory to perform its function of storing and integrating sequential information for comprehension, the mind must be able to extract meaning from the text units with which it is presented. In other words, the meaning of the words on a page must somehow be retrieved from the reader’s mental lexicon. This process is referred to as word recognition. It
consists of two processes: word encoding and lexical access (Just & Carpenter, 1987). In normal reading, these processes occur so rapidly that they seem indistinguishable. However, it is possible to describe situations where only one of the two takes place. For example, the ‘tip-of-the-tongue’ effect can be understood as an example of lexical access without word encoding. The speaker knows there is a word with a certain lexical meaning but for whatever reason is unable to retrieve it. The opposite condition (word encoding without lexical access) happens when a reader encounters an unknown word and forms a representation of the word in the mind (such as when a reader sounds out an unfamiliar word) but is not able to connect that representation to a lexical meaning (Just & Carpenter, 1987). Word recognition is sometimes taken to mean only the word encoding process (Grabe, 2009) but here it will be used in the full sense of word encoding and lexical access as described by Just and Carpenter (1987).

During the 1960s, 1970s and early 1980s, when top-down models of reading prevailed, little research was done on word recognition because many researchers (see, for example, Goodman, 1967) believed that the reader needed only to sample enough information from the text to confirm his or her hypotheses. This model proposes that the reader does not actually need to recognize the majority of the words on a page, rendering word recognition skills secondary to background knowledge and other extra-textual factors for reading comprehension. This began to change in the 1980s and in recent years numerous studies on both L1 reading (Adams, 1999; Juel, 1988; Perfetti, 1999, 2007; Perfetti, Landi, & Oakhill, 2005; Pressley, 2006; Stanovich, 2000) and L2 reading (Koda, 1996, 2005) indicating that skilled readers process almost every content word in a text and that word recognition plays a significant role in reading comprehension.
Word Recognition and Comprehension in L2

The renewed interest in word recognition continues to this day as studies investigate various aspects of word recognition in both L1 and L2. Of particular interest for the current research are the studies that deal with word recognition and comprehension in L2.

Nassaji (2003b) evaluated a group of 60 advanced learners of L2 English on measures of reading comprehension, word recognition, and syntactic, semantic, phonological and orthographic processing skills to see which skills could help identify the less-proficient readers within the group. In addition to syntactic and semantic processes, which are higher-level comprehension processes and have been shown to distinguish between skilled and less-skilled readers, word recognition, including orthographic and phonological processing, was also a reliable indicator of reading ability, even within this advanced group. Nassaji interpreted his findings to indicate that lower-level processes such as word recognition are important even for advanced L2 readers and should not be neglected in advanced L2 instruction.

Van Geldern et al. (2004) used data from a longitudinal study of the Dutch (L1) and English (L2) reading and writing skills of 397 Dutch adolescents to attempt to create models of L1 and L2 reading ability. They found that L2 word recognition speed (as measured by a lexical decision task) was highly correlated with L2 reading comprehension. However, when other linguistic knowledge factors (in particular vocabulary knowledge) were taken into account, word recognition speed did not make a unique contribution to the regression model explaining reading comprehension. Van Geldern et al. interpreted this to mean that individual differences in skills such as word recognition speed may be interdependent with L2 linguistic knowledge. In a way, this conclusion seems tautological as complete word recognition, including lexical access, cannot take place if a word’s meaning is unknown to the reader. Therefore, vocabulary knowledge and
word recognition skills are bound to be closely related, as later researchers have found (see below).

A plausible extension of the fact that word recognition is a requirement for reading comprehension is the idea that improvements in word recognition skills could lead to improved comprehension. Fukkink, Hulstijn and Simis (2005) tested this hypothesis with a group of L2 learners of English. They found that training in word recognition skills did improve word recognition speed but this improvement did not lead to increased reading comprehension. They concluded that, while it is apparent that some unspecified level of word recognition skills are required for comprehension to take place, there is no reason to suppose a strong causal link between faster word recognition speed and higher-level reading comprehension after a learner has reached that threshold, due to the complexity of higher-level comprehension (Fukkink et al., 2005).

Pasquarella, Gottardo and Grant (2012) compared the relative contribution of decoding skills (which by their definition includes word recognition) and vocabulary knowledge to reading comprehension. Their subjects were a group of English L1 adolescents and a group of English L2 adolescents from a variety of L1 backgrounds. In their factorial analysis, they found that decoding skills and vocabulary knowledge were highly inter-related factors and could perhaps both be reflections of a more fundamental general English proficiency factor affecting reading comprehension. Decoding, vocabulary knowledge and the product of those two factors were all shown to be significant factors in reading comprehension (Pasquarella et al., 2012). These results fit well with the Simple View of reading, a paradigm explaining reading comprehension originally developed for L1. The Simple View of reading states that reading comprehension can be explained in terms of decoding skills and listening comprehension (Grabe, 2009).
These recent studies serve as important reminders that word recognition is a necessary but not sufficient skill for reading comprehension in L2. Learners must also have a sufficient level of L2 ability, which often develops in tandem with word recognition skills, in order for written material to be comprehended (Pasquarella et al., 2012).

**Phonological Information and Word Recognition**

While the studies mentioned above make it clear that word recognition is an important part of reading comprehension, they do not tell us how word recognition itself is accomplished. This has been a fruitful area for psycholinguistic research and decades of research have culminated in two competing paradigms of word recognition (at least for English): the Dual Route model and connectionist models. Both paradigms posit two routes for word recognition: orthographic and phonological. In the orthographic route, the written word directly triggers lexical access. In the phonological route, the orthographic form is first transformed into a speech code which is then used to retrieve the word meaning from memory. In simple terms, the Dual Route model proposes that these two word recognition routes “race” each other with both processes being carried out simultaneously. The connectionist models propose that phonological, orthographic and semantic sources of information all interact to lead to word recognition (Grabe, 2009). The speech code generated by the phonological route is referred to as the pre-lexical speech code because it is generated prior to lexical access (Just & Carpenter, 1987).

The phonological route to word recognition is particularly important for beginning readers. Phonological awareness has been shown to be correlated with later reading proficiency for children learning to read in English as an L1 (Ehri et al., 2001) and is generally thought to have a significant role in reading development for both L1 and L2 readers (Haigh, Savage, Erdos, & Genesee, 2011). Phonological awareness and the ability to translate orthographic forms to
phonographic representations has been termed the sine qua non of reading acquisition (Share, 1995). Phonological skills are valuable not just for beginning readers. Increased phonological activity is also seen in skilled readers tackling challenging reading tasks (Just & Carpenter, 1987).

**Logography or Phonography? Framing the Debate**

The vast majority of studies of reading in general and word recognition in particular, included those mentioned above, have been conducted using English as an L1 or L2 (Share, 2008). Such an emphasis on one particular language invariably raises questions as to the applicability of the results to typologically and orthographically different languages. Japanese and Chinese, because they are different from English orthographically (and in the case of Japanese, typologically different as well), create an interesting opportunity to test theories originally developed based on data from English.

According to the Orthographic Depth Hypothesis, languages that have a more regular relationship between the sounds and symbols of the language, such as Finnish, are characterized as shallow, whereas languages with a more irregular relationship between the sounds of the language and the symbols used to express them are described as deep orthographies. Among alphabetic orthographies, English is one of the deepest (Grabe, 2009; Koda, 2005; Share, 2008). In fact, some researchers have expressed concern that word processing research has focused so much on a language that, compared to other alphabetic orthographies, is an “outlier” with a much deeper orthography than most other alphabetic languages (Share, 2008). For teachers and researchers of Chinese and Japanese, however, the focus on English in research is regrettable for another reason, namely, that English is a much shallower orthography than either Chinese hanzi or Japanese kanji.
To better understand the differences between the two types of orthographies, it is useful to consider what each orthography is striving to represent. Whereas English is an alphabetic orthography with orthographic representation at the level of the phoneme (with many irregularities, that is, exceptions to letter-phoneme correspondences, leading to its designation as a deep orthography), Chinese hanzi and Japanese kanji (and, it can be assumed, by extension, Korean hanja) are often characterized as logographic, with representation taking place at the level of the morpheme or word, making phonology irrelevant (or secondary) to orthographic information in the word recognition process (Chen, Yamauchi, Tamaoka, & Vaid, 2007; Hansen, 1993; Shen & Forster, 1999; Wong & Chen, 1999). Others object quite strenuously to this definition and insist that the evidence supports the view that hanzi, kanji and hanja (as the current study investigates Japanese reading, all three will hereafter be subsumed under the term kanji) primarily represent the sounds of the languages they are used to represent, rather than meanings, and that word access via the phonological route is highly likely, or even obligatory (Matsunaga, 1995, 1996; Perfetti, 2003; Unger, 1993, 2004).

The question of the nature of kanji as primarily ideographic or primarily phonographic is a debate of considerable intensity, and more than a little vitriol, that has far-reaching effects for all scholars interested in the areas of the world where these scripts are used (Lurie, 2006). Because the debate reaches back to the very foundations of linguistic study (i.e. the definition of language itself), I would argue that this is an issue that should be of concern to all scholars with any interest in language whatsoever, and most especially to those studying the pedagogy of East Asian languages.

Analyzing and considering the arguments put forth by both sides of the controversy is a particularly onerous task because it can seem as though the various researchers are writing at
cross-purposes. This is no surprise when you consider that the debate, dating back to the 1930s, is about more than just what to call Chinese characters. It concerns core definitions about language and the legitimacy and boundaries of various academic disciplines. If it seems as though the partisans in the debate are working from very different epistemological frameworks, it is because they are. In the words of one scholar, the issue is a “disciplinary collision” such as one can expect in a field, such as Asian Studies, that includes such wide ranging disciplines as sociology, economics, art history and literature (Lurie, 2006).

One of the most prominent critiques, since Creel (1936), of the idea that kanji primarily represent sound comes from Hansen (1993). His position has been described as a critique of linguistic scholarship from the viewpoint of philosophy (Lurie, 2006). The responses to his argument (see, for example, Matsunaga (1996) and Unger (1993)) are even more pronounced in their disciplinarity. It would seem that the arguments over the nature of kanji are just one battle in the “struggle for authority over language” between linguistics and other disciplines in both the humanities and social sciences (Lurie, 2006).

As the theoretical framework for the current study is based on psychological and linguistic theories of language processing, I will not review the historical, philosophical or other arguments put forth by both sides of the logography/phonography debate that are not related to psychology or linguistics. Suffice it to say that if the proponents of logography were correct in their assumptions, there would be little reason to suspect that phonological support, such as the kind used in this study, would have any impact on lexical inferencing or comprehension. DeFrancis (1984), Hansen (1993) or Unger (2004) are helpful references for those who would like to read further on other aspects of the debate.
Logography or Phonography? Evidence from Psychology and Linguistics

Proponents of the idea that Chinese and Japanese characters are primarily a phonographic orthography, as opposed to a logography, cite several strains of research investigating how characters are processed. Particularly in the realm of phonological processing, there is ample evidence for similarities across orthographies and perhaps even some universal principles undergirding written language acquisition regardless of orthography (Perfetti, 2003). As mentioned previously, current word recognition theory, primarily developed based on experiments with English, recognizes both the phonological route to word recognition and orthographic route to word recognition. Various theories differ in respect to the relative importance they place on the two routes but all major theories recognize that both phonological and orthographic information play a part in word recognition. In contrast, however, some scholars have proposed a “direct access” theory of word recognition for kanji characters based on the supposition that characters provide a more direct access to meaning than an alphabetic orthography does (Shen & Forster, 1999; Wong & Chen, 1999).

