Implicit, Eclipsed, but Functional:

The Development of Orthographic Knowledge in Early Readers

by

Tanya Kaefer

Department of Psychology & Neuroscience
Duke University

Date: ______________________
Approved: ______________________

Gary Feng, Supervisor

Reiko Mazuka

Makeba Wilbourn

Jennifer Smith

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology & Neuroscience in the Graduate School of Duke University

2009
ABSTRACT

Implicit, Eclipsed, but Functional:
The Development of Orthographic Knowledge in Early Readers

by

Tanya Kaefer

Department of Psychology & Neuroscience
Duke University

Date: ______________________

Approved:

___________________________
Gary Feng, Supervisor

___________________________
Reiko Mazuka

___________________________
Makeba Wilbourn

___________________________
Jennifer Smith

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology & Neuroscience in the Graduate School of Duke University

2009
Abstract

Although most models of reading development present orthographic knowledge as a more advanced and later developing form of knowledge than phonological knowledge, this dissertation presents a model of the development of orthographic knowledge in which generalized orthographic knowledge, the knowledge of symbol patterns within and across words, develops early, at the same time as phonological knowledge and before lexicalized representations of a whole word. However, because phonological and generalized orthographic knowledge are not fully integrated, phonological knowledge masks orthographic knowledge in typical measures of literacy.

In study 1 pre-readers’ knowledge of the elements that make up words was tested using eye-tracking as a measure of implicit knowledge. We find that pre-reading children as young as 3 have implicit orthographic knowledge regarding the elements that make up words. This supports the prediction that generalized orthographic knowledge develops before lexicalized knowledge.

In study 2, children’s creative spellings were used to gauge children’s implicit knowledge of letter patterns in a naturalistic setting. We find that kindergarteners in particular tend to rely on phonology over orthography when the two are in conflict. This supports the hypothesis that phonological knowledge can mask orthographic knowledge.
In study 3, children were asked to decode non-words and their implicit knowledge of letter patterns was measured using eye tracking. I found that early readers show some implicit knowledge when decoding. This supports the hypothesis that generalized orthographic knowledge can be measured in literacy tasks under certain testing conditions.

In study 4, children’s phonological and orthographic knowledge was tested directly by asking children to sound out and select the best word. Results show that sensitivity to orthographic violations is decreased when phonology is introduced. This is a direct test of the hypothesis that phonological knowledge can mask orthographic knowledge, and findings support this hypothesis.

These results suggest that pre-readers show generalized orthographic knowledge before lexicalized knowledge and concurrently with phonological knowledge. Furthermore, this generalized orthographic knowledge initially presents itself implicitly, and in many early literacy tasks the orthographic domain is dominated by phonological concerns. Essentially, orthographic and phonological knowledge develop at the same time; however, until children learn to integrate the two dimensions of written language, they rely on one source over the other.
Dedication

For my family
Contents

Abstract ......................................................................................................................................... iv
List of Tables ................................................................................................................................. xi
List of Figures .............................................................................................................................. xii
Acknowledgements ...................................................................................................................xiv

1. General Introduction ................................................................................................................ 1
   1.1 Generalized Orthographic Knowledge ......................................................................... 3
   1.2 Generalized Orthographic knowledge and English .................................................. 11
   1.3 The representation of orthographic Knowledge ........................................................ 15
   1.4 The development of orthographic knowledge ........................................................... 17
      1.4.1 Early generalized orthographic knowledge ......................................................... 17
      1.4.2 Stage theories of orthographic knowledge development .................................. 21
      1.4.3 Mechanisms of the development of Generalized Orthographic Knowledge ... 23
   1.5 Interaction between Phonology and Orthography .................................................... 28
   1.5 Implicit and Explicit Knowledge .................................................................................. 36
   1.6 The Current Study .......................................................................................................... 38

2. Pre-Readers' orthographic knowledge ................................................................................ 40
   2.1 Methods ........................................................................................................................... 43
      2.1.1 Participants ................................................................................................................. 43
      2.1.2 Apparatus ................................................................................................................... 44
      2.1.3 Materials ..................................................................................................................... 45
4.1.4 Behavioural coding ................................................................................................... 90
4.2 Results .............................................................................................................................. 91
4.2.1 Behavioural results.................................................................................................... 91
4.2.1 Word Initial Doublets ............................................................................................... 92
4.2.2 Word Initial Stop/Stop Clusters .............................................................................. 98
4.2.4 Word Final Stop/Liquid Clusters .......................................................................... 104
4.3 Discussion...................................................................................................................... 110

5. Orthographic knowledge in decoding and judgment ..................................................... 115
5.1 Method ........................................................................................................................... 118
5.1.1 Participants............................................................................................................... 118
5.1.2 Materials ................................................................................................................... 118
5.1.3 Procedure.................................................................................................................. 119
5.2 Results ............................................................................................................................ 120
5.2.1 Explicit orthographic knowledge.......................................................................... 120
5.2.2 Eye-tracking results ................................................................................................. 123
5.2.2.1 Subtle illegal letter combinations ................................................................... 123
5.2.2.2 Salient illegal letter combinations .................................................................. 127
5.2.2.3 Implicit knowledge and reading skill ........................................................... 130
5.3 Discussion...................................................................................................................... 131

6. General discussion ................................................................................................................ 136
6.1 Conceptual Model of Generalized Orthographic Knowledge ............................... 136
6.2 Generalized Orthographic Knowledge and Phonological Knowledge .......... 138
List of Tables

Table 1: Correlation of explicit tasks with emergent literacy tasks........................................53

Table 2: Correlation of implicit looking time with emergent literacy ........................................54

Table 3: Standardized regression coefficients for emergent literacy tasks and explicit orthographic knowledge controlling for age .................................................................55

Table 4: Standardized regression coefficients of emergent literacy and implicit orthographic knowledge....................................................................................................................56

Table 5: Correlation of proportion of spelling mistakes (sound to letter correspondence) with writing scores ........................................................................................................73

Table 6: Correlation between proportion misspelled morphological suffixes and literacy scores........................................................................................................................................78

Table 7: Correlations between literacy tests and morphological suffix spelling by frequency ......................................................................................................................................79

Table 10: Looking time correlated with reading subtests, partialed for decoding strategy ......................................................................................................................................131

Table 11: Stimuli list for Chapter 2 ...........................................................................................145

Table 12: Stimuli list for Chapter 4 ...........................................................................................147

Table 13: Stimuli list for Chapter 5 ...........................................................................................148
List of Figures

Figure 1: Conceptual model of generalized orthographic knowledge in reading development ................................................................. 5

Figure 2: Percent of children answering ‘yes’ to the question “is this a word?” by condition and age (plus standard error). A * indicates significance at the .05 level in comparison with the pseudo-word condition. ................................................................. 48

Figure 3: Average amount of time children spend looking at the symbol string by condition and age (plus standard error). A * indicates significance at the .05 level in comparison with the pseudo-word condition. ........................................................................ 51

Figure 4: Proportion of spellings omitting the vowel in a stop/vowel/stop combination or a stop/vowel/liquid combination ................................................................................................................. 69

Figure 5: Average proportion of omission errors in stop/vowel/stop combinations and stop/vowel liquid combinations by grade, frequency and combination ........................................................................ 71

Figure 6: Longitudinal subjects change in performance by condition and frequency ...... 72

Figure 7: Proportion of errors in the morphological suffix by phonological match and grade ............................................................................................................................................. 74

Figure 8: Proportion of errors by grade, phonological match, and frequency ............. 75

Figure 9: Types of errors made by phonological match and age .................................. 77

Figure 10: Looking time at initial letters by condition and decoding strategy .............. 93

Figure 11: Percent of non-blenders looking at each letter by time .............................. 94

Figure 12: Percent of blenders looking at each letter by time ..................................... 95

Figure 13: Proportion looking at first 2 letters by strategy, condition, and time .......... 97

Figure 14: Amount of time looking at the first 2 letters in a word by strategy and condition .................................................................................. 99

Figure 15: Percent of non-blenders looking at each letter by time .............................. 100
Figure 16: Percent of blenders looking at each letter by time ................................. 101
Figure 17: Proportion of participants looking at first 2 letters by decoding strategy, condition, and time .......................................................... 103
Figure 18: Amount of time looking at last two letters by strategy and condition .......... 105
Figure 19: Percent of non-blenders looking at each letter by time ............................ 106
Figure 20: Percent of blenders looking at each letter by time .................................... 107
Figure 21: Proportion of participants looking at last 2 letters by decoding strategy, condition, and time .......................................................... 109
Figure 22: Performance on behavioural selection task by condition and phonology ..... 121
Figure 23: Performance on behavioural task by condition, phonology and decoding strategy .......................................................... 122
Figure 24: Looking time at non-word by legality, condition, and strategy ................. 124
Figure 25: Proportion of subjects looking at the illegal non-word by time, phonology and strategy ................................................................. 126
Figure 26: Time spent looking at non-words by task, strategy and legality ............... 128
Figure 27: Proportion of subjects looking at illegal non-word across trial time ......... 129
Acknowledgements

Thank you to my advisor, Gary Feng, for his comments and help on this dissertation. Thank you also to Amanda Byrd, Buff Gibson, and Maya McKeand-Peraza for their aid in recruitment and running participants, as well as all the research assistants who transcribed and coded the spelling work. Finally, thank you to the students, parents, and administrators who participated in these studies. This work was partially funded by the Sulzberger social policy fellowship provided by Duke University Center for Child and Family Policy.
1. General Introduction

Learning to read in English is a task governed by contradictions (Share, 2008). A letter makes a certain sound, except when it does not; certain letter combinations don’t go together, except when they do, and certain sounds are represented with certain letters, except when they are not. Given all the irregularities and exceptions in the English writing system (Venezky, 1999), it is often less amazing that many children struggle with learning to read than that most children pick it up at all. Learning how children begin the process of learning these rules before they have formal instruction, gives us the benefit of knowing where the average child starts in his or her understanding of written language. Knowing this gives us the ability to evaluate and adapt the curriculum to their strengths and pick up slack where they may have weaknesses.

The current studies depart from traditional research in a number of ways. First, most studies of orthographic knowledge focus on children’s recognition of the representation of individual words, that is, lexicalized orthographic knowledge. I focus on the knowledge of generalized orthographic patterns that repeat across many words, and expand the definition of orthographic knowledge to include knowledge about the symbols that make up words. Second, because of this focus on lexicalized knowledge, reading based on orthographic patterns is often seen as more advanced and emerging later than reading based on phonological links between letter and sound. Consequently
most studies on children’s orthographic processing focus on post-emerging readers, and relatively little attention is paid to beginning readers’ and pre-readers’ knowledge of orthographic patterns. I hypothesize that when we consider generalized orthographic knowledge, we’ll see development begin much earlier than previously thought. To that end, I focus on orthographic knowledge of pre-school to kindergarten aged children who are either non-readers or beginning readers. Third, traditional studies focus on children’s explicit orthographic knowledge, knowledge about patterns in words that they can elucidate verbally. There is reason, however, to believe that much of children’s early orthographic knowledge is implicit; thus, I assess children’s implicit orthographic knowledge as a means of determining the earliest appearance of this knowledge. Finally, most studies focus on a stage like transition from emphasis on phonological correspondence of letters to an emphasis on orthographic patterns. I, on the other hand, believe that both phonological and orthographic knowledge are operational among beginning readers, and I, therefore, do not presume a stage-like transition from phonological to orthographic strategies; instead I show that even pre-readers possess some basic orthographic knowledge, and, depending on the task requirement, beginning readers can and do apply their orthographic knowledge to literacy tasks.

The current studies intend to explore children’s knowledge of the patterns of letters within words, by examining children’s earliest approaches to written language through the progression of orthographic knowledge from implicit to explicit awareness.
in both reading and spelling. I hypothesize that beginning readers have emerging knowledge of generalized orthographic patterns, which develops before children have a solid grasp of lexicalized orthographic knowledge. This knowledge develops at the same time as knowledge of phoneme-grapheme correspondence, but because the two sources of knowledge have yet to be integrated, children selectively rely on one source of information more than the other. I hypothesize that children will show basic orthographic knowledge before they learn to read words phonologically, but that, as they begin to learn to read, the focus on phonology in literacy tasks will mask their orthographic knowledge. In order to test this hypothesis I conducted four studies examining children’s early orthographic knowledge in varying phonological contexts. I combined behavioural and eye-tracking methodologies as well as experimental and naturalistic settings in order to determine the earliest expression of orthographic knowledge as well as how this knowledge can be manipulated by the phonological context of an individual literacy task.