Shen and Forster (1999) carried out two experiments using masked priming. In the first experiment, a naming task, subjects were shown a masking pattern followed by a prime for a very brief amount of time (50 milliseconds). The masking pattern was a visual pattern presented before the prime in order to make the subsequent prime less visible to the subjects. They were then shown a target character which they named out loud. Shen and Forster observed that the naming was facilitated by both phonological and orthographic primes. None of the subjects reported seeing the prime before the target, minimizing the possibility that naming of the prime was responsible for the priming effects. This result provides evidence that both phonological and orthographic routes to word recognition operate in Chinese (Shen & Forster, 1999).
The second experiment was a lexical decision task. After being shown a prime for a brief amount of time (no subjects reported seeing the prime), subjects were shown a target character and decided whether the character was a real Chinese character or not. In this task, only orthographic, and not phonological priming was observed. Shen and Forster interpreted their results to show that phonological priming was a result only of the phonological nature of the task in the first experiment and that phonology does not play a role in lexical decisions, unlike in alphabetic languages (Shen & Forster, 1999).

Wong and Chen (1999) approached the question from a different angle by using eye movement data to detect disrupted reading caused by an altered character in a reading passage (Wong & Chen, 1999). A key character in a brief (40 to 60 character) reading passage was replaced with an orthographically similar homophone character, an orthographically dissimilar homophone character, an orthographically similar but phonologically dissimilar character or an orthographically and phonologically dissimilar character. They found that orthographic dissimilarity caused early disruption in reading, while phonological dissimilarity caused much weaker disruption and caused it at a later time than orthographic dissimilarity did. Their results support the direct access hypothesis for Chinese reading.

However, other researchers have come to a decidedly different conclusion about word recognition in Chinese. Xu, Pollatsek and Potter (1999) investigated phonological and orthographic similarity effects in a semantic judgment task. Their subjects, 15 native Chinese-speaking adults, were briefly shown a character and were then shown another character that was either semantically related or was a distractor. The distractors were orthographically similar homophones, orthographically dissimilar homophones, orthographically similar non-homophones or orthographically dissimilar non-homophones to the target characters. The
subjects were instructed to judge quickly and accurately whether or not the second character shown to them was semantically related to the first.

Their results showed that there was interference from both the orthographically similar and the orthographically dissimilar homophones and this interference was statistically significant. However, they also noted that there was even more interference from orthographically similar non-homophones than from unrelated characters or orthographically dissimilar homophones. They compared their results with previous results from English studies and concluded that the relatively high number of non-homophone orthographic errors shows that phonological activation is faster in English than Chinese, due to the relative phonological transparency of the Chinese script, but that there is no reason to suppose a different method of access to word meaning than is used to explain English word recognition (Xu et al., 1999).

Spinks, Liu, Perfetti and Tan (2000) conducted experiments using the Stroop paradigm to investigate phonological activation in Chinese. They used color characters, orthographically dissimilar homophones and semantic associates of the target color characters. They found that homophones create Stroop interference, a finding that is consistent with either a dual route or connectionist model of word recognition but inconsistent with a direct access model. Guo, Peng and Liu (2005) found similar results with their Stroop test involving Chinese speakers ranging in age from 7 to 23. They further found that lower-level readers exhibit stronger Stroop effects, indicative of an important role for phonology in early Chinese reading. In a similar vein, Lee and Chan (2000) conducted Stroop tests with a group of Chinese speakers and a group of English speakers and found similar patterns of interference for both groups. They concluded that word recognition processes are the same in both languages.
Japanese offers an even more interesting opportunity for researching the relationship between phonology and orthography. Unlike Chinese, where most characters have one only pronunciation (in a given dialect) and each character corresponds to a syllable, Japanese kanji usually have more than one pronunciation and these pronunciations are often polysyllabic. Japanese also uses an orthographically shallow syllabary which allows for interesting comparisons between orthographies of different depths within the same language (Kess & Miyamoto, 1999). In spite of this, there are relatively few studies (compared to Chinese) addressing phonology and orthography in word recognition in Japanese.

All but one of the studies reviewed in preparation for the current study that specifically investigated Japanese word recognition indicated that kanji word recognition processes can make use of both phonological and orthographic information. Chen, Yamauchi, Tamaoka, and Vaid (2007), however, performed an experiment with lexical decision tasks in which native Japanese readers were shown primes that were homophones with the target kanji, semantically related non-homophones or semantically and phonologically unrelated primes. In the first experiment, the subjects were shown the primes as kanji. In the second experiment, they were shown hiragana transcriptions of the primes. They reported that there was no homophonic priming effect in the kanji condition, though there was in the hiragana condition. (Chen et al., 2007)

On the opposite end of the spectrum, Flaherty (1993) proposed that phonological information is triggered before semantic information in reading kanji. She compared the performance of Japanese and English speaking adults on a word and photograph categorization task. They found identical patterns of naming and categorization speed for both the English and Japanese groups, regardless of whether or not the words shown the Japanese group were in kanji or kana. Both groups named words faster than pictures but were able to categorize the pictures
faster. This would seem to indicate that a pre-lexical route was used for both Japanese and English readers in this task.

Flores D’Arcais, Saito and Kawakami (1995) also used a priming task, but came to the opposite conclusion about the importance of phonological and semantic information in kanji. They found that showing a character’s phonetic radical as a prime facilitated kanji naming much more than showing the semantic radical did. It is important to note, however, that this effect could be due to the nature of the task, as the task may affect which processing skills are used (Shen & Forster, 1999). This type of experiment should be replicated with a task that does not require naming in order to increase the strength of its argument.

Matsunaga (1995) used an eye-tracking study to investigate the interference caused by homophonic and non-homophonic kanji errors. She hypothesized that non-homophone errors would cause more interference in reading than homophone errors, indicating that phonological information is an important part of word access when reading kanji. Her results supported this hypothesis and she concluded that kanji primarily triggers phonological information. Matsunaga and Crosby (1997) also used eye-tracking technology to determine whether or not there is a relationship between phonology and spatial ability in reading kanji. If kanji are primarily processed orthographically, a Japanese reader who was not spatially adept might make more use of phonology when reading. However, those with good spatial skills should use phonological information less when reading. The results of the study, however, suggested that phonological coding played an important role in reading kanji, regardless of spatial orientation.

Many studies on word recognition in kanji in recent years have come to the conclusion that both phonology and orthography are possible pathways to word recognition in Japanese though some emphasize one route over the other. The results of Wydell, Patterson and
Humphrey’s (1993) semantic categorization experiments and Wydell’s (1998) naming experiments suggest that both orthographic and phonological information play a role in kanji word recognition, particularly at the word (as opposed to the character) level. Sakuma, Sasanuma, Tatsumi and Masaki (1998) replicated Van Orden’s (1987) seminal study using kanji. The original study demonstrated that interference effects in lexical decision tasks in English could be caused by both homophonic and orthographic similarity. Sakuma et al. (1998) found these same results using kanji, indicating that both phonological and orthographic information play a role in kanji word identification much as they do in English.

In a similar vein, Morita and Tamaoka (2001) also used a lexical decision task and found phonological interference indicating that phonological information was activated when reading kanji. At the same time, however, they did not find a homophone effect when subjects were given a proofreading task, suggesting that phonological information did not play a primary role in proofreading. In a later study, Tamaoka (2005) again found evidence that phonological information is automatically activated when reading kanji words.

Morita and Saito (2007) used a synonym judgment task and a semantic decision task to investigate the role of phonological processing in kanji word recognition. Specifically, they wanted to investigate the role of articulatory suppression in these tasks. Articulatory suppression is a technique whereby subjects repeat an unrelated sound (such as “the”) while performing a linguistic task. Articulatory suppression is thought to suppress some kinds of phonological processes but not others. They found patterns of interference that indicate phonological processing mechanisms similar to those previously reported for English (Morita & Saito, 2007).
Phonology and Learning to Read Kanji

In recent years it appears that the gap between the two sides of the debate is narrowing, as most researchers recognize the possibility of lexical access proceeding from both orthographic and phonological information. Even for alphabetic orthographies there is still discussion over the relative importance and timing of phonological and orthographic word recognition processes in fluent readers (Daneman, Reingold, & Davidson, 1995; Grabe, 2009; Taft & van Graan, 1998). In addition to the question of the phonological route to kanji word recognition, another salient question, from a Japanese pedagogical standpoint, is how characters are learned and processed by beginning readers. Compared to the data on fluent readers, relatively little research has been done in this area. Most of the previous research in this area has looked at Chinese children but a few studies have also investigated Japanese children, Chinese L2 learners and Japanese L2 learners.

Chow, McBride-Chang and Burgess (2005) investigated the relationship between L1 Chinese children’s Chinese phonological skills and their reading ability in both their L1 and L2 English. They tested their subjects on three different measures of phonological awareness, rapid automatized naming, and short-term verbal memory. Their results showed these skills were associated with reading abilities in Chinese. Even controlling for visual skills, phonological awareness explained significant variance. This supports the idea that phonological awareness is important for learning to read across orthographies, not just in alphabetic languages. Furthermore, the results showed a positive relationship between phonological awareness in Chinese (L1) and concurrent and subsequent English (L2) reading abilities, demonstrating that phonological skills may transfer even to languages with very different characteristics. This further suggests the universality of phonological processing skills for reading acquisition.
In another experiment with Chinese schoolchildren, He, Wang and Anderson (2005) demonstrated that the children made use of phonological information within characters to help them remember new characters. This was done even when characters where irregular (that is the phonological information provided within the character did not perfectly match the actual pronunciation of the character). Ehri (1991) found similar results for U.S. children utilizing spelling information in irregular words (as cited in He et al., 2005).

Tamaoka, Leong, and Hatta (1992) used an articulatory suppression paradigm to investigate phonological processing when making semantic decisions. They found that, regardless of script type, articulatory suppression interfered with semantic decisions, indicating that phonological recoding takes place when making semantic decisions at the sentence level. They pointed out that this could have to do with the need to keep information in working memory via the phonologic code in order for sentence processing to take place. In other words, while the results of their study do not indicate how word recognition takes place when reading kanji words, they do suggest that the phonologic code could be necessary for other reasons, an observation that will be discussed further below (Tamaoka et al., 1992).

**Differences in Learners Based on Orthographic Background**

Some researchers have investigated the ways in which reading processes differ among various learners. This research has revealed differences between learners from Chinese L1 backgrounds and English L1 backgrounds that could be a result of dominant word recognition or text processing modes used by those learners in their L1.

Chikamatsu (1996) studied the response times of Chinese L1 and English L1 learners of L2 Japanese on a lexical decision task. The subjects were required to distinguish between words and non-words written in hiragana and katakana. Some of the words normally encountered in
katakana were presented in hiragana and some of the words normally encountered in hiragana were presented in katakana. English L1 and Chinese L1 subjects displayed different patterns of response time latencies, indicating that the Chinese L1 subjects used more visual information when reading kana while the English L1 learners used more phonological information. The results seem to indicate that L1 background does affect word recognition processes in L2 and that readers from so-called logographic scripts rely on visual information more while those from alphabetic backgrounds rely on phonological information more.

Two other studies of English L1 learners have also indicated the importance of phonology for English-speaking learners when reading characters. In a study looking at English L1 learners of Chinese, the ability to pronounce (or read) characters was associated with the ability to give an L1 gloss for the characters (Everson, 1998). In addition, other research has indicated that L1 English learners of Japanese are more likely to remember a kanji character if the radical is pronounceable, whereas phonological accessibility of radicals had no effect on L1 Chinese learners recollection of novel Japanese kanji. This could indicate that the way L1 English learners process kanji is related to the way they process their L1, with phonological processing being dominant (Mori, 1998). It could also simply be a result of Chinese learners having a greater depth of knowledge of the parts that make up characters. This could have allowed them to integrate the new characters into their mental character framework better than the English L1 learners who did not have as much familiarity with characters and thus relied on the one aspect of the characters that was readily apparent to them, namely, the radical with the same shape as a known katakana character.

Of course, variation in the way learners process script may be influenced by other factors in addition to orthographic background. Toyoda and McNamara (2011) found that English L1
Japanese learners showed differences in processing skills depending on their overall kanji ability. Native Japanese readers and Japanese L2 learners of a variety of levels performed a semantic judgment task in which they decided how strongly related two kanji characters were. Some characters shared component parts (radicals) while others did not. The response patterns of advanced learners approximated those of native readers better than lower-level learners did, indicating that, with continued experience, learners begin to process characters more like native readers. Chikamatsu (2006) also found differences between the way lower-level and advanced learners process kanji. The more advanced learners seemed to rely more on visual information than the lower level learners did. However, the differences between the two groups were not statistically significant.

**Phonology, Orthography and Japanese Language Teaching**

As is apparent from the discussion above, it is possible to find research to support either a direct route theory of word recognition, thus supporting the logographic view of characters, or a dual-route (also compatible with connectionist models) of word access for kanji, supporting the idea of kanji as a phonography. Opinions on both sides seem deeply ingrained and resistant to change. Word recognition may or may not require phonological processes for fluent readers (as pointed out previously, this is also under debate for alphabetic languages as well). However, the research regarding beginning readers, especially native English speaking learners, seems to support more strongly the view that phonology is essential for learning to read so-called logographic languages.

**The significance of the phonography/logography debate.** These debates have real consequences for teachers and learners of Japanese. As explained above, there is considerable evidence that phonology plays an important role in reading kanji. At least one method of learning
kanji in popular use today tends to divorce the sound associated with the kanji from its shape. Students are taught how to write the kanji by creating a mental story to go along with the shape. These methods encourage learners to directly associate the form of kanji form with a keyword in the L1 through various mnemonic devices. Kanji are presented in isolation with little or no context and pronunciation is not studied until after the written forms of all common-use characters have been learned (Heisig, 1987). This is particularly concerning because the role of phonological information is not limited just to word recognition. Phonological information also plays an important role in text comprehension.

**Phonology and Comprehension Processes**

In addition to the pre-lexical speech code used in word recognition, it appears that even when word meaning is accessed directly from orthography (as can happen in fluent reading), speech codes are still activated. Speech codes that are activated after word recognition, rather than as a pathway to word recognition are referred to as post-lexical speech codes. Post-lexical speech codes are believed to help keep information in working memory, allowing comprehension to take place. (Johnston, Thompson, Fletcher-Flinn, & Holligan, 1995; Just & Carpenter, 1987). This has been shown to be true among readers of a variety of languages, leading one researcher to call phonological decoding “perhaps the most indispensable competence for reading acquisition in all languages” (Koda, 2005, p. 34).

When reading phonologically opaque scripts, such as Japanese kanji, unknown characters would naturally hinder a reader’s ability to generate a phonological code. Japanese, however, can be written with phonologically transparent kana. In fact, furigana are often used in materials written for children and even in texts for adult readers to clarify difficult words or unusual readings. Many Japanese language educators believe that furigana can facilitate fluent reading of
texts with many unknown kanji, particularly when the text contains words that are already in the reader’s oral vocabulary (Koda, 1992; Nara & Noda, 2003). Since the phonological code has other facilitative effects besides word identification, it is possible that providing the phonological code for unknown words could provide additional benefits to language learners in connection with comprehension processes. The current study aims to investigate this question, specifically in reference to lexical inferencing.

**Lexical Inferencing: A Key Strategy**

The discussion about word recognition above has focused on what a person does to find a word in their mental lexicon. But what does a person do when they do not have a given word already in their mental lexicon?

In normal reading, either in L1 or L2, readers regularly make decisions about which of several possible meanings of a known word are most appropriate for a given context. They also encounter unknown words. When encountering unknown words, learners have several options. They may choose to skip a word they do not know and try to make sense of the text without the word. They may ask another person the meaning or look it up in a dictionary, if such resources are available. Another option is to guess what the word means in that context. The informed guesses that readers make about the meaning of a word in context are called lexical inferences (Haastrup, 1991; Wesche & Paribakht, 2009). Lexical inferencing is asubset of more general inferencing procedures that have been defined as the “connections people make when attempting to reach an interpretation of what they read or hear” (Brown & Yule, 1983, p. 265).

Since the early 1970s, when research on lexical inferencing began, it has been closely connected with research on incidental (non-intentional) vocabulary learning. Research on vocabulary learning in both L1 and L2 has suggested that an important amount of vocabulary is
learned incidentally through exposure to words in context (Bengeleil & Paribakht, 2004; Bensoussan & Laufer, 1984; Everson & Kuriya, 1998; Haastrup, 1991; Harley & Hart, 2000; Krashen, 1989, 1993; Nagy, Herman, & Anderson, 1985; Nassaji, 2003a; Qian, 2004; Shimizu & Green, 2002; Sternberg, 1987; Wesche & Paribakht, 2009). Inferencing is important to discussions of pedagogy because making a correct inference about the meaning of an unknown word encountered in context is often the first step in the process of learning the word. If a learner fails to make an inference or infers incorrectly, it is impossible for the learner to begin the process of correctly learning the new vocabulary item from context (Wesche & Paribakht, 2009).

Before beginning an in-depth discussion of lexical inferencing, it is important to clarify that making the claims that vocabulary can be learned incidentally from context and that inferencing is an important part of this process is not the same as claiming that the best way to teach specific vocabulary is through context (Sternberg, 1987). Critics of the concept of incidental vocabulary learning through reading often go about trying to dispute the claim by comparing learning of specific vocabulary items via reading only versus via direct instruction methods (Laufer, 2003; Sternberg, 1987). Though the most ardent proponents of incidental learning might make the case for dispensing with direct instruction and relying on incidental learning through reading for all vocabulary acquisition (see for example Krashen, 1989), the more widely followed approach focuses on how context affects lexical inferencing and word learning with the end goal of teaching learners how to learn better from context so as to take better advantage of the vocabulary building potential of incidental learning through reading (Sternberg, 1987).

This is especially significant since inferencing has been shown to be a commonly used strategy. Everson and Kuriya (1998) identify at least 21 strategies used by learners of Japanese
when reading. They found that learners used a lexical inferencing strategy (referred to by Everson and Kuriya as “Hypothesizing Word Meaning”) as much as 40% of the time they encountered an unknown word. Harley (2000) found that inferencing was a preferred strategy for high-scoring students in L2 French classes. In Qian’s (2004) study of Chinese and Korean students learning English, inferencing was a commonly reported strategy.

Lexical inferencing itself is not monolithic in nature. Rather, it is accomplished by using various strategies. These strategies include such things as using context (both global and local), using background knowledge, looking for grammatical clues and looking for clues within the word itself. Qian (2004) found that students’ self-reported use of these various lexical inferencing strategies did not match what they actual used when exposed to a reading task. Students tended to report higher use of context-based inferencing and lower use of morphological and syntactic-based inferencing than they demonstrated when observed.

Nassaji (2003a) outlined a variety of these strategies and points out that successful inferencing depends not so much on the number or type of strategies used but rather the quality of the strategy use. In other words, regardless of how many or which strategies a student uses, the information (from the text, from background knowledge, from the word itself) used in inferencing must be accurate and be used appropriately if learners are to be able to correctly guess the meaning of an unknown word.

This is not to say that learners attempt to infer the meaning of all unknown words they encounter. Research shows that learners tend to ignore many unknown words (Bensoussan & Laufer, 1984; Paribakht & Wesche, 1999). This should not come as a surprise when we consider that many contexts simply do not give a reader enough information to guess an unknown word. In addition, reader factors may discourage inferencing. Bengeleil and Paribakht (2004) found
that learners with a higher reading proficiency were more likely to make lexical inferences than lower-level learners. This should come as no surprise either since a better knowledge of the language would mean that a learner has more information on which to base her inferences. In her study of ESL learners’ inferencing, Haynes (1993) pointed out that native English speakers have access to extensive knowledge of word collocations as well as complex syntactic knowledge that they can use in addition to their knowledge of grapheme-meaning and sound-meaning connections. ESL learners often do not have complex collocation and syntactic knowledge and thus must rely on their knowledge of shape-meaning and sound-meaning connections. In the case of learners of Japanese, when confronted with a kanji word with an unknown pronunciation, they are limited to only their knowledge of shape-meaning connections.

Looking specifically at Japanese, Mori and Nagy (1999) and Mori (2003) investigated the contribution of morphological information and context to the lexical inferencing of unknown kanji words. Both studies investigated how Japanese learners used both morphological and contextual clues in guessing unknown two-kanji compound words. The unknown words used as targets in both studies were semantically semi-transparent. That is, the meanings of the target words were semantically related to the meanings of the individual kanji that made up the word, but only in a general way. An example of a semantically semi-transparent word used by Mori and Nagy is 月食 (lunar eclipse) consisting of the characters 月 (moon) and 食 (eat). An example of a semantically opaque word is 皮肉 (sarcasm), which is semantically unrelated to either of its component characters 皮 (skin) or 肉 (meat). By contrast, 月光 (moonlight) is a semantically transparent word whose meaning can be derived directly from the meanings of its component characters 月 (moon) and 光 (light). The meanings of the semantically semi-transparent target words could not necessarily be deduced from the individual meanings of the
component kanji alone, though it is possible that knowledge of the meanings of the component kanji could help steer the reader in the right direction for inferencing. Both studies found that morphological analysis (applying the meanings of the individual kanji in a target word to the lexical inferencing) and context use were separate strategies whose use was not correlated. Some participants were better at using one strategy than the other and higher students proficiency (as determined by course exam and standardized test scores) were better able to use context successfully (Mori & Nagy, 1999). Additionally, combining the clues from context and morphology resulted in more accurate inferencing than using only one strategy, further strengthening the position that context and morphology provide different information to the reader (Mori, 2003). These results taken together suggest that learners might benefit from practice using both context and morphological information to make lexical inferences.

Kondo-Brown (2006a) investigated whether or not context plays a role in English L1 students’ ability to infer the meanings of kanji vocabulary and determined that it did indeed seem to be an important factor in learners’ inferencing. She also observed that complete or partial phonological knowledge of a kanji vocabulary word was positively correlated with correct inferencing of meaning. Interestingly, when the subject had no phonological knowledge of the kanji word, they were most likely to not guess at all.

Despite the importance of inferencing to the process of reading in an L2 and the potential facilitative effects of furigana for Japanese reading instruction, no research has yet looked directly at the link between the two. This gap in our existing knowledge gave rise to the following research questions:

1. Does the presence of furigana affect how often participants attempt to infer the meaning of unknown kanji words?
2. Is there a difference in the quality of inferences when furigana is present?

**Perceptions of Text Difficulty**

In addition to affecting inferencing, it is also possible that rendering the entire text pronounceable could also have an effect on learners’ perceptions of text difficulty. There has been little research investigating the effect of script choice on perceptions of difficulty or affect in general.