1.1 Generalized Orthographic Knowledge

Orthographic knowledge includes “both knowledge of the actual spelling of particular words and higher level conceptual skills, such as the recognition of the properties of words and sequences and typical positions of letters in English” (Siegel, Share, & Geva, 1995). That is, orthographic knowledge has two elements. Both lexicalized knowledge of individual word representations and generalized knowledge
of patterns across words contribute to orthographic knowledge. I focus on the second element of this definition and expand it to include even more basic knowledge about the orthographic patterns that make up words. Generalized orthographic knowledge encompasses all the associated knowledge about symbolic patterns in words that are distinct from associations with sound. In short, generalized orthographic knowledge can be thought of as an abstract knowledge of the orthographic rules and conventions that govern reading and spelling at the sub-lexical level, distinct from sound. For example, knowledge that doublets do not occur at the beginning of the word in English, would be generalized orthographic knowledge. Similarly, knowledge that ‘8’ cannot be part of a word, in Standard English, is generalized orthographic knowledge. Even knowledge that words that end in the /k/ phoneme are usually spelled with a ‘ck’ rather than a ‘k’ or a ‘kc’ or any other combination that could be pronounced /k/, is generalized orthographic knowledge.

Thus, in addition to knowledge about patterns of letters within words, I propose that orthographic knowledge should include rules that govern what kind of symbols make up English words. Feng and Kaefer (under review) argue that children must learn that written language is socially prescribed before they can learn about the combinations within a language. That is, they must learn that only letters can be used to make up words before they learn that these letters must be combined in certain ways to make up words. In this way, generalized orthographic knowledge encompasses a range of
information. In its broadest form, it can include information about what is or is not a legitimate orthographic symbol, while, in its most specified form, it includes knowledge about which patterns of letters can occur and which cannot.

I propose a conceptual model in which generalized orthographic knowledge appears very early in reading development, and continues to build throughout development in more complex forms. Figure 1 shows this conceptual model depicting the development of generalized orthographic knowledge in reading development.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Operational Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized orthographic knowledge</td>
<td>Knowledge of symbol and letter patterns within and across words</td>
</tr>
<tr>
<td>Letter Knowledge</td>
<td>Letter names or sounds</td>
</tr>
<tr>
<td>Phonological knowledge</td>
<td>Knowledge of the sounds of words and sound-letter correspondence</td>
</tr>
<tr>
<td>Lexicalized Orthographic Knowledge</td>
<td>Knowledge of the pattern of letters associated with a particular word</td>
</tr>
</tbody>
</table>

Figure 1: Conceptual model of generalized orthographic knowledge in reading development
In the interest of simplicity, each branch of knowledge is expressed as a distinct and concrete area; of course, in reality, each of these skill sets grows and influences and overlaps with one another. In particular, though generalized orthographic knowledge is shown to occur very early in its broadest form, its nature changes as children gain more experience with the reading process, the shading is an attempt to indicate the growth from a simplistic to a rich and complex source of knowledge. While children may not know about certain patterns of letters before they have significant phonological knowledge, I propose that they show the earliest form of generalized orthographic knowledge at this time and this knowledge continues to develop concurrently with phonological knowledge. This early generalized orthographic knowledge develops before children enter school and gain direct instruction in reading, but continues to develop for much of the lifespan.

Letter knowledge is proposed as a necessary precursor to generalized orthographic knowledge. In its broadest form, generalized orthographic knowledge consists of the understanding of which symbols are appropriate to make up words. In order to gain this generalized orthographic knowledge, children must have a basic level of understanding of letters (i.e., letter knowledge). As the building blocks of words, letter knowledge is an important precursor for generalized orthographic knowledge. If a child is not readily familiar with letters, he or she is likely to consider an “8” as an acceptable symbol for a word. If, however, they have enough knowledge of letters to
accurately distinguish between a letter and another type of symbol, then this ability to discriminate between symbols is a basic form of generalized orthographic knowledge.

Further, as children gain more experience with written language, and receive direct instruction, this basic form of generalized orthographic knowledge develops into a more advanced orthographic knowledge of the patterns of letters within words. This more advanced knowledge can range from the knowledge that ‘k’ cannot be doubled at the end of words to the knowledge that most words ending in /k/ are spelled with a ‘ck’. This knowledge develops concurrently with phonological knowledge, but before lexicalized orthographic knowledge. That is, as children are learning to phonologically decode and spell written words they are also encoding patterns of words and learning which combinations are appropriate and which are illegal.

Phonological knowledge is defined here as the combination of knowledge of the sound elements in words, traditionally known as phonological awareness, and the knowledge of which letter corresponds to said sound, traditionally known as letter-sound correspondence, or phoneme-grapheme correspondence. For the purposes of this dissertation I will refer to this combination of knowledge as ‘phonological knowledge’, and the extent to which a particular word or spelling corresponds to this idealized system will be referred to as ‘phonological legality’. In this way phonological knowledge can act as shorthand for everything children know about the correspondence between spoken and written language at the sublexical level. For example, a child who spells a
word like ‘has’ as ‘haz’ will be said to be relying on phonological knowledge in their production of words. As an extension of this definition, a word that cannot be pronounced as written, like ‘dtam’, will be said to be phonologically illegal, such that it goes against the ‘phonological knowledge’ children have about the breakdown of sounds and the correspondence between sounds and letters. I argue that phonological knowledge develops concurrently with generalized orthographic knowledge. That is, as children are learning to phonologically decode and spell written words they are also encoding patterns of words and learning which combinations are appropriate and which are illegal.

To distinguish generalized orthographic knowledge, lexicalized orthographic knowledge, and phonological knowledge, an example may be beneficial. If you were attempting to spell the word ‘luck’ you could use a number of sources of information to produce it. You could remember that the word is spelled ‘l-u-c-k’ (lexicalized orthographic knowledge), or you could try sounding out each phoneme and associate that with individual representations, perhaps ending up with ‘l-u-k’ (phonological knowledge), or, finally, you could try sounding out each phoneme, associate that with individual representations, and then remember that after ‘u’ the /k/ phoneme is usually represented by a ‘ck’ rather than just a ‘k’ (phonological knowledge combined with generalized orthographic knowledge). Knowing that ‘ck’ can be used to reflect the /k/ sound is phonological knowledge, but knowing that it’s the most likely choice in a
context like this is generalized orthographic knowledge. Similarly, knowing the sound or name associated with a letter is letter knowledge but, being able to put those sounds and letters together in a combination is phonological knowledge. Finally, lexicalized orthographic knowledge is a representation of the pattern of letters of a whole word associated with all the inherent semantic associations.

Furthermore, I predict that generalized orthographic knowledge develops before lexicalized orthographic knowledge. Children’s sight word reading is generally considered to develop after phonological knowledge (Ehri, 1995), but when we consider smaller units, it’s likely that generalized knowledge can be developed before children have lexicalized representation. For example, imagine children encounter a series of words in succession, ‘leash’, ‘learn’, ‘log’. For each of these they must develop an individual representation of the word, as well as the generalized knowledge that words begin with a single consonant. At this point, they have only one example of each word from which to develop a lexicalized representation, but three examples of words beginning with a single ‘l’ (as opposed to a doubled ‘l’, for example). Later, when they encounter a novel word that begins with the phoneme /l/ they can predict that it will be spelled with a single ‘l’ based on their generalized orthographic knowledge, although they can not have a lexicalized representation of the word. Eventually, of course, children will encounter a word like ‘llama’ and their generalized knowledge will not help; they must develop a lexicalized representation. Because this will be the only
example they see with a doubled letter at the beginning of the word it shouldn’t affect their processing of frequency for general orthographic knowledge.

This proposed conceptual model also provides some insights into the possible mechanisms by which generalized orthographic knowledge develops. If the current set of studies show that a more basic generalized knowledge can develop before lexicalized orthographic knowledge, it suggests that any mechanism for the development of this knowledge must not have lexical orthographic knowledge as a prerequisite. Of the current theories of development of orthographic knowledge, the most promising theory is that children are sensitive to the statistical regularities inherent in the patterns of written language. A statistical learning mechanism would fit this criterion and provides a promising explanation for the development or progression of basic to more advanced levels of generalized orthographic knowledge.

If, as I predict, generalized orthographic knowledge develops concurrently with phonological knowledge and before lexicalized orthographic knowledge, it raises the question as to why studies of children’s reading and spelling development have shown children relying on phonological knowledge before orthographic knowledge (Treiman, 1993). I propose that certain literacy tasks lend themselves to a phonological approach and when children engage in these tasks their orthographic knowledge is overshadowed by the phonological demands of the task. Once children are able to integrate their
phonological and orthographic knowledge, and thus have a fully specified representation of a word, the masking of orthographic knowledge no longer occurs.

The current set of studies tests several aspects of the proposed model. First, I examine the origin of generalized orthographic knowledge by testing pre-reading children’s knowledge of the patterns in non-words. The current model predicts that I will find some generalized orthographic knowledge in pre-readers. Next I examine why, if this knowledge develops early, we do not see more evidence of it in early readers’ attempts at decoding and spelling. This model predicts that phonological knowledge may mask generalized orthographic knowledge in certain contexts.

1.2 Generalized Orthographic knowledge and English

Generalized orthographic knowledge is important for reading, particularly in a language like English, whose structure attempts to incorporate a number of different influences. The goal of any writing system is to replicate the communicative function provided by spoken language across the barriers of space and time. To improve translatability between mediums, each written language usually provides some link back to the phonological information of the language (DeFrancis, 1989). Rather than memorizing the visual patterns for each individual word one encounters, alphabetic systems, like English, use a combinatorial system that breaks oral language down into sounds and attempts to represent each of those sounds. Linguistically, phonemes are these individual sounds in a language, and phonology is the pattern and organization of
these sounds. Ideally, all one would only have to know is the individual phonemes of a word and their corresponding letters in order to re-assemble an oral word in the written medium.

Although phonological knowledge is vital in learning to read in most languages, the extent to which one can learn to read based on phonological knowledge alone varies by writing system (Seymour, Aro, & Erskine, 2003). For ‘shallow’ orthographies, like Finnish, the link to phonology is the primary goal of the writing system. In which case, grapheme-phoneme correspondence is extremely regular. Each phoneme has a representative letter and all that is required for learning to read is the knowledge of this correspondence and the ability to blend the sounds together. English, on the other hand, has a deeper orthography, which means it does more than reflect the sound of spoken language, and so the grapheme-phoneme correspondence is deeper. For example, English writing conveys morphological information (Chomsky & Halle, 1968), grammatical information (Richard L. Venezky, 1970), and information about word origin (Richard L. Venezky, 1999). Thus, a word like ‘quiche’ sacrifices sound-to-spelling correspondence in order to preserve the etymological link to the original French word. Though they seem somewhat arbitrary each of these sacrifices preserves the communicability of a language as broadly used and far reaching as English. Thus, the process of reading differs based on the needs of a language (Frost, Katz, & Bentin, 1987; Katz & Feldman, 1983; Share, 2008). It is because of these variations from sound-to-
spelling correspondence that orthographic knowledge, particularly generalized orthographic knowledge, is vital for reading processes.