Dewey (2004) found that learners’ views regarding the choice of script in their textbooks correlated strongly with their teacher’s views. However, Dewey was specifically investigating textbooks using Romanization and textbooks using kana or a kana/kanji mix. His study does not refer to the use of furigana as a factor influencing student or teacher views.

Kirwan (2005) investigated the effect of a textbook using furigana on a group of Australian high school students. His study looked at both incidental learning and affect. Kirwan found that after one semester of Japanese in high school, where students were required to learn hiragana and katakana but did not do any explicit study of kanji, his students using a textbook with furigana learned the readings of 7.8% and the English meanings of 7.1% of the kanji used in the textbook. As Kirwan points out, this amount is impressive when one considers that no extra effort was expended on the part of the teacher (other than providing the furigana textbook) or the students to achieve this extra learning. Kirwan also found that students preferred not to have furigana over all kanji words. In particular, intermediate and advanced learners did not want furigana over kanji that they had previously been taught or simple kanji that appeared multiple times in a text. Kirwan did not address student perceptions of text difficulty as related to furigana use. These results, though not generalizable because the study was conducted with an intact class with no control group, are an interesting starting point for further research in this area.
Kondo-Brown (2006b) found that self-perception of reading ability and perceived difficulty in reading kanji were correlated with reading comprehension and kanji knowledge. Mori, Sato and Shimizu (2007) had similar findings. They found that student belief that kanji learning required special abilities was negatively correlated with some of the measures of kanji ability used in their study, such as the ability to contextually appropriate sounds and radical awareness.

In light of these results we can see that student perceptions of difficulty are an important factor to consider in reading ability and kanji achievement. The third and fourth research questions will investigate the relationship between furigana use and perceptions of text difficulty.

3. How does the presence of furigana affect a learner’s perception of text difficulty?

4. How does the presence of furigana affect the extent to which a learner likes a text?
Chapter 3: Method

This study employs elements of Kondo-Brown’s (2006a) methodology for investigating inferences made in context. Specifically, participants took a decontextualized kanji test to determine known words. They then completed a think-aloud protocol which was transcribed and analyzed to determine when participants made lexical inferencing attempts and when those attempts were successful. This information, along with a questionnaire completed after the think-aloud activity, provided the needed information to analyze the effect of the independent variable on the dependent variables.

Variables

The independent variable in this study is the presence or absence of furigana. The dependent variable for the first research question is the mean percentage of unknown target words that the participants attempted to infer. The dependent variable for the second question is the mean percentage of unknown target words that the participants correctly inferred. Though the percentages are the primary variables analyzed for the first and second research questions, counts of the inferencing attempts and correct inferences are also included in order to provide a frame of reference for the percentages. The dependent variables for the third question are the Likert scale rating of text difficulty, the Likert scale rating for kanji difficulty and the self-assessed percent understood. The dependent variable for the fourth question is the Likert scale rating for how well the student liked the text. Data for the first and second dependent variable were collected via a vocabulary test and a think-aloud protocol. The data for the third and fourth dependent variables were collected via a questionnaire administered after the think-aloud protocol.
Participants

Participants were volunteers recruited from 300-level Japanese classes at Brigham Young University. They ranged in age from 19 to 24 years old (Table 1). Fifteen were male and seven were female (Figure 1).

Table 1

*Ages of Participants in Years*

<table>
<thead>
<tr>
<th></th>
<th>Control (n=11)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>Maximum</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>1.3</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>22.5</td>
<td>21.3</td>
</tr>
</tbody>
</table>

Figure 1

*Gender by Group*

Participants varied in the length of time they had spent in Japan and the highest level of class they had taken at BYU. As L2 reading proficiency and vocabulary knowledge have been shown to affect lexical inferencing, participants were assigned to the control and experimental (furigana) groups using stratified random group assignment based on the highest level class they had taken and whether or not they had spent more than a year living in Japan.
The control and furigana groups both consisted of nine participants that had spent one year or longer in Japan and two participants that had spent less than one year in Japan (Figure 2). Though one year was originally chosen as the cut-off point, in effect the cut-off was 16 months, since 16 months was the shortest length of time any of the participants in the year-or-longer group had been in Japan (see Figure 2). All of the students who had spent more than one year in Japan went there to serve Japanese-speaking missions for The Church of Jesus Christ of Latter-day Saints. These participants have a language learning background that is unique in comparison to other university students. Missionaries study Japanese for a short time (generally ten weeks) in the U.S. before they go to Japan. They then spend 16 to 22 months living in Japan where they interact with Japanese people on a daily basis and continue to study informally. Since study of the written language is not emphasized, missionaries generally have much larger oral vocabularies than written vocabularies. None of the four participants in the less-than-a-year-group had ever been to Japan. The mean length of time spent in Japan (in months) for the two groups was compared using an independent samples $t$-test (Table 2) and no statistically significant difference was found between the means for the two groups ($df=20$, $t=-.440$, $p=.66$).

Figure 2

*Time Spent in Japan*
Table 2

*Time Spent in Japan in Months*

<table>
<thead>
<tr>
<th></th>
<th>Control (n=11)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>SD</td>
<td>8.8</td>
<td>8.9</td>
</tr>
<tr>
<td>Mean</td>
<td>16.4</td>
<td>18</td>
</tr>
</tbody>
</table>

Participants were asked to report the Japanese class that they were currently enrolled in or the highest numbered class they had taken at BYU, as some 300-level classes do not have to be taken in order of course number. All students reported that the course in which they were currently enrolled was the highest numbered course they had taken. Eleven participants reported that they were taking Japanese 300, an intermediate reading, writing and grammar course that is generally the first course taken by students who have served Japanese-speaking missions. Six participants were in Japanese 302, the next course in sequence after 300, which focuses on reading and writing. Five students were in Japanese 321, which is taken after 302 and focuses on reading authentic documentary-style texts, particularly newspapers. Due to uneven numbers of students from two of the classes, it was not possible to have an identical distribution of participants from each class in the two groups. See Figure 3 for the distribution of participants by class level within the control and experimental groups.
Treatment

Participants first met with the researcher individually to take a vocabulary test to determine if they had previous knowledge of the target kanji compounds (Appendix A). Approximately a week after the pretest, participants met again with the researcher individually to record the think-aloud protocol and complete the questionnaire. Appendix B contains the passage used as the prompt in the think-aloud protocol and Appendix C contains the questionnaire.

Instruments

Decontextualized kanji pretest. The first procedure in the study was the decontextualized kanji pretest. In order to minimize the effects of sensitization to the target compounds, non-target distracters were also included in the assessment. The assessment consisted of 16 target words and 34 non-target words. The reading and meaning for each kanji word were scored separately. Each item was marked as correct, partially correct, incorrect, or not attempted for the reading and correct, incorrect, or not attempted for the meaning. Readings
were marked *partially correct* if the correct reading was given for one of the characters but not for the other. Meanings were marked correct if they matched one of the meanings for the compound given in a commonly used online Japanese dictionary (Electronic Dictionary Research and Development Group, 2011). *Partially correct* was a possible score for the meanings as well. However, no partially correct answers were found.

**Think-aloud protocol.** Approximately one week after taking the vocabulary assessment, participants met with the researcher again for the think-aloud protocol. A one-week time lapse was chosen as it was thought to be long enough to minimize any sensitization effects from the vocabulary assessment but short enough that the possibility of subjects being exposed to the target words in the interim was minimal.

It is acknowledged that think-aloud protocols may influence the data collected by altering comprehension processes, in some cases possibly impeding and in some cases possibly aiding comprehension (Horiba, 1990; Nisbett & Wilson, 1977). However, under certain circumstances, think-aloud protocols may be the most effective means of gathering information about a specific cognitive process. In particular, think-aloud protocols are most useful for studying higher-level thinking skills that are involved in the comprehension of text, as opposed to lower-level processing skills such as recognition of words or letters which, in skilled reading, generally take place too rapidly to be brought to conscious awareness. Higher-level processes such as making inferences or predictions are assumed to be more available to the reader’s conscious, and thus more readily verbalized, because they are slower and outputs of the processes themselves are generally verbal (Olson, Duffy, & Mack, 1984). It is important to note, however, that some researchers have suggested that people may not actually be able to observe their own thinking processes and that attempts to verbalize this process may be only partially successful (Nisbett &
Wilson, 1977). The use of a think-aloud task for this study is modeled after Kondo-Brown's (2006a) use of the procedure for collecting verbal data on lexical inferencing attempts.

Shin’ichi Hoshi’s (1971) short story “Favorite Watch” was adapted for use in the think-aloud protocol. The story is sometimes used in Japanese 322, a course focused on reading modern Japanese short stories in the original Japanese, which is taken after Japanese 302 but may be taken before or after 321. Though none of the participants had yet taken Japanese 322, as a precaution they were asked if they had read the story and none of the participants reported having read the story previously. The adapted version, as used in the experiment, can be found in Appendix B. Sixteen words in the text where chosen as target words. Three of the words were judged to be unlikely to be known by participants. Thirteen of the sixteen were replaced with contextually appropriate synonyms that were assumed to be unknown to the participants. The instrument was then pilot-tested with BYU students with a similar learning background as the participants. The textual changes were made by a native Japanese speaker and the story was read by another native speaker to insure that the changes did not interfere with the original meaning of the story.

Each sentence in the story was printed on a separate piece of paper. The control group received the sentences without furigana. The experimental group received sentences that had furigana over every word. Participants were instructed to read each sentence out loud and then say what they thought the sentence meant in English. This procedure is not a traditional think-aloud in that the participants were not required to comment on their thought processes per se, but rather on their understanding of the meaning of the Japanese passage. Both the specific procedure used in this study and the usage of the term think-aloud protocol to describe the procedure are modeled after Kondo-Brown (2006a). The participants read out loud in order to
show their knowledge of word pronunciation and to facilitate comparison of the results with those found by Kondo-Brown (2006a). Though reading out loud and translating may be considered unnatural, the participants were familiar with the task because it was commonly used in their language classes. Participants appeared to be providing a direct translation of the text, minimizing the possibility that they made mental inferences but did not vocalize them.

Participants had access to the sentences while providing the oral translation. Blank pieces of paper were inserted after every sentence to remind the participant to say the meaning of the sentence in English before moving on. The think-aloud protocols were recorded and later transcribed for analysis. The transcription was done at the word level in accordance with recommendations made by Schiffrin, Tannen and Hamilton (2003). The transcriptions were analyzed by two highly proficient speakers of Japanese. Differences of opinion were resolved through consultation and the raters achieved 100% agreement on the ratings. The meaning and reading for each target word were marked as inferred correctly, inferred partially correctly, inferred incorrectly, or not attempted. To determine whether or not the participant had attempted to give the meaning of the target word, the Japanese sentence from the text and the corresponding English sentence generated by the student were compared. Raters looked for the word or words in the English sentence that corresponded to the word in the Japanese sentence and judged whether or not the English word had the same meaning as the Japanese word. Meanings were marked partially correct if they were not accurate translations of the target word but were very close in meaning and still maintained the essential meaning of the text. If no corresponding word could be found, the meaning was marked as not attempted. Readings were marked in the same manner as the pretest. Samples from the protocols with an explanation of the scoring can be found in Appendix D.
**Questionnaire.** After completing the think aloud protocol, participants completed a questionnaire (Appendix C). The questionnaire had three Likert scale items asking how difficult the participant thought the text was, how difficult they thought the kanji in the text was and how much they liked the text. Participants also wrote what percentage of the text they thought they understood.

**Data Analysis**

The data was analyzed using IBM SPSS statistical software. The means of the following key indicators were compared using independent samples *t*-tests in order to determine if there were statistically significant differences between the means of the two groups: percentage of inferencing attempts, percentage of successful inferences, self-reported percentage of understanding, Likert score for text difficulty, Likert score for kanji difficulty and Likert score for how well the participant liked the text. Levene’s Test of Equality of Variances was conducted and it was determined that equal variances could be assumed for both groups for all indicators.
Chapter 4: Results and Discussion

Research Question 1

The first research question addresses whether or not the rate of inferences is affected by the presence of furigana. To calculate the percentage of inference attempts, the total number of opportunities for inferencing was first calculated for each participant. This was done by subtracting the number of words whose meaning was answered correctly on the pretest from the number of target words (16). This number was used as the denominator when calculating each participant’s percentage of inferencing attempts. Since the denominator used in calculating each participant’s percentage of inference attempts would vary depending on how many of the target words were already known, the distribution of pretest scores was analyzed to determine if the two groups could be considered equivalent in this respect. Figure 4 summarizes the pretest scores of the participants in each group. Descriptive statistics for the mean number of correct answers on the pretest can be found in Table 3.

Figure 4

*Number of Correct Answers on the Pretest*
As the figure illustrates, the Furigana group had more students who knew one or more of the target words prior to the experiment than the control group did. The furigana group also has a higher mean than the control group. The results of a *t*-test did not show a statistically significant difference in the means of the two groups at the \( p < .05 \) \( (t = -1.44, \, df = 20, \, p = .17) \), indicating that difference in the mean number of correct answers between the groups is most likely due to chance.

As indicated in Table 3, the control group had slightly more opportunities to make inferences, because they knew fewer words on the pretest than the furigana group did. The total numbers of inference opportunities (i.e. unknown target words) and inference attempts for each group can be seen in Table 4 and Table 5 respectively. The difference in the number of inference opportunities is not statistically significant at the \( p < .05 \) level \( (t = 1.44, \, df = 20, \, p = .165) \) but the number of actual inference attempts is \( (t = -2.25, \, df = 20, \, p = .036) \).

### Table 3

**Number of Correct Answers on the Pretest**

<table>
<thead>
<tr>
<th></th>
<th>Control ((n=11))</th>
<th>Furigana ((n=11))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (for group)</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>SD</td>
<td>1.25</td>
<td>1.68</td>
</tr>
<tr>
<td>Mean</td>
<td>.82</td>
<td>1.73</td>
</tr>
</tbody>
</table>

### Table 4

**Number of Inference Opportunities**

<table>
<thead>
<tr>
<th></th>
<th>Control ((n=11))</th>
<th>Furigana ((n=11))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>167</td>
<td>157</td>
</tr>
<tr>
<td>Minimum</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Maximum</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>SD</td>
<td>1.25</td>
<td>1.68</td>
</tr>
<tr>
<td>Mean</td>
<td>15.18</td>
<td>14.27</td>
</tr>
</tbody>
</table>
Table 5

*Number of Inference Attempts*

<table>
<thead>
<tr>
<th></th>
<th>Control (n=11)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>74</td>
<td>100</td>
</tr>
<tr>
<td>Minimum</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Maximum</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>SD</td>
<td>2.90</td>
<td>1.92</td>
</tr>
<tr>
<td>Mean</td>
<td>6.73</td>
<td>9.09</td>
</tr>
</tbody>
</table>

In order to account for the differences in the number of inferencing opportunities each participant had, percentages are used in later comparisons for investigating statistical significance, rather than raw scores. The total number of lexical inference attempts each participant made during the think-aloud protocol was divided by the number of unknown target words (as determined from the pretest) to obtain the percentage of inference attempts made. Figure 5 shows the percentage of unknown target words that the participants attempted to infer and the percentage they did not attempt to infer.

Figure 5

*Summary of Mean Percentages*
At a glance, it appears that there may be a significant difference between the two groups on the percentage of target words attempted. Table 6 contains the descriptive statistics for both groups.

Table 6

<table>
<thead>
<tr>
<th></th>
<th>Control (n=11)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>19%</td>
<td>38%</td>
</tr>
<tr>
<td>Maximum</td>
<td>73%</td>
<td>92%</td>
</tr>
<tr>
<td>SD</td>
<td>19.84%</td>
<td>18.55%</td>
</tr>
<tr>
<td>Mean</td>
<td>44.88%</td>
<td>65.35%</td>
</tr>
</tbody>
</table>

The ranges of the percentage of inference attempts made (for unknown target words) for both the control group and the furigana group were similar. However, the minimum, maximum and mean for the experimental (furigana) group were higher than those for the control group, as outlined in Table 6. A two-tailed independent samples $t$-test showed that the difference in means was significant at the $p<.05$ level ($t=-2.50$, $df=20$, $p=.02$). Thus, it appears that the presence of furigana does affect how often participants attempt to infer the meaning of unknown kanji words.

Both the control and the furigana groups exhibited a similar pattern with a standard deviation of almost 20% for both groups. One possible explanation for this can be found in participants’ perceptions of text difficulty (see the discussion of Research Question 3 below). Sternberg (1987) found that how difficult a text is perceived to be is a factor in whether or not readers choose to attempt to make a lexical inference. The participants in this study exhibited a wide range in scores both in the perceived difficulty of the text and, more particularly, in their perceived comprehension. This difference in self-assessed comprehension and perception of difficulty may explain the high standard deviation for the percentage of lexical inferencing attempts. The correlation between self-assessed comprehension and the percentage of inference attempts is statistically significant ($r=.67$, $p=.001$). Correlations between these and other study
variables can be found in Table 7. The results displayed in this table will be further examined in the discussion for Research Question 3.

Table 7

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Group</td>
<td>—</td>
<td>.01</td>
<td>.31</td>
<td>.29</td>
<td>.49*</td>
<td>.59**</td>
<td>-.29</td>
<td>-.45*</td>
<td>-.37</td>
</tr>
<tr>
<td>2. Months in Japan</td>
<td>.01</td>
<td>—</td>
<td>.50*</td>
<td>.37</td>
<td>.57**</td>
<td>.41</td>
<td>-.21</td>
<td>-.35</td>
<td>.17</td>
</tr>
<tr>
<td>3. Pretest Score</td>
<td>.31</td>
<td>.50*</td>
<td>—</td>
<td>.74**</td>
<td>.67**</td>
<td>.45*</td>
<td>-.41</td>
<td>-.63**</td>
<td>-.19</td>
</tr>
<tr>
<td>4. % understood</td>
<td>.29</td>
<td>.37</td>
<td>.74**</td>
<td>—</td>
<td>.67**</td>
<td>.57**</td>
<td>-.76**</td>
<td>-.63**</td>
<td>-.02</td>
</tr>
<tr>
<td>5. % Inf. Attempts</td>
<td>.49*</td>
<td>.57**</td>
<td>.67**</td>
<td>.67**</td>
<td>—</td>
<td>.56**</td>
<td>-.54**</td>
<td>-.69**</td>
<td>-.25</td>
</tr>
<tr>
<td>6. % Correct Inf.</td>
<td>.59**</td>
<td>.41</td>
<td>.45*</td>
<td>.57**</td>
<td>.56**</td>
<td>—</td>
<td>-.48*</td>
<td>-.42</td>
<td>.11</td>
</tr>
<tr>
<td>7. Story Difficulty</td>
<td>-.29</td>
<td>-.21</td>
<td>-.41</td>
<td>-.76**</td>
<td>-.54**</td>
<td>-.48*</td>
<td>—</td>
<td>.75**</td>
<td>-.11</td>
</tr>
<tr>
<td>8. Kanji Difficulty</td>
<td>-.45*</td>
<td>-.35</td>
<td>-.63**</td>
<td>-.63**</td>
<td>-.69**</td>
<td>-.42</td>
<td>.75**</td>
<td>—</td>
<td>.11</td>
</tr>
<tr>
<td>9. How well liked</td>
<td>-.37</td>
<td>.17</td>
<td>-.19</td>
<td>-.02</td>
<td>-.25</td>
<td>.11</td>
<td>-.11</td>
<td>.11</td>
<td>—</td>
</tr>
</tbody>
</table>

* — p<.05, ** — p<.01

Due to the relatively low $p$-value found when comparing the mean difference of the pretest scores and in light of the fact that reading proficiency and vocabulary are known to affect inferencing, further analysis was done to determine if the number of correct answers on the pretest was related to the number of inferences each participant made. When an ANCOVA was calculated with pretest score included as covariate, group was no longer significant at the $p<.05$ level ($F=3.611, p=.07$). The pretest score, however, was significant ($F=11.99, p=.003$). If the score on the pretest is interpreted to represent L2 reading proficiency or, more specifically, vocabulary knowledge, the results confirm previous research regarding the effect of ability on inferencing (See Table 8).
Table 8

<table>
<thead>
<tr>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>.516(^a)</td>
<td>2</td>
<td>.258</td>
<td>10.837</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.379</td>
<td>1</td>
<td>1.379</td>
<td>57.951</td>
</tr>
<tr>
<td>Group</td>
<td>.086</td>
<td>1</td>
<td>.086</td>
<td>3.611</td>
</tr>
<tr>
<td>Pretest Score</td>
<td>.285</td>
<td>1</td>
<td>.285</td>
<td>11.992</td>
</tr>
<tr>
<td>Error</td>
<td>.452</td>
<td>19</td>
<td>.024</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7.650</td>
<td>22</td>
<td>11.992</td>
<td>.003</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>.968</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) \( R^2 = .533 \) (Adjusted \( R^2 = .484 \))

Class level was assumed to be a factor affecting reading proficiency and vocabulary knowledge that would need to be controlled for through stratified random assignment, since more proficient students are usually more likely to make inferences and their inferences are more likely to be accurate and class was assumed to be directly connected to proficiency. Descriptive statistics by class are presented in Table 9. A one-way ANOVA did not find any statistical significance in the differences in the means of the three classes \((F= .812, p= .459)\). Other factors not related to class, such as number of correct answers on the pretest and length of time in Japan appear to be more closely related to the percentage of inference attempts.

Table 9

<table>
<thead>
<tr>
<th>Percentage of Inference Attempts by Class</th>
<th>300 ((n=11))</th>
<th>302 ((n=6))</th>
<th>321 ((n=5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>19%</td>
<td>36%</td>
<td>25%</td>
</tr>
<tr>
<td>Maximum</td>
<td>82%</td>
<td>92%</td>
<td>73%</td>
</tr>
<tr>
<td>(SD)</td>
<td>21.59%</td>
<td>22.97%</td>
<td>20.13%</td>
</tr>
<tr>
<td>Mean</td>
<td>40.29%</td>
<td>61.30%</td>
<td>60.62%</td>
</tr>
</tbody>
</table>

The other variable that was controlled for through stratified random assignment was length of time spent in Japan. Descriptive statistics for participants who had been to Japan for longer than a year and those who had been to Japan for less than a year can be found in Table 10.
As expected, a t-test comparing the means for the two groups was statistically significant ($t=2.36$, $df=20$, $p=.029$).