Take the example of morphology. Written English attempts to maintain morphological links between words; thus, ‘health’ maintains the link ‘heal’ despite the differences in pronunciation (Chomsky & Halle, 1968). These kind of relationships are quite common and it has been estimated that for every word a child learns, one to three additional words should be understandable, based on morphological ties (Nagy & Anderson, 1984). Children in 3rd through 6th grade have been shown to use morphological information in spelling, such that learning to spell a novel word that has morphological ties to a known word is easier than learning to spell a novel word that can only be learned through rote memorization (Waters, Bruck, & Malus-Abramowitz, 1988). Similarly, 4th, 6th, and 8th graders are more likely to spell a derived word correctly when they have previously spelled the root correctly (Carlisle, 1988). This suggests children who have some knowledge of reading and spelling are able to use generalized orthographic knowledge based on the morphological connections between words, to decide how to read and spell. The focus of English on preserving these morphological ties allows for connections between words that are based on meaning. This takes the written symbol system away from a simple reflection of the phonology of English and engages it as a reflection of the many diverse historical and social influences of English.
Despite the value of having a writing system broad enough to include the diversity
of the English speaking population, and the social and historical forces that created it,
the drawback of such a system is that children learning to read and write in the system
can have difficulty (Share, 2008). If children paid attention to only sound-to-spelling
correspondence, and for each phoneme selected the grapheme that is used most often,
73% of graphemes would be correct; however, since most words contain more than one
grapheme, most words would be spelled incorrectly (Hanna, United States. Office of
Education., & Stanford University., 1966). The focus of English writing on morphological
consistency, preservation of original spellings and other conventions means that
children have to follow multiple rules at different levels with conflicting goals. Thus, in
order to successfully read and write in English, readers must incorporate generalized
orthographic knowledge as well as phonological information in order to maximize their
potential. Focusing on memorizing a pattern of letters for each word fails to capitalize on
the ties between spoken and written English, and, by the same token, focusing only on
the reproduction of individual sounds fails to account for the many complexities of the
English language. In order to successfully read and write in English children must
develop sensitivity to patterns of letters that occur across words and be able to
incorporate this knowledge with their phonological knowledge.
1.3 The representation of orthographic Knowledge

In order to distinguish generalized orthographic knowledge from lexicalized orthographic knowledge, it is important to understand how this knowledge is represented. Research in adults has focused on their acquisition of lexicalized orthographic knowledge, and this research hints at orthographic representations stored in an abstract form at the word level. For example, Besner, Coltheart & Davelaar (1984) showed that adults were slower to label lowercase and uppercase versions of non-words as perceptually different than homophonic non-words. They suggest that because the non-words used contained no semantic information, the reason they paired the two perceptually different forms of the non-word was by accessing abstract orthographic codes that are independent of the perceptual cues of individual words. Similarly, McLelland (1976) showed that the word superiority effect, the phenomenon that words are remembered faster in a recognition task than non-words, appears even if the match is a mixed case condition.

Although these studies provide evidence that orthographic information in the mental lexicon is not tied to a simple perceptual encoding of the word form, they haven’t shown that this knowledge is more than perceptual processing at the letter level, rather than at the word level. Theoretically adults could be responding to the individual letters in the pattern, rather than the pattern itself. Bowers, Arguin & Bub (1996) provide evidence against this claim in a masked priming task. They found that letters of different
cases do not prime for one another, it is only words or non-words - patterns of letters -
that prime across perceptually different stimuli. Similarly, Brown, Sharma & Kirsner
(1984) find that priming occurs across scripts for Hindu/Urdu words. Although the
spoken forms of the words are the same, the scripts are different, so any priming effects
are attributed to abstract orthographic knowledge. Because the scripts contain different
letters, this cannot be a simple letter level encoding that appears at the word level also.
By the same token, Bowers and Michita (1998) find that words written in different
scripts of Japanese, hiragana and kanji, prime one another in a masked priming task.
Again this suggests that there must be an abstract orthographic representation for
individual words that transcends perceptual encoding.

These studies suggest that this knowledge is not simply a visual or perceptual
representation of form, nor is it a representation of letters. Instead, this shows evidence
that readers represent a pattern of letters that they then retrieve to identify a word.
These studies examine knowledge about patterns of letters and their association with a
specific word, but the current studies expand this to include knowledge about patterns
of letters that occur across words. Theoretically, if lexicalized knowledge is a
representation of a pattern of letters associated with a specific word, then generalized
knowledge can be thought of as a representation of patterns that occur across words
within a language. Like lexicalized knowledge, it is not simple memory for a visual
stimulus, but a representation of a pattern of letters; however, unlike lexicalized
knowledge it is not associated with a specific word or meaning. Instead, this representation is an abstraction of the patterns within a language.

1.4 The development of orthographic knowledge

The proposed model of the development of generalized orthographic knowledge makes several predictions. First, that generalized orthographic knowledge develops very early in reading development. Second, that it continues to develop concurrently with phonological knowledge and, finally, that it primarily develops before lexicalized orthographic knowledge. This section outlines how current theories of development consider the origin and development of generalized orthographic knowledge, as well as how it reacts with lexicalized orthographic knowledge and phonological knowledge.

1.4.1 Early generalized orthographic knowledge

Although it was once thought children had to have an understanding of letter-to-sound correspondences before they could develop orthographic knowledge (Frith, 1986), when we consider the most basic form of generalized orthographic knowledge, knowledge of what symbols make up words, there is evidence that pre-reading children can show orthographic knowledge. When 5-year-old children are presented with different types of printed stimuli, they are able to label printed words and pseudo-words as writing but refuse to label symbol strings, Chinese characters, and shapes as writing (Lavine, 1977). In another study, when 4-and-5-year-olds were shown printed stimuli along with pictures, squiggles, and shapes, they were able to identify print and
cursive writing as words, but labeled squiggles and pictures as non-word (Bialystok, 1995). These findings suggest that even very young children have an early orthographic understanding of what kinds of symbols belong in a word. The young age of these children suggests that this knowledge develops before children have lexical orthographic knowledge, and even before children have much experience phonologically decoding written words.

Once children gain more experience with written language, there is evidence that their knowledge base grows to include more detailed, but still generalized, knowledge. In addition to being able to identify the symbols that go into words children gain an ability to recognize appropriate patterns of these symbols. Zivian & Samuels (1986) examined this conceptual knowledge of orthographic legality by asking 9-year-old children to choose which of a pair of non-words were “most like a word”. They found that children were more likely to identify orthographically correct letter patterns as ‘good’ words than orthographically incorrect letter patterns. It should be noted, however, that in this study the illegal orthographic combinations were also unpronounceable, thus children who were sensitive to the phonological make-up of the non-word could have used that information to make a decision about which was the ‘best’ word. This limitation was addressed in future studies that looked at illegal orthographic combinations that were also pronounceable. In particular, word initial consonant doublings are a good example of a generalized orthographic rule that crosses
a representation of individual words and refers to an abstract representation of a particular pattern. When kindergarteners are presented with pairs of letter strings, one of which had a doubled letter at the end (e.g. STUFF) and one that had a double letter at the beginning (e.g. FFULP) they recognize that doubled letters better at the end of words than the beginning. However, if they are presented with a pair of words in which one has a legal double at the end (e.g. STUFF), and one has an illegal double at the end (STUKK) they perform at random. First graders, however, are able to identify which doublings are appropriate at the end of words (FF vs KK) (Cassar & Treiman, 1997). Generalized orthographic knowledge is also inferred from tasks that do not require explicit judgments. Wright and Ehri (2007) reinforced this finding by having both kindergarteners and first graders learn to read and recognize real, meaningful words with modified spellings. Children of both age groups required fewer trials to learn orthographically legal spellings than orthographically illegal ones. That is, children have an easier time learning to read and spell a word like ‘PADD’ than a word like ‘RRAG’. Even more impressively, they showed that kindergarten and first grade children, who are relatively advanced readers, will spontaneously regularize the orthography in a spelling test when writing previously learned illegal non-words. That is, when asked to spell a word like ‘RRAG’ they were more likely to produce ‘RAGG’ than ‘RRAG’.

Further evidence for the existence of this type of generalized orthographic knowledge comes from research in other alphabetical languages. Every alphabetic
language has a finite amount of letters, so words get their meaning from patterns of these letters. Recurring patterns of letters occur in every alphabetic language and experience with them should allow children and experienced readers to recognize orthographically legal and illegal patterns. Pacton, Perruchet, Fayol and Cleereman (2001) showed French speaking first through fourth graders two non-words and asked them to select which would make the better word. They found that children in each grade were more likely to select words containing high frequency doublets than low frequency doublets (e.g. mm vs. cc), even though each of the consonants occurred just as frequently individually (e.g. c and m occur equally frequently). Similarly, Lehtonen & Bryant (2005) showed that, in order to complete a sentence, Finnish speaking children were more likely to select a non-word with a single consonant in the word initial position, than a non-word with a doubled consonant in the word initial position. This is particularly interesting because, in Finnish, doublets fulfill a genuine phonological purpose, by indicating a long consonant (or vowel). In the word-initial position, however, there are no doublets, irrespective of the consonant or vowel length. That is, although a word initial consonant may be long when preceded by certain words (the second person singular imperative, for example), it is always represented by a single letter, whereas a word medial consonant is represented by a doubled letter if it is long and a single letter if it is short. The finding that children recognize that word-initial consonants are always a single letter, irrespective of the length of the sound, strongly
suggests that they have a conceptualized knowledge about legal and illegal letter patterns. These studies suggest that children show knowledge about letter patterns shortly after they know about letters, and before they have much experience reading words.

1.4.2 Stage theories of orthographic knowledge development

Traditional developmental theories focus on the progression of knowledge through stages. Most stage theories of reading predict that phonological knowledge precedes orthographic knowledge, though most studies have focused on lexicalized orthographic knowledge. Ehri’s theory of reading development (Ehri, 1995, 2005) focuses on the development of orthographic representations as ‘sight word reading’. Learning to read by sight automatically is essentially the process of accessing a lexicalized orthographic representation and associating it directly with meaning. She theorizes that children learn to recognize individual words in their environment first, creating visual representations for things like ‘Pepsi’. As children receive direct instruction in reading, she theorizes that this representation gets broken down on a phonological level. At this point children begin to be able to sound out new words and the process of sounding out these words leads to the representation of their orthographic patterns. In this way Ehri sees the development of orthographic knowledge as something that happens initially in a simplified form and then slowly develops again as children begin to associate phonology and orthography. Orthographic
knowledge is primarily considered lexicalized in nature, and develops primarily after phonological knowledge is well established.

Similarly, Treiman’s (1997) stage theory of spelling development explains children’s progression from a reliance on phonological knowledge to a knowledge of orthographic rules in four stages. These stages are pre-alphabetic, alphabetic, orthographic, and morphologic, with changes occurring over time. The first stage involves the internalization of sounds in normal spelling, which tests children’s understanding of speech and its relation to print. During the second phase, children tend towards more conventional spellings. At this point they have a larger understanding and are increasingly frequently able to distinguish between correct sound spellings, and when they are correctly used in context. The third stage involves fast learning about correct letter patterns as they begin to generalize their specific knowledge. Finally, children begin to understand the many rules in the English language, and can recognize the consistencies and inconsistencies. This theory suggests phonological knowledge comes well before orthographic knowledge, though, with the inclusion of several interim steps in which children use some spelling conventions, it does allow for the existence of some generalized knowledge that influences children’s production.

The hallmark of stage theories is the quantitative shift in knowledge. Children progress from phonological knowledge to orthographic knowledge. I propose that
evidence for this kind of shift is weakened by the inclusion of generalized orthographic knowledge. According to my conceptual model, generalized orthographic knowledge begins early, and develops concurrently with phonological knowledge. Thus, the shift observed by stage theorists is less a progression from phonological to orthographic knowledge, and more an integration of phonological and orthographic knowledge.

1.4.3 Mechanisms of the development of Generalized Orthographic Knowledge

If, as I predict, generalized orthographic knowledge is shown to develop with phonological knowledge and before lexical knowledge, then it suggests several implications for the mechanisms of development of orthographic knowledge. Any mechanism of development must not require extensive phonological or lexicalized orthographic knowledge. Thus when considering current theories, it is important to focus on the cognitive prerequisites in children.

One theory is that children’s specific knowledge of individual words influences their understanding of general letter patterns through analogy (Goswami, 1988). Goswami and colleagues argue that once children have an orthographic representation of a word they can apply that representation to similar orthographic patterns (Goswami & Byrant, 1992; Goswami, 1993; Ziegler & Goswami, 2005). For example, children can use their knowledge of the word ‘peak’ to read the word ‘beak’. They tested this by having children decode a ‘code word’ that they said would be helpful in decoding other words in the task, in this example, ‘peak’. They found that when children were given a
code word they decoded similar words, like ‘beak’, faster and more accurately than children who had not been given the analogical word to start. Theoretically this can be used as a way of developing general knowledge from specific knowledge. Understanding that ‘peak’ is pronounced /pik/ leads to a general orthographic understanding of the pattern ‘eak’.