Table 10

<table>
<thead>
<tr>
<th>Percentage of Inference by Length of Time in Japan</th>
<th>Less than 1 year ($n=4$)</th>
<th>1 year of longer ($n=18$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>24%</td>
<td>19%</td>
</tr>
<tr>
<td>Maximum</td>
<td>44%</td>
<td>92%</td>
</tr>
<tr>
<td>$SD$</td>
<td>8.07%</td>
<td>20.84%</td>
</tr>
<tr>
<td>Mean</td>
<td>34.38%</td>
<td>59.72%</td>
</tr>
</tbody>
</table>

In light of this analysis, the answer to the first research question appears to be that furigana did not have a statistically significant effect on how often the participants in this study attempted lexical inferencing with unknown kanji words once vocabulary knowledge is taken into account. However, the results were approaching statistical significance.

**Research Question 2**

The second research question addresses whether or not lexical inferences are more accurate when participants have furigana than when they do not. As indicated in Table 5, the minimum number of inferencing attempts made by a participant was three. The data from the one participant who made only three inference attempts was not included in the analysis of percentage of correct inference attempts because, with so few inference attempts to judge from, there is a greater possibility of an extreme value (such as 0% or 100%) happening by chance. The scores for the remaining 21 participants (11 in the furigana group and ten in the control group) were included in the analysis.

Figure 6 shows the mean percentages (out of inference attempts) of correct, partially correct and incorrect inferences for each group. The percentages were calculated by dividing each participant’s number of correct inferences, partially correct inferences and incorrect inferences, respectively, by the number of attempts they made at inferencing.
Figure 6

**Summary of Mean Percentages of Correct, Partially Correct and Incorrect Inferences**

![Graph showing percentages of correct, partially correct, and incorrect inferences for control and furigana groups.]

It appears that there are considerable differences between the control and furigana percentage of correct inferences and the percentage of incorrect inferences. Table 11 gives the descriptive statistics for the mean number of correct inferences made by the two groups. Table 12 shows the descriptive statistics for the mean percentage of correct inferences (out of total inference attempts) for each group.

Table 11

<table>
<thead>
<tr>
<th></th>
<th>Control (n=10)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (for the group)</td>
<td>24</td>
<td>42</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Maximum</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>1.43</td>
<td>2.17</td>
</tr>
<tr>
<td>Mean</td>
<td>2.4</td>
<td>4.91</td>
</tr>
</tbody>
</table>

Table 12

<table>
<thead>
<tr>
<th></th>
<th>Control (n=10)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>14%</td>
<td>25%</td>
</tr>
<tr>
<td>Maximum</td>
<td>50%</td>
<td>82%</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>12.78%</td>
<td>15.20%</td>
</tr>
<tr>
<td>Mean</td>
<td>32.70%</td>
<td>52.50%</td>
</tr>
</tbody>
</table>
The means for the two groups were compared using a two-tailed independent samples $t$-test. The difference in the means was found to be significant at the $p<.05$ level ($t=-3.232$, $df=20$, $p=.004$). Thus it appears that the answer to the second research question is that furigana does affect the accuracy of lexical inferences. It appears that participants in the furigana group were more likely to make a correct inference of the meaning of an unknown target word than participants in the control group without access to furigana.

However, since the number of words known on the pretest was found to be a confounding factor when investigating the percentage of inferences made, an ANCOVA was performed with the percentages of correct inferences as the dependent variable and number of correct answers on the pretest as a covariate. The results were very different from those for the percentage of inferences. In this analysis, group appeared to be the most important factor in relation to the percentage of correct inferences. Table 13 contains the results from the ANCOVA.

Table 13

<table>
<thead>
<tr>
<th>Type III Sum of Squares</th>
<th>$df$</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>.265$^a$</td>
<td>2</td>
<td>6.995</td>
<td>6.995</td>
</tr>
<tr>
<td>Intercept</td>
<td>.846</td>
<td>1</td>
<td>44.721</td>
<td>44.721</td>
</tr>
<tr>
<td>Group</td>
<td>.138</td>
<td>1</td>
<td>7.296</td>
<td>7.296</td>
</tr>
<tr>
<td>Pretest Score</td>
<td>.051</td>
<td>1</td>
<td>2.671</td>
<td>2.671</td>
</tr>
<tr>
<td>Error</td>
<td>.360</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4.622</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>.624</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$a. R^2 = .424$ (Adjusted $R^2 = .363$)

The difference between the percentages of partially correct inferences for each group appears small but was further analyzed to determine if there was a statistically significant difference between groups as there was with correct inferences. The percentage of inferences judged to be partially correct was calculated for each participant in the same manner as for
correct inferences and a mean percentage for each group was calculated. See Table 14 for

descriptive statistics for the raw scores and Table 15 for descriptive statistics for the percentage
of partially correct inferences out of total inference attempts. The difference in the means was
not statistically significant ($t=-.78, df=20, p=.45$).

Table 14

*Number of Partially Correct Inferences*

<table>
<thead>
<tr>
<th></th>
<th>Control ($n=10$)</th>
<th>Furigana ($n=11$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (for the group)</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$SD$</td>
<td>.71</td>
<td>.87</td>
</tr>
<tr>
<td>Mean</td>
<td>.50</td>
<td>.82</td>
</tr>
</tbody>
</table>

Table 15

*Percentage of Partially Correct Inferences (Out of Total Inference Attempts)*

<table>
<thead>
<tr>
<th></th>
<th>Control ($n=10$)</th>
<th>Furigana ($n=11$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Maximum</td>
<td>20%</td>
<td>33%</td>
</tr>
<tr>
<td>$SD$</td>
<td>8.56%</td>
<td>10.04%</td>
</tr>
<tr>
<td>Mean</td>
<td>6.16%</td>
<td>9.35%</td>
</tr>
</tbody>
</table>

The means of the percentage of incorrect inferences were also analyzed. Table 16 shows
descriptive statistics for the number of incorrect inferences and Table 17 shows descriptive
statistics for the mean percentage of incorrect inferences. The $t$-test performed on the means of
the percentage of incorrect inferences was statistically significant at the $p<.05$ level ($t=3.25,$
$df=20, \ p=.004$).

Table 16

*Number of Incorrect Inferences*

<table>
<thead>
<tr>
<th></th>
<th>Control ($n=10$)</th>
<th>Furigana ($n=11$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.62</td>
<td>1.57</td>
</tr>
<tr>
<td>Mean</td>
<td>4.20</td>
<td>3.36</td>
</tr>
</tbody>
</table>
Table 17

**Percentage of Incorrect Inferences (Out of Total Inference Attempts)**

<table>
<thead>
<tr>
<th></th>
<th>Control (n=10)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>38%</td>
<td>11%</td>
</tr>
<tr>
<td>Maximum</td>
<td>86%</td>
<td>75%</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>14.32%</td>
<td>19.70%</td>
</tr>
<tr>
<td>Mean</td>
<td>61.14%</td>
<td>38.19%</td>
</tr>
</tbody>
</table>

It is unclear why the differences in the mean percentages of correct and incorrect inferences were significant but the difference in mean percentages of partially correct inferences was not significant. One possibility is the numbers of inferences judged partially correct are so low as to make statistical comparisons difficult. Lexical inferencing studies of native speakers of French and Persian (Wesche & Paribakht, 2009) have found similarly low percentages of partially correct inferences but also very low percentages of incorrect inferences. Studies dealing with L2 inferences (such as Kondo-Brown, 2006a) have tended to find higher percentages of partially correct inferences and lower levels of correct inferences than the present study. This is likely due to different standards of correctness being applied in the various studies. However, it is impossible to be certain about this since no study could be found that detailed exactly what standards were used to determine partial correctness.

Kondo-Brown (2006a) found that when participants had at least partial access to the phonology of an unknown kanji word (i.e. they could correctly pronounce at least one of the characters in the word), they made more accurate inferences than when they did not have any knowledge of the phonology of the word. She interpreted this as meaning that “more successful inferencing is likely to occur when the students had some knowledge about the target kanji words, including phonological information” (p. 141). This result conforms with this finding in that participants who were provided with the phonology of the word made more correct inferences than those that did not. However, Kondo-Brown’s study is fundamentally different
from the present study in that she did not provide the participants with the phonological information. Thus, her results are not necessarily indicative of the benefits of phonological accessibility per se, but rather indicate that readers with some knowledge (even phonological knowledge) of a word or a character within a word are more likely to infer the meaning of that word successfully.

The results from the control group were analyzed to see if they exhibited patterns similar to those found by Kondo-Brown (2006a). First, the 176 responses from the control group (11 participants x 16 target words) were evaluated to see into which of the patterns in Table 18 they fit. As can be seen in the table, more than a third of the target words that could be pronounced correctly during the think-aloud protocol were known on the pretest. Clearly, knowing how to pronounce a word is associated with knowing its meaning. In contrast, none of the words for which the pronunciation was unknown were answered correctly on the pretest.

Table 18
Analysis of Control Group Responses (All Target Words)

<table>
<thead>
<tr>
<th></th>
<th>Pronunciation Known</th>
<th>Pronunciation Partially Known</th>
<th>Pronunciation Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferred Correctly</td>
<td>5 (29.41%)</td>
<td>12 (21.82%)</td>
<td>8 (7.69%)</td>
</tr>
<tr>
<td>Inferred Partially Correctly</td>
<td>1 (5.88%)</td>
<td>2 (3.64%)</td>
<td>2 (1.92%)</td>
</tr>
<tr>
<td>Inferred Incorrectly</td>
<td>2 (11.76%)</td>
<td>19 (34.55%)</td>
<td>23 (22.12%)</td>
</tr>
<tr>
<td>Not inferred</td>
<td>3 (17.65%)</td>
<td>19 (34.55%)</td>
<td>71 (68.27%)</td>
</tr>
<tr>
<td>Known</td>
<td>6 (35.29%)</td>
<td>3 (5.45%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Totals</td>
<td>17 (100%)</td>
<td>55 (100%)</td>
<td>104 (100%)</td>
</tr>
</tbody>
</table>

Though Kondo-Brown also administered a decontextualized test (similar to the pretest in this study), she did not report the numbers of known words for each of the categories of phonological accessibility. Rather, she reported only on the correct inferences, incorrect inferences and non-attempts. Furthermore, she did not have a partially correct category. To make
a more direct comparison with Kondo-Brown’s results possible, the known words were removed
from the sample and all partially correct words were subsumed in the incorrect category. The
results are displayed in Table 19 and Figure 7.

Table 19

*Analysis of Control Group Responses (Unknown Words only, no partial scores)*

<table>
<thead>
<tr>
<th></th>
<th>Pronunciation Known</th>
<th>Pronunciation Partially Known</th>
<th>Pronunciation Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferred Correctly</td>
<td>5 (45.45%)</td>
<td>12 (23.08%)</td>
<td>8 (7.69%)</td>
</tr>
<tr>
<td>Inferred Incorrectly</td>
<td>3 (27.27%)</td>
<td>21 (40.38%)</td>
<td>25 (24.04%)</td>
</tr>
<tr>
<td>Not inferred</td>
<td>3 (27.27%)</td>
<td>19 (36.54%)</td>
<td>71 (68.27%)</td>
</tr>
<tr>
<td>Totals</td>
<td>11 (100%)</td>
<td>52 (100%)</td>
<td>104 (100%)</td>
</tr>
</tbody>
</table>

Figure 7

*Analysis of Control Group Responses (Unknown Words Only, No Partial Scores)*

If the pronunciation of the entire word was known, the rate of correct inferencing was
almost 50% whereas the rate of correct inferencing when pronunciation was not known was
under 10%. Kondo-Brown’s data exhibited a similar, though slightly more pronounced, pattern
of more accurate inferencing the more phonological information the participant knew. A $\chi^2$ analysis on the data showed that phonological accessibility and inference are related ($\chi^2=22.64$, $df=4$, $p=.0001$). Cramer’s V was calculated to measure the strength of association. The result ($V=.26$) shows a strong relationship between the two factors.