There has been some evidence that analogical transfer can happen early in literacy development. Martinet, Valdois and Fayol (2004) tested children with tightly controlled lexical exposure on their orthographic knowledge through traditional and non-word spelling. They found that three months into literacy training, children show both a frequency effect on the words they have seen frequently and an analogy effect in the spelling of non-words. This suggests that children rely on analogy right from the beginning of literacy development. This would lend support to analogical transfer as a mechanism for the development of conceptual or generalized orthographic knowledge.

Analogical reasoning assumes a fairly well-established pre-existing lexicon, and posits that children use the orthographic patterns in this pre-existing lexicon to decipher new words. It makes no provision for how this lexicon develops, and focuses specifically on the acquisition of orthography without a direct prediction of how it interacts with phonological knowledge. Ehri and Robbins (1992) showed that analogical transfer only seems to occur when children already have a grasp of sound-to-spelling correspondence, which means that it’s not clear whether children are using analogical
transfer to build their orthographic knowledge or exploiting analogical transfer for efficiency of decoding. Nation, Allen & Hulme (2001) showed that orthographic transfer occurred whether the code-word was provided visually or orally. This suggests that the analogical transfer may not even be strictly orthographic, but rather a matter of phonological priming. This makes the relationship between phonology and orthography unclear in analogical theories of orthographic knowledge development. Initial studies ignored the phonological context in which children learn to read and follow-ups indicate that phonological knowledge is a necessary requirement for applying analogical reasoning to orthographic learning.

Another mechanism that may explain the development of conceptual orthographic knowledge is statistical learning (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). On a basic level, proponents of a statistical learning mechanism suggest that children process patterns they see over and over again and learn to anticipate frequent patterns. Some work has begun to show that children can use this process to develop orthographic knowledge from what they see in their environment. Theoretically children are sensitive to the relative frequency of letter combinations that they encounter in the environment, and thus they begin to recognize frequent combinations as legal and infrequent combinations as illegal. If this is the case, then they should use this knowledge when deciding how to produce written words. Some evidence for this kind of learning has shown that children are able to recognize
and generalize probabilistic patterns in their environment and use this information as they are learning to produce written language. Treiman and Kessler (2006) showed that children use consonant context to spell vowels. For example, “/æ/” is typically spelled as ‘a’ when it follows /w/, as in wand, but as ‘o’ when it follows other consonants, as in pond. They showed that children have sensitivity to this kind of contextual information and can utilize it when spelling, to select the correct vowel in a particular context. By the same token, Treiman, Kessler and Hayes (2005) showed that children use vowel context when spelling consonants. In fact, the effects of the consistency of vowel spellings in context influence children’s success in spelling vowel patterns over and above the complexity or frequency of the pattern. (Caravolas, Kessler, Hulme, & Snowling, 2005). In this way, children use contextual orthographic cues to help them determine the appropriate spelling of a word. Children are sensitive to the context of individual letters, or, in other words, they are paying attention to the patterns and combinations in which letters operate and they are able to use their knowledge of these patterns to produce written language. Lete, Peereman, and Fayol (2008) provide support for this interpretation when they find that not only do children use context when deciding how to spell, but as their orthographic knowledge increases so does their reliance on contextual information. As children gain more information about appropriate patterns in written language they rely more and more on these patterns to produce written language. Statistical learning theories of spelling and orthographic
patterns are fairly new, and as such there is significantly less evidence, either in support
or against, than for other theories. The relationship between phonological knowledge
and the statistical acquisition of generalized orthographic knowledge hasn’t been
examined. It is unclear whether and how phonological knowledge could help
acquisition, or if the two processes do not affect each other. Similarly, the relationship
between how lexicalized and generalized orthographic knowledge develop has not been
examined. Theoretically, however, because statistical learning theory hasn’t fully
elucidated the prerequisites for orthographic knowledge there is room to argue that
generalized orthographic knowledge could have neither phonological nor lexicalized
pre-requisites.

Neither analogy theorists, nor statistical learning theorists focus on the divide
between lexicalized and generalized orthographic knowledge, but from their research
we can extend their findings to examine how both lexicalized and generalized
knowledge develops. For analogy theories, the necessity of pre-existing lexicalized
representations means that lexical orthographic knowledge must develop first, and then
children can use these representations as example to draw analogies from. Statistical
learning theory, on the other hand, has no prerequisites. It also doesn’t have any
requirements for the size of the pattern learned, which gives the option of dealing with
units smaller than a whole word. Thus, if children use statistical learning, they may have
a grasp of some generalized knowledge before lexicalized representations. If the
predictions made by the current set of studies are borne out, this would provide
evidence for statistical learning theory as the most promising model for the
development of orthographic knowledge.

1.5 Interaction between Phonology and Orthography

It is generally acknowledged that the representation of a single word is
comprised of three constituent elements: orthography (spelling), phonology
(pronunciation) and semantics (meaning) (Coltheart, Curtis, Atkins, & Haller, 1993;
Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perfetti & Liu, 2005; Perfetti, Liu, &
Tan, 2005; Plaut & Booth, 2000; Plaut, McClelland, Seidenberg, & Patterson, 1996; Wang,
Perfetti, & Liu, 2005). Perfetti & Hart (2001) argue that reading skill is the extent to
which a reader’s lexicon contains fully integrated and specified representations. That is,
reading failure is due to representations that do not fully integrate the meaning,
phonology, and orthography of a word. If, as I predict, generalized orthographic
knowledge develops before lexicalized orthographic knowledge, and is used by children
from their earliest encounters with written words, then there is a potential conflict
between early orthographic and phonological knowledge, if they do not integrate the
knowledge as it develops.

Interestingly, the integration and activation of these constituents occurs
differently across writing systems. For example, in alphabetic systems word-level
phonology is activated before an individual has completed specifying individual letter
units, but in Chinese, word level orthography is not completed until the full character is specified (Perfetti & Tan, 1998). Similarly, studies of adult second language learners show that, when an English speaker is learning Chinese, they integrate form and meaning before they learn to integrate phonology (Liu, Wang, & Perfetti, 2007), likely because phonology plays a less direct role in non-alphabetic languages. This is supported by findings that, when second language learners learn a language with an alphabetic script, they integrate phonology and orthography at roughly the same time (Brysbaert, Van Dyck, & Van de Poel, 1999; Van Wijnendaele & Brysbaert, 2002). This evidence from second language learners focuses on knowledge about spoken language and written language. In first language learners, spoken language is learned well before written language. So the question is whether, with this lag in information, readers incorporate the two streams of knowledge from the beginning, or phonological knowledge of the link between individual letters and sounds develop along one track while knowledge of the link between patterns of letters and sounds develops along another.

In order to understand how experienced readers of English deal with the integration of phonology and orthography, streams of information, researchers have created several models of reading that incorporate both orthographic and phonological dimensions. According to dual route models, every time readers encounter a word they have two ways of accessing the corresponding oral word. If readers have a lexicalized
orthographic representation, they can access a lexical orthographic representation and retrieve its stored pronunciation. That is, for a word that one sees frequently, like ‘orthography’, the pattern of letters is stored in memory and associated with the spoken word and meaning. If the word is less familiar, knowledge of sound-to-letter correspondence will be the better and faster route, by assembling the pronunciation by combining the individual pieces (Coltheart, et al., 1993; Coltheart, et al., 2001). In this way the model examines the reading process in terms of memory for lexical form, or assembly of individual parts through phonological knowledge.

On the other hand, connectionist models of reading development propose that readers of English use a single network of association, such that an orthographic representation activates a number of potential pronunciations and the strongest connection is typically the selected one (Seidenberg & McClelland, 1989). As children age, the associations become more elaborate, branching out from a letter-to-sound association being the dominant one to larger orthographic units dominating the model (Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994). In this way as children age the single network grows more and more complex as children gain experience with reading.

Dual route models do not explicitly refer to generalized orthographic knowledge. Because it is at a sub-lexical level, and requires an assembly of parts, generalized orthographic knowledge, as I’ve outlined it, would likely be considered part
of the ‘indirect route’ which also contains grapheme-phoneme correspondence. In this way dual route theories don’t have any prediction for the integration of generalized orthographic knowledge and phonology. Connectionist modeling theories also do not explicitly reference generalized orthographic knowledge, but, because these theories center around the gradual learning of lexicalized orthographic knowledge, generalized orthographic knowledge would seem to be a hidden layer to lexicalized orthographic knowledge. In this case the interaction between phonology and orthography is represented, but because generalized orthographic knowledge is not distinguished from lexicalized knowledge, it is difficult to distinguish their separate interactions with phonology. The proposed conceptual model argues that generalized orthographic knowledge will interact with phonology in a distinct way from lexicalized orthographic knowledge, such that it will develop at the same time as phonological knowledge, but because the two are not integrated, this generalized knowledge will be masked.

Traditional theories of orthographic knowledge frequently argue that phonological knowledge comes before orthographic knowledge, but these theories typically refer to lexicalized orthographic knowledge. In terms of generalized orthographic knowledge, the interaction between phonology and orthography is complex and fluctuates with age. Children gain experience with many aspects of phonology before they learn to speak, but if we consider phonological knowledge (as it pertains to reading development) as including the knowledge of phoneme-grapheme
correspondence, then a unique pattern of knowledge occurs. Early generalized orthographic knowledge aids in the development of phonological knowledge and this in turn leads to increases in orthographic knowledge. The two processes work cyclically to create a comprehensive representation of both phonology and orthography.

Evidence for this cyclical relationship between orthography and phonology has been demonstrated in research on children’s earliest experiences with written language. Children’s earliest word experience is often recognizing their own name in print; before children can read they have been known to recognize letters in their own name, particularly the first letter (Treiman, Cohen, Mulqueeny, Kessler, & Schechtman, 2007). Then, as children explore other literacy activities, Treiman et al (2008), showed that children can apply the sounds of the letters in their name to novel words. Once children know the names of letters, this knowledge can act as a precursor to letter sound knowledge (Foulin, 2005). Thus children’s early experiences identifying their names and learning the names of the letters in their name can begin the process of integrating phonological and orthographic knowledge.

Once children have a basic understanding of sound-to-symbol correspondence, theories on the interaction between phonology and orthography branches in two directions. Although the proposed model suggests that children’s phonological knowledge can mask orthographic knowledge in certain contexts, some theories argue that children can use this phonological knowledge to increase their orthographic
knowledge. Other theories contend that children’s difficulty in recognizing and producing orthographically unusual words is their focus on sound-letter correspondence. On the side that phonology benefits orthography, the self-teaching hypothesis (Jorm & Share, 1983) states that the process of decoding or ‘sounding out’ words gives children the opportunity to attend to and process the orthographic pattern in words. It is generally measured by exposing children to novel words and, following a short or long delay, measuring whether they can select the word they learned from a series of phonologically similar words (Nation, Angell, & Castles, 2007). Oftentimes, a single encounter with a novel word or non-word is enough for children to encode orthographic detail (Share, 2004). In this way, focusing on the phonology of a word can be beneficial to the development of orthographic knowledge. However, it seems that this has less to do with the phonological mechanisms children are using to process written language, and more to do with the focus required by phonologically based decoding. Like decoding, spelling requires focus on individual letters and letter order. When children spell a word, they have a greater memory for its orthographic form and are, again, more likely to pick it out from a series of homophones (Shahar-yames & Share, 2008).

Thus, in this theory, children’s orthographic knowledge (recognizing their own name in print) contributes to their phonological knowledge (the sounds of the letters in their name) which in turn again contributes to their orthographic knowledge (through
self-teaching). In addition to self teaching there is other evidence of phonology positively affecting orthography, and vice versa. Adults find it easier to perform phoneme awareness tasks when the letters and sounds correspond directly than when they don’t (Castles, Holmes, Neath, & Kinoshita, 2003). Ricketts, Bishop and Nation (2008) showed the same effects in children, wherein children remember the orthography of consistent spelling-sound mappings (TROM to rhyme with PROM) better than inconsistent spelling-sound mappings (TROM to rhyme with HOME). Learning individual orthography relies, to a certain extent, on phonological transfer.

In contrast to results shown by the above studies, research on children’s orthographic production – i.e., spelling – has often shown that phonology can override focus on orthographic combinations. As a result, a focus on sound-letter correspondence can mask the orthographic knowledge children have. When asked to create specific spelling combinations, children often use phonology to guide their spelling. Children will often spell clusters like /sp/ (as in spin), with voiced stops (like sbin). In spoken language, the /p/ sound in the initial cluster often sounds more like a /b/ due to the early onset time of the voicing on the vowel and the lack of aspiration between the consonant and the vowel. This directly leads to children’s producing the sound they hear more than paying attention to the orthographic spelling of the word (Hannam, Fraser, & Byrne, 2007). Bowman & Treiman (2008) also found that a discrepancy between orthography and phonology can lead to difficulty learning letter patterns. Children learn
to spell words with a strong sound-spelling link quicker than words with an arbitrary spelling sound link. Additionally, Stage and Wagner (1992) examined young children’s spelling of non-words and found that, for kindergartners, a combination of phonological knowledge and working memory accounted for most of children’s spelling choices. For older children, these factors did not exclusively explain children’s spelling or the link between children’s spelling and decoding.

Increasing the complexity of the relationship, Booth & Cho (2007) used imaging data to show the development of orthographic and phonological representations. They presented 9-15-year-old children with pairs of spoken words and asked them whether the rimes were orthographically the same. This allowed them to have pairs with both orthographic and phonological mismatches (‘mint’ - ‘pint’, ‘jazz’ – ‘has’). They found that older children had more elaborate system for mapping the link between phonology and orthography, and that children who performed the task better, specifically when there was a conflict between orthography and phonology, had more activation in the left inferior frontal gyrus, an area implicated in the hierarchical structuring of cognitive processes (Jubault, Ody, & Koechlin, 2007). This suggests that judgments about orthography are not only about matching orthography, but also includes a process of overcoming a conflict between orthography and phonology.

The evidence seems to show that some phonologically related tasks can aid in the development of orthographic knowledge, while, in other tasks, an emphasis on
phonology can come at the expense of orthography. According to the timeline outlined above, orthography feeds phonology, which then feeds orthography as children gain skill in decoding. I intend to show that at this precise time when children have some letter knowledge and some phonological knowledge, they also have generalized orthographic knowledge and are using phonological knowledge in tandem with their orthographic knowledge. Orthographic and phonological knowledge feed into one another, and this dissertation intends to show that phonological knowledge can mask generalized orthographic knowledge at this point in development as this knowledge develops from implicit to explicit awareness.

1.5 Implicit and Explicit Knowledge

I hypothesized that orthographic knowledge is best understood as a spectrum of insights and understandings that range from the observation that English words are composed of letters to particular letter combinations. While some of these may be subject to conscious reflection, at least among adults and more experienced readers, a large portion of our orthographic knowledge may be implicit. In particular, I hypothesize that the initial state of orthographic knowledge is mostly implicit.

Explicit knowledge can be described consciously, whereas implicit knowledge is not available to conscious awareness, but can be measured in other ways. In memory development, for example, when children have seen a picture before, they are quicker to identify a blurry version of it, even though they have no explicit recollection of seeing
According to Karmiloff-Smith’s representational redescription theory, most knowledge begins implicitly and through a series of representational redescriptions, in which the mind represents its own representations, it becomes available to conscious, explicit awareness (A. Karmiloff-Smith, 1991, 1992).

There are many examples of children’s knowledge progressing from implicit to explicit knowledge. For example, by 5-years-of-age, children know implicitly the rules of when to use ‘a’ and when to use ‘the’ in a sentence, but it is not until 9-or-10-years-of-age that children become consciously aware of when each is appropriate (A. Karmiloff-Smith, 1979). Additionally this has been shown in mathematics, where children will use a mathematical shortcut before they are able to explicitly say that they are using it (Siegler & Stern, 1998), as well as in conservation tasks, where children can use gesture to indicate a correct answer before they can verbally perform the task (Church & Goldin-Meadow, 1986). Finally, eye-tracking has been used to measure children’s implicit knowledge of theory of mind, demonstrating that preschool aged children have an implicit recognition of false-belief before they are able to explicitly pass a false belief task (Clements & Perner, 1994; Ruffman, Garnham, Import, & Connolly, 2001). Thus, according to this theory, where the methods exist to measure implicit knowledge, we should be able to measure children’s earliest knowledge about a concept. In this case, I intend to utilize eye-tracking to examine children’s earliest orthographic knowledge.
1.6 The Current Study

In a series of four studies this dissertation examines the development of children’s generalized orthographic knowledge as it interacts with phonological knowledge and progresses from implicit to explicit knowledge. In the first study, I expand on previous research that demonstrates that children use certain orthographic rules to make decisions about word quality to show that pre-reading children have implicit orthographic knowledge with very little background knowledge of letters and phoneme-grapheme correspondence. A methodological innovation of this study is the use of eye-tracking to gauge the earliest point at which children recognize a difference between a legal and an illegal symbols string. In the second study, I expand on this assessment of children’s implicit orthographic knowledge by examining children’s creative spellings. I hypothesize that children’s invented spellings are reflections of the underlying orthographic knowledge, and, as their reading ability increases, their ability to balance the phonological and orthographic demands increases. In a third study, I hypothesize that, when the task demands phonological decoding, beginning readers will shift attention to individual letters, which hinders access to preexisting orthographic knowledge, while more experienced readers will demonstrate orthographic knowledge. In the last study, the role of phonology in masking implicit orthographic knowledge is tested in a looking task that is similar to that in study 1. I examine how a child’s implicit and explicit knowledge of orthography changes when phonology is introduced. I
hypothesize that adding phonology to a primarily orthographic task will hinder children’s ability to recognize legal and illegal words.
2. Pre-Readers’ orthographic knowledge

If generalized orthographic knowledge and phonological knowledge develop simultaneously we should see evidence of orthographic knowledge among children who are learning the alphabet, before they can read. This study tests pre-readers’ implicit and explicit orthographic knowledge. Being able to read implies lexical orthographic representations. Evidence of orthographic knowledge among pre-readers would suggest that lexicalized orthographic knowledge is not a necessary condition for generalized knowledge about the orthography.

As children begin to learn to read, they already show sophisticated – but often tacit – understanding about the combinations of letters that make words. Cassar and Treiman (1997) showed that early readers have some knowledge of legal and illegal patterns in print, specifically that first and second graders recognize that doublets do not come at the beginning of the word. Additionally, there is evidence that although children can make some explicit judgments about orthographic legality, much of their orthographic knowledge is implicit. When Finnish children begin school they already know that word initial doublets are illegal, though they cannot articulate what a doublet does in Finnish script (Lehtonen & Bryant, 2005). Wright and Ehri (2007) reinforced this finding by having both kindergarteners and first graders learn to read and recognize real, meaningful words with modified spellings. Children of both age groups required fewer trials to learn orthographically legal spellings than orthographically illegal ones.
Pre-reading children also show some fundamental understanding of what makes written words. Lavine (1977) showed that English-speaking 5-year-olds did have a certain sense of what made writing. Letter combinations, number combinations, and other basic linear combinations of simple symbols were classified as words; whereas, scribbles, Chinese characters, and strings of recognizable shapes were classified as non-words. Similarly, Bialystok (1995) showed that when tests require explicit responses, pre-reading 5-year-olds show orthographic knowledge of what symbols make English words, whereas 4-year-olds correctly categorize real writing but are at chance with symbols that are not part of the English orthography.

Although the aforementioned studies with pre-school aged children used explicit judgments, there is reason to believe that much of this general orthographic knowledge is implicit in nature. If children are sensitive to the relative frequency of letter combinations in their environment, as proposed by statistical learning studies (Hayes Treiman, & Kessler, 1996; Treiman & Kessler, 2006), then they may be able to recognize an illegal or unusual combination before they are able to verbally identify it. Research by Lavine (1977) and Bialystok (1995) strongly suggests that it may be profitable to explore implicit orthographic knowledge among young pre-school children.

In the current study I exploit children’s sensitivity to illegal or unlikely combinations by measuring the time they spend looking at symbol strings. If children encounter a combination of symbols that violates their implicit expectations about the
English orthography, they should look longer at it (Baillargeon, 1987). Thus a difference
in looking times between legal and illegal combinations would suggest an *implicit*
understanding of orthographic knowledge, which is independent of whether or not
children can *explicitly* state that this combination is inappropriate. Moreover, the
increased looking time should concentrate on the specific combinations of symbols that
violate the English orthography.

In this study, I created four levels of general orthographic violations, testing
children’s understanding of both fundamental combinatorial rules, and pattern based
rules like illegal letter combinations: 1) Whole-word symbolic violations, for example, an
entire string of numbers; 2) Part-word symbolic violations, for example, a number
embedded in a letter string; 3) combinatorial violations, for example, a single letter
repeated for an entire string; and 4) pattern violations, for example, a string of varied
letters that violates English spelling rules. In order to ensure that children are using
generalized orthographic knowledge to make decisions about these patterns,
participants were limited to pre-readers, tested before they have lexicalized
orthographic knowledge. The youngest group studied in previous reports is 4-year-olds.
Here I tested even younger children. Additionally, non-words were used to ensure that
lexicalized knowledge could not be used to make judgments.

Furthermore, to link children’s emergent orthographic knowledge with their
early reading skills, I take several measures of early literacy. I tested children’s letter
name knowledge, phonological awareness, and symbol understanding as measures of literacy skill. Past research points to strong links between letter knowledge and orthographic knowledge (Treiman, 1993). The connection between phonological awareness and orthographic knowledge is also well documented (Bus, 1999). In addition to these two emergent literacy skills, recent studies highlight the importance of early conceptual understanding of the nature of writing (Bialystok, 1993; Bialystok & Martin, 2003). Understanding that a written word has the same meaning in varying contexts is an important prerequisite of learning to read (Feng & Kaefer, under review), and, in order to measure this component of emergent literacy, we utilize Bialystok’s moving word task as a measure of the understanding of written words as symbols.

I hypothesize that, as in previous studies, 5-year-olds will be able to explicitly reject symbolic violations of orthography, such as numbers replacing letters, but be unable to explicitly state that pattern violations, like an illegal consonant cluster, is wrong. Based on Bialystok’s (1995) finding with 4-year-olds, I also hypothesize that 3-year-olds will be unable to explicitly identify symbolic, combinatorial, or pattern violations of orthography but that they will show an implicit understanding of the symbolic and combinatorial violations.

2.1 Methods

2.1.1 Participants
Twenty-two 3-year-old (13 girls 9 boys, mean age 3.3 years) and twenty-two 5-year-old (17 girls, 5 boys, mean age 5.5 years) typically developing native English speakers participated in this study. They were identified as non-reading by their parents and were unable to read any of the stimuli in the study. An additional 3-year-old was tested and not included in data analysis because of a language delay, as well as five 5-year-olds, who were excluded due to high reading ability. The ethnic composition of subjects reflected the surrounding community, with 14% of subjects of African American descent, 5% of Hispanic descent, and 2% of Asian descent.

2.1.2 Apparatus

Children were presented with a series of pseudo-words and symbol strings on a computer monitor attached to a Tobii eye-tracking system by Tobii Technology AB (Tobii 1750). It is a remote eye tracking system that has no contact with the reader. The system uses infrared cameras to automatically capture eye images from a reading distance. The system uses infrared video cameras to automatically acquire the pupil positions at a sampling rate of 50Hz. The typical spatial accuracy, measured by repeated calibrations, is approximately 1 visual degree. A 5-point calibration was used in all studies.
2.1.3 Materials

The symbol strings presented to children were as follows. Appendix 1 shows all words used in the study:

1) pseudo-words – these were letter strings like ‘WURGS’ that could have been a legal English word;

2) illegal letter strings – these were composed of letters, but contained illegal combinations of letters like ‘YLRKS’ that violate the English orthography;

3) repeated letter strings – these were composed of the same letter repeated 5 times for example ‘AAAAA’;

4) symbol strings – these were composed of a random assortment of symbols, for example ‘#%&*@’

5) number strings - these were composed of a random assortment of numbers like ‘89274’

6) letter/symbol strings – these were composed of one symbol surrounded by letters, for example ‘TR#MP’;

7) letter/number strings - these were composed of one number surrounded by letters, like ‘A8PER’.