**Research Question 3**

To answer the third research question, means for participants’ ratings on Likert scale questions about story difficulty and kanji difficulty and participant assessment of the percentage of the text that they understood were compared. Table 20 contains descriptive statistics for these three measures. The control group appeared to find the story slightly more difficult than the group without furigana. However, the difference in the means was not significant at the $p<.05$ level ($t=-1.327$, $df=20$, $p=.20$). The difference between the kanji difficulty is significant at the $p<.05$ level ($t=2.279$, $df=20$, $p=.034$), indicating that the group without furigana perceived the kanji in the story to be more difficult than the group with furigana. The means for the percentage of the story understood also appear to be different but the difference is not statistically significant. Taken together these three measures indicate that there may be a difference in perceptions of difficulty, particularly for kanji, but not for the story as a whole. With these mixed results it is difficult to give an unequivocal answer to research question 3. The mixed results indicate that further research is needed on this topic.

As would be expected, a participant’s subjective ratings of the percent understood correlated with the rating of both kanji and story difficulty. Correlation coefficients can be found in Table 7.
Table 20

Perceptions of Difficulty  

<table>
<thead>
<tr>
<th></th>
<th>Control (n=11)*</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Story difficulty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Maximum</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>SD</td>
<td>.9</td>
<td>.7</td>
</tr>
<tr>
<td>Mean</td>
<td>4</td>
<td>3.6</td>
</tr>
<tr>
<td><strong>Kanji difficulty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Maximum</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>SD</td>
<td>.9</td>
<td>.8</td>
</tr>
<tr>
<td>Mean</td>
<td>4.18</td>
<td>3.4</td>
</tr>
<tr>
<td><strong>Percentage understood</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Maximum</td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>SD</td>
<td>22.5%</td>
<td>22.3%</td>
</tr>
<tr>
<td>Mean</td>
<td>34%</td>
<td>47.3%</td>
</tr>
</tbody>
</table>

*One participant in the control group did not answer the “percentage understood” question. For this indicator only n=10, for the control group.

Research Question 4

The fourth and final research question addresses learner affect, namely, whether or not the presence of furigana affects how much participants like the text. This question was assessed by having the participants respond to the question “Did you like the story?” using a 5-point Likert scale with 1 being “did not like at all” and 5 being “liked a lot.” Table 21 summarizes the descriptive statistics for both groups. The control group appeared to like the story more than the furigana group. However, the difference in the means is not statistically significant at the $p<.05$ level. Thus, it appears that there is no relationship between the presence of furigana and how well the participants liked the story.
Table 21

‘Did You Like the Story’ Likert Scale

<table>
<thead>
<tr>
<th></th>
<th>Control (n=11)</th>
<th>Furigana (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Maximum</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>SD</td>
<td>.75</td>
<td>.67</td>
</tr>
<tr>
<td>Mean</td>
<td>3.18</td>
<td>2.64</td>
</tr>
</tbody>
</table>

Pearson correlations were calculated for the Likert scale ratings of how well the participants liked the story and the pretest score, the percent understood, the percentage of inference attempts, the percentage of correct inferences, and the Likert scale ratings for story and kanji difficulty (Table 7). None of these variables was found to have a statistically significant correlation with the ratings of how well the story was liked.

These results indicate that furigana does not affect how much the participants in this study liked the story. Furthermore, the extent to which participants liked the story does not appear to be related to how well they think they understand the story or how difficult they perceive the story and kanji within the story to be.
Chapter 5: Discussion and Conclusions

Summary of Results

This study found that furigana affected the accuracy of lexical inferences as well as participants’ perception of kanji difficulty in a text. Furigana also seemed to affect the rate at which the participants infer, though once participant scores on the vocabulary pretest were taken into account, the effect was not statistically significant at the \( p < .05 \) level. The effect was approaching significance, however, and further research is needed in this area.

The presence of furigana did appear to affect the accuracy of inferences. Participants with access to furigana were more accurate in their inferences than those without. The mean percentage of accurate inferences for the control group was 32.70\%, whereas the mean for the group with furigana was 52.50\%. Thus it appears that access to phonological information improves the accuracy of inferences, at least for the participants in this study. In addition to this new finding, an analysis of the control group's responses confirmed previous research (Kondo-Brown, 2006a) showing that knowledge of character readings is related to successful inferencing. The control group participants in this study, as well as the participants in Kondo-Brown's study, were more likely to correctly infer the meaning of an unknown word if they were able to correctly write even part of the reading on the pretest.

The results for the third research question were mixed. The mean rating for story difficulty was slightly higher for the control group than for the furigana group. However, the difference was not statistically significant. The difference in the mean rating for kanji difficulty, however, was statistically significant with the control group having a higher mean rating for kanji difficulty than the furigana group. As for the fourth research question, furigana did not appear to affect how much the participants liked the text.
No statistically significant correlations were found between how well the participants liked the text and the other variables. However, statistically significant correlations were found between some of the other variables. Correlations between self-assessed percentage understood and the story difficulty rating, the kanji difficulty rating, the score on the pretest, the percentage of inference attempts made, and the percentage of correct inferences were all found to be statistically significant.

**Limitations of This Research**

Like most research studies, this study has a few characteristics that could limit the generalizability of the findings. Research participants were volunteers and thus may not be a truly representative sample of the population of BYU students studying Japanese. The small sample size (n=22) is also a limitation for the generalizability of the findings. As mentioned previously, the majority of the participants served Japanese-speaking missions for The Church of Jesus Christ of Latter-day Saints. These participants have unique language learning background and have received much more exposure to Japanese than most university learners and generally have larger vocabularies and better speaking and listening abilities, compared to reading ability, than students in the same classes who have not served missions in Japan. Thus, though the target words were chosen because they were unlikely to be familiar to the participants in either written or spoken form and the instrument was pilot-tested with a small group of students with a similar learning background, it is possible that the participants had some previous experience with the spoken form of the target words. If that were the case, it could mean that furigana only facilitates inferencing insofar as it triggers word knowledge that the participants already had. Further research is needed in this area, in particular with participants that more closely reflect the language learning experience of the average U.S. university student.
The missionary experience of the majority of the study participants might also affect their likelihood of making lexical inferences. Anecdotal evidence from teachers who have worked with students with missionary experience and those without it suggests that students who have had the mission experience may be more willing to attempt a task even when they do not have perfect comprehension of the materials than those who have not had that experience, thus they might be more likely to attempt lexical inferencing than the average learner. However, it must be noted that, in a t-test comparing the means of only those participants who did not know any of the target words on the pretest, the difference in the mean percentage of inference attempts between those who had spent a year or more in Japan and those who had spent less than a year in Japan was not significant. These statistics must be interpreted cautiously due to the small numbers of participants in the sample size. However, it is possible that time spent in Japan, be it for a mission or otherwise, may be a relevant factor affecting the rate of inferences only because it is related to L2 proficiency and not due to any specific characteristics of people who learned Japanese on a mission.

A further limitation of the study is that it was not possible to accurately gauge the Japanese proficiency and vocabulary knowledge of the students. This is an important limitation since proficiency is a significant factor correlated with the rate and accuracy of guesses (Bengeleil & Paribakht, 2004). In future studies, proficiency should be determined using an unrelated, validated measure. Perhaps because most learners in foreign language situations tend to learn the written language at the same time as the spoken language, little attention has been paid to the potential impact of spoken language proficiency, independent of written language proficiency. In order to better understand the relative effects of different types of proficiencies
on lexical inferencing, future studies would ideally obtain measurements for both written and oral language proficiency.

The passage used in the think-aloud protocol was pilot-tested with a group of students with a similar Japanese learning background and it was determined that the passage was of the appropriate difficulty level. However, it is possible that the story itself chosen for the inferencing task may have been too difficult for the participants, particularly those in the control group who on average knew fewer target words (though the difference in means was not statistically significant). Researchers have reported that second language readers must know 95-99\% of words in a text in order to be able to comprehend the text and make contextually appropriate lexical inferences (Wesche & Paribakht, 2009). In other words, at most only one out of every twenty words encountered in a text should be new to the reader if the reader is expected to be able to comprehend and make lexical inferences.

There are some challenges in applying these guidelines to Japanese texts because the number of words is not usually used as a measure of text length in Japanese (due to the challenge of determining word boundaries in conventionally written Japanese). Generally, text length is measured in characters. The length of the text used in the experiment is 1200 characters. The average length of the target words was two characters and the average number of unknown target words was 14.73. Therefore, on average the characters in the unknown target words account for 2.5\% of the characters in the passage, making the percentage of unknown characters in the target words well within the 95-99\% guidelines for comprehensibility. However, the raters reported that no student was able to translate all non-target words. In other words, at least some of the non-target words were unknown to each participant. This study did not incorporate a measure to determine participant knowledge of all the words known in the passage so it is not possible to
say for certain, but there is a possibility that the percentage of unknown words in the passage exceeds the guidelines for comprehensibility stated above. Since vocabulary knowledge was not controlled for, the possibility also exists that the two groups were not equivalent in this respect, which would influence the interpretation of the results. The study design would be strengthened significantly by the addition of a general vocabulary measure, preferably one assessing both depth and breadth of vocabulary knowledge, to adequately account for this potentially confounding variable.

The think-aloud protocol used to collect data also has limitations. Asking participants to stop reading and provide the meaning in English necessarily interrupts the flow of reading the passage. Furthermore, the absence of a word in a subject’s protocol does not necessarily mean that he or she did not make any guess about the meaning of the word in question. It is possible that the subject simply did not think the information contained in the word was important enough to include in his or her oral translation. In addition, subjects may have simply forgotten to include some words due to the working memory burden of the cognitively complex sight translation task, not because they had not made a mental guess of the meaning of the target word.

On the other hand, the nature of the task in the think-aloud protocol may have actually encouraged subjects to make more guesses than they would have under normal reading circumstances. It might have felt awkward or uncomfortable for subjects to say that they did not know a certain word or to leave obvious holes in the text by leaving unknown words out. Despite these limitations to the think-aloud protocol, it remains an important way to glimpse the comprehension processes that are taking place while a person reads.
Implications of the Results for Theory and Pedagogy

Like previous lexical inferencing studies, the participants in this study chose not to attempt to infer the meaning of unknown target words much of the time (Bensoussan & Laufer, 1984; Paribakht & Wesche, 1999). However, students reading the phonologically accessible text were more likely to attempt lexical inferencing and were more likely to guess correctly. This indicates that, at least for deep orthographies such as Japanese kanji, phonological accessibility may also be a factor in lexical inferencing, in addition to already known factors such as L2 proficiency, vocabulary knowledge or context usefulness. This should be explored further in future research.

Because a certain level of comprehension is generally required before a reader can attempt lexical inferencing (Wesche & Paribakht, 2009), the effect of the phonological transparency afforded by furigana on lexical inferencing can give us an indirect look at the impact of phonology on reading comprehension. The results show that phonological accessibility has a positive impact on the quality of lexical inferences, and may also affect the rate of inferences. This result aligns with previous research suggesting that phonology is an important part of reading comprehension processes in general. The reasons for why this is the case are not clear, but theories of working memory provide one plausible explanation. By allowing the reader easy access to the phonological code, which is better suited to holding information in working memory, furigana may improve the reader’s ability to process a sentence in order to come up with a guess for an unknown word.