Children’s emergent literacy skills were measured using 3 tasks. Phonological awareness was measured using the CTOPP elision task (Wagner, Torgesen, & Rachotte, 1999), which asks children to remove sounds from words. For example, children are
asked to remove the “brush” from “toothbrush” or the “d” from “powder”. Letter knowledge was measured using Marie Clay’s (1977) test of letter knowledge in which children are asked about each of the letters one at a time and asked to name them. Finally children’s understanding of written words as symbols was measured using the moving word paradigm (Apperly, Williams, & Williams, 2004; Bialystok, 1991, 1995, 1999; Bialystok & Martin, 2003). The moving word task measures whether children understand that a word means the same thing in different contexts. In this task, children are shown a word, like ‘MOUSE’, and the experimenter reads it aloud. The word is then placed in front of a corresponding picture. A disturbance on the screen moves the word in front of another picture, and, at this point, the child is asked what the word says. If the children know that it still says mouse, it suggests they have and understanding of how words function as symbols.

2.1.4 Procedure

Children were introduced to a cartoon bear named Fiffi. Children were told that unlike a normal bear, Fiffi only liked to eat words, and before the child and experimenter played any other games I would have to find some words for Fiffi to eat. To begin, children were given an example of a word (“CAT”) and a non-word (“#%$”) and instructed that Fiffi only liked to eat the word because it means something. Children were then presented with the pseudo-words and symbol strings, one at a time, in
random order, and asked whether they thought that was a word and whether Fifi
would want to eat it.

After the lexical decision task was complete, children performed the emergent
literacy tasks. First, children performed the moving word task. After the moving word
task, children performed the phonological awareness task and the letter knowledge task
in random order.

2.2 Results
2.2.1 Behavioural results
I measured pre-reading children’s behavioural responses to the question: “Is this
a word?” Pseudo-words are phonologically and orthographically legal. Given that these
children could not read, they have every reason to label pseudo-words as words. Thus,
children’s responses to all other symbol strings are compared to their responses to
pseudo-words. If children are significantly less likely to label a symbol string as a word
compared to a pseudo-word, it suggests that they explicitly detect the orthographic
violation. A 6 by 2 repeated measures analysis of variance (ANOVA) showed a
significant main effect of condition, $F(6, 37) = 21.78, p < .001$, a main effect of age, $F(1, 42) = 13.29, p < .001$ and a condition by age interaction $F(6, 37) = 10.57, p < .001$. Simple
contrasts, comparing each condition to the pseudo-word condition, showed a significant
main effect of Condition and a significant Condition by Age interaction for each
condition but illegal letter strings. That is, in comparison to the pseudo-word condition, children are overall less likely to label letter strings, \( F(1,42) = 56.05, p < .001 \), symbol strings, \( F(1,42) = 100.734, p < .001 \), number strings, \( F(1,42) = 54.53, p < .001 \), letter/symbol strings, \( F(1,42) = 70.052, p < .001 \), and letter/number strings, \( F(1,42) = 40.83, p < .001 \).

Similarly, in comparison to the pseudo-word condition this effect varies by age for letter strings, \( F(1,42) = 11.14, p = .002 \), symbol strings, \( F(1,42) = 6.41, p = .015 \), number strings, \( F(1,42) = 22.24, p < .001 \), letter/symbol strings, \( F(1,42) = 30.07, p < .001 \), and letter/number strings, \( F(1,42) = 18.42, p < .001 \). Figure 2 shows children’s behavioural responses.

![Figure 2: Percent of children answering ‘yes’ to the question “is this a word?” by condition and age (plus standard error). A * indicates significance at the .05 level in comparison with the pseudo-word condition.](image)

Three-year-olds: Follow-up t-tests show that 3-year-olds judged that illegal letter combinations, such as YLRKS, were a ‘word’ just as often as a pseudo-word such as
WURGS, \( t(21) = .826, p = .418 \). They also judged that letter/number combinations, such as A8PER, and letter/symbol combinations, such as TR#MP, were words as often as pseudo-words, \( t(21) = 1.86, p = .08 \) and, \( t(21) = 1.86, p = .08 \) respectively. However, they did not label symbol strings or single letter strings as words, \( t(21) = 4.85, p < .001 \) and \( t(21) = 3.28, p = .004 \), respectively. Three-year-olds were also marginally significantly unlikely to label number strings as words as often as pseudo-words, \( t(21) = 2.017, p = .06 \).

*Five-year-olds*: Similarly, Follow up t-tests show that 5-year-olds judged that illegal letter combinations, such as YLRKS, were a ‘word’ just as often as a pseudo-word such as WURGS \( t(21) = 1.00, p = .349 \). However they rejected letter/number combinations such as A8PER, letter/symbol combinations, like TR#MP, symbol strings, like *%&#@, number strings, like 89247, or single letter strings, such as AAAAA \( t(21) = 6.472, p < .001 \); \( t(21) = 10.81, p < .001 \); \( t(21) = 9.91, p < .001 \); \( t(21) = 8.07, p < .001 \); and \( t(21) = 6.988, p < .001 \) respectively.

Taken together, the behavioural results suggest that 5-year-olds understand explicitly that anything composed of non-letters is not a word but believed any combination of varied letters is a word. Three-year-olds, on the other hand, believe that anything with letters is appropriate to make up a word, even if it has other elements in it.
2.2.2 Eye-tracking results

When children are presented with a symbol string composed of either letter combinations, numbers, symbols or a repeated letter, their looking time at the average individual symbol depends on the type of symbol. A difference in looking times between conditions suggests an implicit knowledge of the violation of orthographic rules. A 6 by 2 repeated measures ANOVA showed a significant main effect of Condition $F(6, 37) = 6.82, p < .001$, and a marginally significant Age by Condition interaction $F(6, 37) = 2.003, p = .09$. A set of simple contrasts comparing each condition to the pseudo-word condition, showed a significant main effect of condition for the letter/symbol strings $F(1,42) = 16.34, p < .001$ and the letter/number strings $F(1,42) = 16.90, p < .001$, a marginally significant main effect of condition for illegal letter strings, $F(1,42) = 3.259, p = .078$, and number strings $F(1,42) = 3.38, p = .073$ and a significant condition by age interaction for illegal letter strings, $F(1,42) = 8.06, p = .007$, and repeated letter strings, $F(1,42) = 6.34, p = .016$. Figure 3 shows children’s average looking time at the individual elements of the symbol string.
Figure 3: Average amount of time children spend looking at the symbol string by condition and age (plus standard error). A * indicates significance at the .05 level in comparison with the pseudo-word condition.

*Three-year-olds:* A series of follow up t-tests show that 3-year-olds look longer at the symbol in a letter/symbol combinations, i.e. the ‘#’ in ‘TR#MP’ than they do a letter in the corresponding position in a pseudo-word, i.e. the ‘R’ in ‘WURGS’, $t(21)=2.11, p = .047$. By the same token they look longer at the number in a letter/number combination, i.e. the ‘8’ in ‘A8PER’ than they do a letter in the corresponding position in a pseudo-word, i.e. the ‘U’ in ‘WURGS’, $t(21)=2.12, p = .046$. On the other hand, 3-year-olds looked at the individual elements in a symbol strings, such as &*$#@, number strings, like 89258, and repeated letter strings, such as AAAAA, approximately the same amount of time as the individual elements in pseudo-words $t(21)=.353, p = .728; t(21)=1.41, p = .174; and t(21) = 1.52, p = .144$. Three-year olds also look significantly longer at the
individual elements of pseudo-words, such as WURGS than the individual elements of illegal letter strings like YLRKS $t(21)= 2.537, p = .019$.

*Five-year-olds:* Like 3-year-olds, 5-year-olds looked longer at the non-letter in a letter/symbol combination than they did at an individual element in a corresponding position in a pseudo-word, $t(21)= 2.66, p = .015$. By the same token they looked longer at the number in a letter/number combination than they did in a letter in the corresponding position in a pseudo-word, $t(21)= 3.75, p = .001$. They also looked longer at an individual letter in a repeated letter string, such as AAAAA, than individual elements in a pseudo-word $t(21)= 2.26, p = .034$. On the other hand, 5-year-olds looked at the individual elements in a symbol strings, number string, and illegal letter strings, approximately the same amount of time as the individual elements in pseudo-words $t(21)= .694, p = .495$; $t(21)= 1.31, p = .203$; and $t(21)= 1.284, p = .213$.

**2.2.3 Emergent Literacy**

In order to examine the relationship of both implicit and explicit orthographic knowledge with a more traditional understanding of emergent literacy, I used children’s performance on the emergent literacy tasks to predict children’s implicit and explicit orthographic knowledge as assessed by the lexical decision task.

Table 1 shows zero order correlations between the three emergent literacy tasks and explicit knowledge (judgment of whether it is a word) of the lexical decision tasks.
These results suggest a relationship between each of the emergent literacy tasks and all of the lexical decision tasks in which children show explicit knowledge.

**Table 1: Correlation of explicit tasks with emergent literacy tasks**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Letter Knowledge</th>
<th>Phonological Awareness</th>
<th>Symbolic Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo-word</td>
<td>-.175</td>
<td>-.222</td>
<td>.137</td>
</tr>
<tr>
<td>Symbol string</td>
<td>-.683**</td>
<td>-.381*</td>
<td>-.425**</td>
</tr>
<tr>
<td>Number string</td>
<td>-.642**</td>
<td>-.486**</td>
<td>-.684**</td>
</tr>
<tr>
<td>Repeated letter</td>
<td>-.503**</td>
<td>-.475**</td>
<td>-.348*</td>
</tr>
<tr>
<td>Symbol/letter string</td>
<td>-.612**</td>
<td>-.327*</td>
<td>-.506**</td>
</tr>
<tr>
<td>Number/letter string</td>
<td>-.514**</td>
<td>-.277 †</td>
<td>-.317*</td>
</tr>
<tr>
<td>Illegal combination</td>
<td>-.229</td>
<td>-.175</td>
<td>.157</td>
</tr>
</tbody>
</table>

† p < .10, *p < .05, **p < .01

Table 2 shows zero order correlations between the three emergent literacy tasks and implicit knowledge (looking time) on each of the lexical decision tasks. These results only suggest a relationship between each of the emergent literacy tasks and the knowledge that numbers do not make up words, as well as the knowledge about the letter combinations that are appropriate in words.
Table 2: Correlation of implicit looking time with emergent literacy

<table>
<thead>
<tr>
<th>Condition</th>
<th>Letter Knowledge</th>
<th>Phonological Awareness</th>
<th>Symbolic Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo-word</td>
<td>.120</td>
<td>-.055</td>
<td>.251</td>
</tr>
<tr>
<td>Symbol string</td>
<td>-.02</td>
<td>.053</td>
<td>.087</td>
</tr>
<tr>
<td>Number string</td>
<td>.179</td>
<td>-.014</td>
<td>.264</td>
</tr>
<tr>
<td>Repeated letter</td>
<td>.132</td>
<td>.145</td>
<td>.300†</td>
</tr>
<tr>
<td>Symbol/letter string</td>
<td>.116</td>
<td>.022</td>
<td>.031</td>
</tr>
<tr>
<td>Number/letter string</td>
<td>.268†</td>
<td>.097</td>
<td>.394*</td>
</tr>
<tr>
<td>Illegal combination</td>
<td>.311*</td>
<td>.228</td>
<td>.372*</td>
</tr>
</tbody>
</table>

† p < .10, *p < .05, **p < .01

Although the raw data for explicit knowledge is a series of yes/no questions, once the raw data has been merged into categories each child receives a score on a scale of 0 to 1; thus, table 3 shows results from a series of multiple correlations, with age partialed out. Children’s letter knowledge successfully predicts children’s explicit orthographic knowledge even when accounting for age.
Table 3: Standardized regression coefficients for emergent literacy tasks and explicit orthographic knowledge controlling for age

<table>
<thead>
<tr>
<th>Condition</th>
<th>Age</th>
<th>Letter Knowledge</th>
<th>Phonological Awareness</th>
<th>Symbolic Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo-word</td>
<td>098</td>
<td>-.358†</td>
<td>-.358†</td>
<td>.451*</td>
</tr>
<tr>
<td>Symbol string</td>
<td>162</td>
<td>-.778**</td>
<td>-.029</td>
<td>-.143</td>
</tr>
<tr>
<td>Number string</td>
<td>-.277†</td>
<td>-.319*</td>
<td>-.105</td>
<td>-.237</td>
</tr>
<tr>
<td>Repeated letter</td>
<td>-.220</td>
<td>-.257</td>
<td>-.217</td>
<td>-.014</td>
</tr>
<tr>
<td>Symbol/letter string</td>
<td>-.471*</td>
<td>-.293†</td>
<td>.195</td>
<td>-.20</td>
</tr>
<tr>
<td>Number/letter string</td>
<td>-.415*</td>
<td>-.324†</td>
<td>.071</td>
<td>.05</td>
</tr>
<tr>
<td>Illegal combination</td>
<td>.357</td>
<td>-.521*</td>
<td>-.287</td>
<td>.345</td>
</tr>
</tbody>
</table>

† p < .10, *p < .05, **p < .01

Specifically, having high scores on letter knowledge, predicts labelling non-words as non-words for pseudo-words, symbols, numbers, symbol/letter strings, number/letter strings, and illegal letter combinations; whereas, phonological awareness symbolic meaning only predicts labelling a pseudo-word a non-word.