The results also support the idea that phonology is an important part of kanji word learning, at least for native English speakers. The participants were better able to make a form-meaning connection through correct lexical inferencing in the phonologically transparent
furigana condition. Though this study did not address incidental word learning, it stands to reason that phonologically opaque vocabulary words are less likely to be learned incidentally, since it is less likely that the reader will make a correct lexical inference to begin with.

There is evidence that word learning can take place outside of the classroom setting through extensive exposure to words when reading (Bengeleil & Paribakht, 2004; Bensoussan & Laufer, 1984; Everson & Kuriya, 1998; Haastrup, 1991; Harley & Hart, 2000; Huckin & Coady, 1999; Krashen, 1989, 1993; Nagy, Herman, & Anderson, 1985; Nassaji, 2003a; Qian, 2004; Shimizu & Green, 2002; Sternberg, 1987; Wesche & Paribakht, 2009). Materials can and should be designed to take advantage of the vocabulary-enriching potential of extensive reading and the facilitative effects of furigana. For example, if teachers want to increase the amount of vocabulary learned incidentally through extensive reading activities, they might consider using reading materials with furigana.

**Suggestions for Future Research**

There are several ways in which future research studies could build upon this study. As this was the first study of its kind investigating the role of furigana in the lexical inferencing of unknown kanji compounds, it should be replicated in order to validate the results. In addition to exact replication, variations should be made to increase the generalizability of the results. For example, context is an important factor in inferencing (Wesche & Paribakht, 2009) so the results may be affected by the nature of the context in which the target word appears. Thus, the study procedures should be replicated using different passages, in addition to replication using the same instrument.

Because the rate at which a person guesses unknown words varies greatly from person to person, as indicated by the high standard of deviation of the mean percentage of inference
attempts, a future study might devise a way to have each subject participate in both an experimental furigana condition and a control non-furigana condition in order to compare the rate and quality of guesses on the two conditions within subjects.

Though this study attempted to account for prior knowledge of vocabulary by using the vocabulary pretest, it is still possible that students might have had some level of oral knowledge of the target words that could have been triggered by the presence of furigana in the experimental condition. Future researchers might consider trying to mitigate this possibility by using plausible non-words instead of the actual pronunciation for the characters. Another alternative would be to provide furigana on the pretest and only evaluate students’ knowledge of meaning on the pretest.

Since the sample size was relatively small and homogeneous, with the majority of participants having a similar educational backgrounds and experience in Japan, larger studies with a population more representative of all native-English speaking Japanese learners should be conducted in order to be able to generalize the results. In particular, past studies have shown that L2 proficiency is a leading predictor of the rate at which a learner guesses unknown words. Future studies could look at a wider range of skill levels and passage difficulty relative to skill level.

Furthermore, future studies should look more closely at various aspects of proficiency and their relationships with lexical inferencing. The language learning profile of the majority of participants, with considerably stronger oral language ability in relation to written language ability, is probably not common among U.S. university students. However, these characteristics might be more common among Japanese learners in Japan (Japanese as a Second Language learners) from countries which do not use kanji (e.g. immigrants from Brazil or other South
American countries). It seems logical that many of these learners, by virtue of being afforded large amounts of exposure to the spoken language through their in-country experience, would also have higher oral proficiency in relation to their written language proficiency. Though previous studies on lexical inferencing while reading investigate the relationship between proficiency and inferencing success, the emphasis seems to be on proficiency as judged by written language measures, without taken oral proficiency into account. This study makes a valuable contribution to the field by acknowledging the learning background of the participants and recognizing the potential effects that may have had on the results. It is important that future research takes into account oral language proficiency as well as written proficiency (or proficiency judged by written measures). Future research should investigate learners with a wide variety of language learning backgrounds to ascertain what effects various interventions may have on different types of learners.

In addition to investigating the effect of furigana on guessing the meaning of unknown words specifically, much research still needs to be done on furigana in general. The potential impact of furigana on incidental learning has yet to be studied empirically. Other than the previously mentioned classroom action research by Kirwan (2005), there does not appear to be much research directly investigating furigana. An empirical study investigating incidental learning of kanji readings and meanings would add greatly to the current dialogue. An eye-tracking study to determine where students of different proficiencies are actually looking when they read materials with furigana would also add valuable information to the discussion on furigana and Japanese reading.

Future research could also look at the possible link between furigana and motivation. The current study found no link between how well participants liked the text and whether or not
the text had furigana. However, it did find a link between perceived kanji difficulty and the use of furigana. Because learners are more likely to feel motivated to complete a task when the perceived exertion is lower (Anderson, 1999), it is possible that furigana could increase motivation. This, along with other affective factors associated with furigana, such as student and teacher feelings and beliefs about furigana use, is a possible area of focus for future research.

**Conclusion**

This study was the first of its kind to directly investigate the effects of furigana on Japanese learners’ lexical inferencing. It was found that furigana affects the accuracy of inferences but not the rate at which inferences are made. The presence of furigana also affected participants’ perceptions of the difficulty of the kanji in the text. However, furigana was not found to affect learner perceptions of the difficulty of the text or the extent to which a student liked the story. Though much research is still needed in this field, these results have implications for both pedagogy and theory. In particular, the results indicate that phonological accessibility may be a factor in lexical inferencing, that furigana has a positive impact on the accuracy of lexical inferences, and that furigana may facilitate incidental word learning by increasing the likelihood of accurate lexical inferencing. Issues of phonological accessibility and its effects on reading in Japanese as a foreign language offer fertile ground for much future research.
References


## Appendix A: Kanji Pretest

Instructions: For each Kanji word below, write the reading in romaji or hiragana and a brief English meaning.

<table>
<thead>
<tr>
<th>No.</th>
<th>Kanji</th>
<th>Reading</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>陳列</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>予報</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>本棚</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>断言</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>支度</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>疑惑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>検査</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>戦争</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>受理</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>指導者</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>容器</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>政府</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>教科書</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>叱責</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>金貨</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>中心</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>診断</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>丁重</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>形容</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>転落</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>天皇</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>乗車券</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>故障</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>購入</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>試験</td>
<td>身体</td>
<td>大切</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
</tbody>
</table>
Appendix B: Story Used in the Protocol

Target words are underlined. Original words are provided in parenthesis.

K氏は週末の旅行に出かけるため、支度をととのえていた。

服のポケットのなかでは、ラジオが天気予報を告げていた。

<あすは、よいお天気でしょう・・・>

楽しげに口笛を吹きながら、K氏はハンカチを出し、腕時計を軽くぬぐった。

これは彼のいつもの癖だった。

癖とはいうものの、頭をかくとか耳をつまむとかいう、意味もない仕草（動作）とはちがっていた。

彼はその時計を大切にしていたのだ。

大げさな形容をすれば、愛していたともいえる。

K氏がこれを購入して（買って）から、五年ほどになる。

デパートの時計売場のそばを通過とき、ガラスの容器のなかに陳列された、たくさんの時計の一つがキラリと光った。

ちょうど、女の子にウインクされたような気がした。

また、「あたしを買ってくれない・・・・」と、やさしく、ささやきかけられたようにも思えた。

古代の異国の金貨が、文字盤になっている。

たまたま、入社してはじめてのボーナスを受理した（もらった）日だった。

「よし。買うことにしよう」彼は無意識に（思わず）こうつぶやいた。

それ以来、時計はずっと、K氏と一緒にいる（ともにいる）。

K氏は、身体の一部でもあるかのように扱った。
彼はまだ若く、自分では定期的な健康診断などを受ける気にはならなかったが、時計のほうは定期的に検査に出した。

別なのを使うその数日は、彼にとって、たまらなく孤独な（さびしい）日だった。

しかし、そのため、錯乱したり（狂ったり）することはまったくなかった。

進みすぎもせず、おくれもせず、正確な時刻を、忠実に知らせつづけてきたのだ。

その時、ラジオが時報の音をたてた。

Ｋ氏は首をかしげた。

「おかしいぞ。時報が狂うとは」

彼にとって、時計のほうに疑惑をかける（疑う）のは、考えられないことだった。

だが、ダイヤルをまわし、ほかの局を調べ、時報が正しいのを知って、あわてた。

もはや、乗車券（切符）を買っておいたバスの、発車時刻にまにあわなくなっている。

彼は時計に叱責した（文句を言った）。

「おい。なんということをしてくれたのだ。これだけ丁重に扱ってやってるのに」

しかし、どうしようもなかった。

Ｋ氏は旅行を断念（中止）し、散歩にでかけた。

そして、ついでに時計店に立ち寄った。

「妙（変）だな。おくれはじめた。せっかくの週末が、ふいになってしまった」

「しかし、このあいだ検査をしたばかりですが・・・」と、時計店の主人は受けとり、機械をのぞきこんでいたが、ふしぎそうな声で答えた。

「変ですね。どこにも故障なんかないようですよ」
「そんなはずはない」 そのとき、ポケットに入れっぱなしになっていたラジオが、こう報道した（告げた）。

＜観光シーズンです。 S山へ行くバスが・・・・・＞

それを聞きながら、K氏は断言した。「おかげで、このバスに乗りそこなったのだ。たしかに、この時計はどうかしている」

しかし、ニュースはそのさきをこう告げていた。＜・・・・・事故のため、谷へ転落して・・・＞
Appendix C: Background Questionnaire

Name: ___________________________________

Age: ___________

Gender: ____________

Highest level Japanese class taken at BYU: ____________________________

Languages spoken: ______________________________________________________________

Time spent in Japan/Japanese study experiences (Please describe the time you have spent in Japan and the time you have spent studying Japanese):

______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________

What percentage of the story (from 0% to 100%) do you think you understood? ________

For each of the questions below, please circle one answer.

<table>
<thead>
<tr>
<th>Very Easy</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

How difficult was the story overall?  
How difficult was the kanji in the story?

<table>
<thead>
<tr>
<th>Did not like at all</th>
<th>Liked a lot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Did you like the story?
Appendix D: Selected Samples of Protocol Scoring

The excerpts from the protocols below provide examples of scoring for reading (pronunciation) and meaning. The first line in each sample is the Japanese text as it appeared in the story. The second line is a transcription (in Hepburn romanization with double vowels representing long vowels) of the participant's reading of the text. The third line is the English translation the participant provided for that sentence. The target word is underline in each line. Correct readings and meanings are provided after the sample as needed.

Sample 1: Correct reading, correct meaning

1. K氏がこれを購入してから、五年ほどになる。
2. K-shi ga kore o koonyuu shite kara gonen hodo ni naru
3. It’s been about 5 years since Mr. K bought this purchased this.

Sample 2: Correct reading, incorrect meaning

1. それを聞きながら、K氏は断言した。 「おかげで、このバスに乗りそこなったのだ。 たしかに、この時計はどうかしている」
3. So as he was listening to that kay min kay he stopped talking*  yeah I don’t really know

*Correct meaning: assert, declare

Sample 3: Correct reading, meaning not attempted

1. 彼は時計に叱責した。
2. Kare wa tokei ni shiseki shita.
3. He shisekki’d* the watch.

*Correct meaning: scold(ed), yell(ed) at
Sample 4: Incorrect reading, incorrect inference

1. 彼は時計に叱責した。

2. kare wa tokei ni nantoka* shita.

3. I think he turned** the watch.

*Correct reading: shisseki (nantoka is a Japanese word meaning "something" or "somehow" and is sometimes used as a filler when a reader encounters unknown characters during oral reading)

**Correct meaning: scold(ed), yell(ed) at