On the other hand, when accounting for age, emergent literacy skills do not significantly predict implicit orthographic knowledge on many tasks. Only symbolic meaning predicts average looking time at a pseudo-word or a number string.
Table 4: Standardized regression coefficients of emergent literacy and implicit orthographic knowledge

<table>
<thead>
<tr>
<th>Condition</th>
<th>Age</th>
<th>Letter Knowledge</th>
<th>Phonological Awareness</th>
<th>Symbolic Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo-word</td>
<td>-.564*</td>
<td>.181</td>
<td>-.109</td>
<td>.540*</td>
</tr>
<tr>
<td>Symbol string</td>
<td>-.298</td>
<td>-.108</td>
<td>.082</td>
<td>.319</td>
</tr>
<tr>
<td>Number string</td>
<td>-.596*</td>
<td>.303</td>
<td>-.073</td>
<td>.469*</td>
</tr>
<tr>
<td>Repeated letter</td>
<td>.064</td>
<td>-.156</td>
<td>-.012</td>
<td>.346</td>
</tr>
<tr>
<td>Symbol/letter string</td>
<td>-.288</td>
<td>.250</td>
<td>.007</td>
<td>.053</td>
</tr>
<tr>
<td>Number/letter string</td>
<td>-.046</td>
<td>.051</td>
<td>-.052</td>
<td>.388</td>
</tr>
<tr>
<td>Illegal combination</td>
<td>-.150</td>
<td>.190</td>
<td>.070</td>
<td>.300</td>
</tr>
</tbody>
</table>

† p < .10,  *p < .05, **p < .01

2.3 Discussion

This study examines the development of pre-reading children’s emergent knowledge of orthography as it manifests in both implicit and explicit tasks. To this end we can draw several conclusions. Concurrent with our hypotheses, and replicating previous findings (Bialystok, 1995; Lavine, 1977), 5-year-olds identify letter strings as ‘words’ while any combinations including non-letters or repeated letters are identified
as non-words. Additionally, I find that although 3-year-olds do not significantly
distinguish between number or symbol-letter combination strings and pseudo-words
when determining a symbol’s quality as a word, they do look significantly longer at a
number or symbol in the combination string than a letter in the same position of a
pseudo-word. This suggests an early implicit orthographic knowledge in 3-year-olds.

Children who have greater emergent literacy skills are more likely to explicitly
claim that words with symbols, numbers, or repeated letters are non-words. By the same
token, children who have greater symbol knowledge are more likely to take their time
examining legal letter combinations and numbers, showing an implicit understanding of
the difference between illegal and legal letter combinations. This suggests that a general
knowledge of the nature of the symbolic properties in words, letter knowledge, and
phonological awareness ties directly to explicit orthographic understanding. On the
implicit side, however, symbolic knowledge is only correlated with implicit knowledge
of legal letter combinations and number combinations. None of these precursors fully
account for children’s performance, which suggests that, although built upon a
foundation of emergent literacy, generalized orthographic knowledge is more than the
sum of several emergent literacy tasks. Knowing that different letters must be paired
together and cannot contain another symbol to classify as a word is a unique skill for
pre-reading children.
Overall these results suggest a progression in pre-reading children’s orthographic knowledge from implicit to explicit understanding. When 3-year-olds are unable to explicitly state that a number/letter combination string makes a poor word, their eye-movement data register the ‘special’ nature of this string, and they look longer at the number than they do with a letter in a pseudo-word that follows English orthography. By the same token, 3-year-olds, who are explicitly able to judge that a symbol string makes a poor word, show no extra looking time at the symbols, compared to a pseudo-word. Once children are aware of the orthographic violation, they can proceed to make explicit judgments without spending additional time to inspect the word.

Limitations to this study include the general bias of children to say ‘yes’ when asked if a non-word makes a good word. Although this can be worked around statistically, it would be preferable if children’s baseline were chance, rather than a specific default answer. In future studies, I utilize a forced choice task so children must choose a specific non-word as better, rather than merely state yes or no. The reliance on verbal responses is another difficulty, as children are likely making their decisions prior to verbalizing them. A process by which children can press a button to select make a decision may streamline our ability to measure children’s knowledge.

This study has shown the earliest signs of children’s fundamental knowledge of orthography. In addition to showing that pre-reading 5-year-olds are able to use
generalized orthographic knowledge to make a decision about the quality of a word, I have shown that even 3-year-olds show an implicit understanding of some of these rules. I have shown that children develop an understanding of what elements make up words before they receive direct instruction in reading, and that this understanding begins implicitly before moving to explicit awareness. This suggests that generalized orthographic knowledge can develop before children have lexical representations for words.
3. Orthographic knowledge in creative spelling

Chapter 2 has shown that children enter kindergarten with some basic orthographic knowledge before they learn to read. I have also shown that some orthographic knowledge is implicit, and children know more than what they demonstrate in behavioural tasks. Considering that children have some orthographic knowledge before they learn to read, we would expect to see generalized orthographic knowledge play a role throughout reading development; however, the nature of the English writing requires that phonological and orthographic information be integrated to have a fully realized representation of written language. Thus, I expect to find that early readers favour one source of information over the other when producing and decoding written language.

The goal of this study is to determine the nature of children’s bias when it comes to the interplay between phonological and orthographic rules governing written language, once they begin to receive direct instruction in reading and have more experience with grapheme-phoneme correspondence rules. Given the implicit nature of generalized orthographic knowledge at this stage of literacy development, it is likely that when children are faced with a task requiring them to translate their spoken language into written words, they focus on translating the sounds one-by-one rather than integrating their knowledge of appropriate combinations. Later in literacy
development, children have to shift from their attention from sound-letter correspondence to larger orthographic units (Ziegler & Goswami, 2005).

Previous studies have shown that children rely heavily on the phonological structure of words when learning to spell (Read, 1986; Treiman, 1993). Stage theories of spelling development also emphasize children spelling based on sound-to-letter correspondence before they learn to take orthographic rules into account (Ehri, 2005; Frith, 1985). Learning how children progress from their reliance on phonological structure to an understanding of orthographic structure when producing written language will give us the ability to determine when they explicitly understand the orthographic rules that govern written language. Unlike reading, writing is the reflection of the child’s communicative intent, and their creative spellings can be used as a method of determining their pre-existing biases when it comes to orthography. When forced to make a decision about the “goodness” of a word children must focus on the understanding of what makes a good word. However, another element of interest is what knowledge children show when they are focused on communicating ideas, rather than thinking directly about word form. Theoretically, if children are surprised to see illegal letter combinations, they should be unwilling to produce them. If, however, they are not focusing on the combination, but just transcribing each sound as they hear it, they may produce letter combinations that adhere to phonological rules rather than orthographic ones. I hypothesize that, when I isolate children’s spelling of individual
combinations in which there is conflict between orthography and phonology, I will find that as children learn to spell they will first follow the phonology then learn the orthography.

In order to examine children’s creative spelling of orthographically and phonologically legal and illegal letter combinations the current study examines children’s creative spellings over 2 years of elementary school. I isolate children’s production of individual clusters and track children’s production of these clusters through their early reading development. I also associate their production with children’s reading and writing scores from kindergarten and first grade.

In order to narrow my analysis, I am focusing on 2 situations in which phonology, defined above as a reflection of sound-to-letter correspondence is in conflict. First, I examine a conflict in which a single sound is represented by multiple letters. Phonologically speaking the transfer between phoneme and grapheme would ideally be one-to-one. There are, however, several instances in English where the orthography requires a single phoneme to be represented by a combination of graphemes. This section examines how children handle situations where two letters are required for one sound. In other words, I am holding the orthographic legality constant and varying the phonological legality of the pairing. The specific examples I look at are:

1) Word final stop/vowel/stop cluster omissions (‘rowbt’ for ‘robot’)– When spelled correctly, stop/vowel/stop clusters are in line with a single sound-letter
correspondence. When the vowel is omitted however, one of the stop consonants must be pronounced as a syllable, which makes errors of this type a violation of this rule. Thus, these clusters are both phonologically and orthographically illegal, and do not occur within syllable in typical spelling.

2) Word final stop/vowel/liquid or stop/liquid/vowel cluster omissions (‘watr’ for ‘water’ or ‘littl’ for ‘little’) – When spelled correctly, these combinations violate the single sound-letter rule. The ‘e’ is superfluous, from a phonological standpoint. Thus, when an error of omission is made, the combination is phonologically better than the correct spelling. Thus, these combinations are phonologically legal, that is they can be pronounced as written, but orthographically illegal. Although a combination like ‘tr’ is legal at the beginning of a word, it doesn’t appear in the word final position.

The next section deals with children’s spelling of morphological suffixes. One of the things English as a written language values over phonology is the marking of morphological structure (Richard L. Venezky, 1999). There is some evidence that children attend to morphology in their spelling such that children are better able to spell words that share a morphological root (Bourassa, Treiman, & Kessler, 2006; Senechal, 2000; Treiman, Cassar, & Zukowski, 1994). In addition, there is evidence from languages like French that children spell morphological suffixes better than words with irregular endings. (Pacton & Deacon, 2008; Pacton, et al., 2001). Although it was at first thought that the use of morphology in spelling was fairly advanced, recent work shows that
children are sensitive to morphology early in the development (Pacton & Deacon, 2008). To that end, morphological suffixes are based on broad rules and are orthographically consistent across different phonological pronunciations of the words. Though there are exceptions, the general orthographic rule accompanying these suffixes is fairly simple and salient. In this way, the spelling of morphological suffixes, are consistent across different phonological pronunciations, but have a clear and easily applicable orthographic rule. For example, the suffix ‘-s’ is added to nouns to indicate plurality, and often represents both /s/ and /z/ at the end of a word. Using conventional sound-to-letter correspondence, when the suffix ‘-s’ is used to represent /z/, it indicates a phonology/orthography mismatch. Similarly, the suffix ‘-ed’ is added to verbs to indicate past tense, and the ‘d’ often represents both /d/ and /t/. Again, when the suffix is used to refer to /t/ it creates a mismatch based on conventional sound-to-letter correspondence. In this analysis I compare children of different age groups on their naturalistic production of both phonologically matched and mismatched morphological suffixes.

In this study, I also examine the difference in performance for high and low frequency words. Several studies have shown that children are better at spelling elements of high frequency words than low frequency words (Caravolas, et al., 2005). If we consider a mechanism like statistical learning or analogy to be the driving force of development of generalized orthographic knowledge then we may expect that children
have more experience with high frequency words, and thus, their generalized orthographic knowledge is based on these words. This may mean that when spelling these words they perform better, due to their familiarity.

This study has an overall hypothesis that children’s general orthographic knowledge will be masked by phonological concerns when there is a conflict between phonology and orthography. This will manifest in specific ways in each of the two analyses. First, children will be less likely to omit a vowel in a stop/vowel/stop combination than a stop/vowel/liquid combination. Because stop/liquid combinations are pronounceable and phonologically appropriate, children will only not produce them if they recognize that these combinations do not occur at the end of words. On the other hand, stop/stop combinations are not pronounceable as written and, as such, are not phonologically legal. Children will not produce these combinations if they have either relevant generalized orthographic knowledge or phonological knowledge. Second, children will spell morphological suffixes better when they are phonologically aligned, and when they are not aligned children will be more likely to make mistakes towards the phonological end. On the frequency side, I expect that children will spell both of the above combinations better in high frequency words than low frequency words.
3.1 Methods

3.1.1 Participants

3.1.1.1 Longitudinal subject pool

Participants were 65 Kindergarten students from Chapel Hill area elementary schools. There were 30 females and 35 males, and the mean age at the time of first testing was 5 years, 8 months. Kindergarten students were selected because Kindergarten is the first opportunity children have to regularly produce creative written language as they are beginning to receive direct instruction in reading. The ethnicity breakdown of the sample was 14% Asian American, 5% African American, and 9% Hispanic.

3.1.1.2 Non-longitudinal subject pool

23 Kindergarten (mean age, 5 years, 6 months at time of first collection, 13 female, 10 male) and 82 First grade (mean age, 5 years, 9 months at time of first collection, 39 female, 43 male) students from Durham and Chapel Hill area elementary schools. Kindergarten students were selected for the same reasons as in the longitudinal sample. First grade students were selected because this is the next step after kindergarten and children still have problems with spelling. In this group the ethnicity breakdown was 18% Asian American, 24% African American, and 11% Hispanic.
3.1.2 Procedures

Teachers were asked to collect and put aside all writing that children produced without direct instruction on form. No control was made for topic, quantity or format of children’s writing. They included productions such as journal entries, short stories, and captions to pictures. Approximately once per month a research assistant collected these samples, and they were scanned in the school and returned immediately thereafter to the classroom. As data were collected, they were put into a computer database. From this database, research assistants transcribed the handwriting using a web-based coding system. The coding system in place uses 52 letter identifiers, both lowercase and capital. There are also symbols for letter reversals, incorrect capitalizations, incorrect spacing, and so on. The writings were initially transcribed as seen on the page. This includes capitalization, spacing (and lack thereof), and mirrored letters. Most importantly, coders transcribed exactly what was written rather than what the child may have intended. In the next step, the dictionary spelling of each word was entered into the database. Children were not penalized for grammatical errors, so just the spelling of the word they wrote was considered. For example, if a child wrote “we was playeeng,” the only error is the spelling of ‘playing’.

3.1.3 Analysis

The computer program GAWK was used to search the transcriptions and pull out all instances of the target spelling combinations from either the direct transcription
or the correct spelling interpretation. All instances of the target spellings were coded by hand into error categories by two researchers.

The current study examines 2 examples of letter combinations that each demonstrates the interaction between phonology and orthography and how the balance can shift between tasks. First I look at examples of situations where there is a phonological violation of the principal of single-sound to letter correspondences. That is, the orthography requires a single phoneme to be represented by a combination of graphemes. Next, I compare children of different age groups on their naturalistic production of both phonologically matched and mismatched morphological suffixes.

3.2 Results

3.2.1 Single sound-single letter correspondence

For these analyses, the longitudinal and non-longitudinal samples are combined. Figure 4 shows the proportion of errors children make on each type of combination in Kindergarten and 1st grade. This shows that kindergarteners, $t(47) = 6.18$, $p < .001$, and 1st graders, $t(82) = 2.74$, $p = .007$ produce stop/liquid clusters far more than stop/stop clusters. However, while both kindergarteners and 1st graders drop the vowel to produce stop/stop clusters about equal proportions of time, $t(170) = 1.38$, $p = .17$,

Kindergarteners drop the vowel to produce significantly more stop/liquid clusters than first graders $t(104) = 3.56$, $p = .001$. 
The next step was to examine the types of words children make errors on. Does a word that occurs frequently become easy to spell orthographically or does the nature of the phonology/orthography tension make the frequency of the individual words irrelevant? In order to test this I isolated the most popular words that end in a stop/vowel/stop combination and ones in a stop/vowel/liquid combination. In the Stop combinations the most popular words are: got, but, get, did, dad, put, big, and dog. These accounted for 51.1% of the stop/vowel/stop spellings. The most frequent of these words, ‘got’, ranks 30th in overall frequency within the corpus. The low frequency group contained all other words that children produced with this cluster. The least frequent words containing stop combinations included words like: bathtub, blacktop, blanket,
bog, bug, god, peridot and jacket. In the liquid combinations the most popular words are: after, little, water, people, sister, winter, computer, November, better, bigger, December, monster, apple, super, chapter, paper and castle. These accounted for 49.9% of the stop/vowel/liquid combinations. The most frequent of these, ‘little’, ranks 96th in overall frequency within the corpus. The low frequency group consists of all the remaining words that children produced with this combination. The least frequent words of this type included: noble, chamber, cucumber, dagger, danger, penetrable, horrible, saddle. For stop/liquid combinations there was no difference in performance between high and low frequency words for either kindergarteners $t(55) = .27 \ p = .79$ or 1st graders $t(85) = .05 \ p = .96$. There was a difference in the proportion of 1st graders errors in stop/stop combinations $t(97) = 3.29 \ p = .001$, with 1st graders making more errors on high frequency words than low frequency words, but there was no difference in this combination for kindergarteners $t(73) = 1.04 \ p = .30$. Figure 5 shows these results.
In order to examine children’s growth directly, I narrowed the subjects to only the longitudinal group who had spellings of these combinations in both kindergarten and first grade. The effect outlined above was replicated with children making fewer errors on stop/vowel/liquid combinations in 1st grade rather than Kindergarten, \( t(21) = 3.32 p = .003 \). Interestingly, when the frequency effects were analyzed, children’s spelling of stop/vowel/liquid clusters improved for low frequency words, \( t(21) = 2.09 p = .049 \), but there was no significant improvement in children’s performance for high frequency words \( t(21) = 1.46 p = .158 \).
To further explore these results and relate them directly to children’s educational attainment, I correlated children’s in-school literacy tests with their errors in these patterns Table 5 shows these results, which indicate that the proportion of mistakes children make in stop/vowel/liquid combinations positively predict letter knowledge and concepts about print at the end of Kindergarten, such that making more errors in stop/vowel/liquid combinations suggests greater letter knowledge and greater knowledge of how print works. However, letter knowledge at the end of Kindergarten negatively predicts proportion of errors in 1st grade, such that the more letters children know the fewer mistakes they make.
Table 5: Correlation of proportion of spelling mistakes (sound to letter correspondence) with writing scores

<table>
<thead>
<tr>
<th></th>
<th>Proportion wrong stop/liquid</th>
<th>Proportion wrong stop/stop</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion wrong stop/liquid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion wrong stop/stop</td>
<td>.135</td>
<td></td>
</tr>
<tr>
<td>Letter Identification (end of K)</td>
<td>.494**</td>
<td>.176</td>
</tr>
<tr>
<td>Concepts about Print (end of K)</td>
<td>.735***</td>
<td>.095</td>
</tr>
<tr>
<td>Kindergarten Writing Score</td>
<td>.045</td>
<td>.092</td>
</tr>
<tr>
<td><strong>G1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion wrong stop/stop</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Letter Identification (end of K)</td>
<td>-.263*</td>
<td>-.018</td>
</tr>
<tr>
<td>Kindergarten Writing Score</td>
<td>-.171</td>
<td>-.006</td>
</tr>
<tr>
<td>Concepts about print (end of K)</td>
<td>.032</td>
<td>.111</td>
</tr>
<tr>
<td>1st Grade Writing Score</td>
<td>.074</td>
<td>.006</td>
</tr>
</tbody>
</table>

*p < .1, ** p < .05, ***p < .01

3.2.2 Morphological suffixes

Figure 7 shows the proportion of errors children make in each grade in situations where the orthographical dedication to marking morphology compromises the sound-to-letter correspondence and situations where it does not. This reinforces several
elements of the balance between orthography and phonology. First, this shows that, overall, children produce morphemic additions correctly, that is, they adhere to the orthography significantly more than half the time, whether there is a match \( t(172) = 24.98, p < .001 \) or a mismatch, \( t(192) = 19.68, p < .001 \) between the phonology and orthography. It also shows that kindergarteners make more errors when there is a phonological mismatch, \( t(63) = 2.117, p = .038 \); whereas, first graders, who have more experience with reading and writing, only make marginally more errors when there is a mismatch \( t(98) = 1.84, p = .069 \).

![Proportion of errors in the morphological suffix by phonological match and grade](image)

**Figure 7: Proportion of errors in the morphological suffix by phonological match and grade**

Breaking down these results by the frequency of the words we see a similar pattern (figure 8). There is no different between the proportion of errors in the matched and mismatched conditions for high frequency words in either Kindergarten, \( t(44) = \).
.203, *p* = .840 or first grade *t*(77) = .202, *p* = .841. For low frequency words, however, kindergarteners, *t*(38) = 2.06, *p* = .046 and first graders *t*(78) = 2.101, *p* = .039 each make more errors in the mismatched condition. This suggests that the difference between the match and mismatched conditions is driven by children’s performance on low-frequency words.

![Figure 8: Proportion of errors by grade, phonological match, and frequency](image)

**Figure 8: Proportion of errors by grade, phonological match, and frequency**

To further explore these errors I broke down children’s overall errors again, this time by the type of errors children make in each grade and matching group. I coded the errors to fall into 3 types. The first type of error, called ‘omission’ occurs when children leave the morphological suffix off the root word, for example, they spell ‘dogs’ as ‘dog’. In the past tense case, an error was classified as an omission when the ‘d’ was omitted, whether or not the ‘e’ was present. The second type of error, called ‘phono’ occurred
when children replaced the final letter with a ‘t’ (for past tense) or a ‘z’ (for plurals). For example, spelling ‘jumped’ as ‘jumpt’. In the match condition, this reflects the amount I expect children to produce this kind of error randomly (for example, spelling played as playt). The final type of error was called ‘other’ it consists of the other associated errors children make, including replacing the final letter with a random letter, and transposing the final 2 letters, for example, spelling ‘dogs’ as ‘dosg’. Here we see that children produce more omissions than any other type of error. That is, they spell ‘dogs’ as ‘dog’.

There is no difference in the proportion of errors that are omissions between words with a phonological match as those with a phonological mismatch for both Kindergarteners, \( t(13) = .546, p = .594 \), or first graders, \( t(21) = .183, p = .857 \). However, first graders replace the letters far more often in suffixes that have a phonological mismatch than when there is a match, \( t(21) = 2.396, p = .026 \). There is no difference in the proportion of replacement between conditions in Kindergarten, \( t(13) = .685, p = .505 \). We see the difference made up in the ‘other’ category of errors, with first graders making more ‘other’ errors in the match than mismatched condition, \( t(21) = 2.943, p = .008 \), and no difference in kindergarteners performance, \( t(13) = .303, p = .766 \). Essentially, though children produce roughly the same amount of mistakes in spelling a morphological suffix whether there is a match in the phonology or not, the type of mistake they make differ. When there is a phonological mismatch, children are much more likely to mistake the required letter for
a more phonologically appropriate one, whereas when there is no mismatch children make errors by dropping the morpheme all together.

![Figure 9: Types of errors made by phonological match and age](image)

Table 6 shows the correlation of the proportion of mistakes made in morphological suffixes when there is a match and mismatch between phonology and orthography measures of literacy development collected by the school. The correlation between the proportion wrong in the match and mismatch is marginally significant for first graders, \( p = .082 \) but non-significant for kindergarteners \( p = .115 \). Letter identification at the end of kindergarten is significantly related to the proportion wrong when there is a match between phonology and orthography for kindergarteners, \( p = .002 \), and marginally, first graders, \( p = .06 \), such that greater knowledge of letters leads to better spelling when there is a match between orthography and phonology. Children’s
spelling in the match condition also predicts writing scores at the end of kindergarten, \( p = .049 \), whereas first graders spelling in the mismatch condition predicts their writing scores at the end of that year, \( p = .043 \). No other correlations were significant. These results reinforce the suggestion of the previous analyses that children learn to spell in the match condition sooner than they learn to spell in the mismatch condition.

Table 6: Correlation between proportion misspelled morphological suffixes and literacy scores.

<table>
<thead>
<tr>
<th></th>
<th>Proportion wrong match</th>
<th>Proportion wrong mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion wrong mismatch</td>
<td>.199</td>
<td></td>
</tr>
<tr>
<td>Letter Identification (end of K)</td>
<td>-.581***</td>
<td>-.102</td>
</tr>
<tr>
<td>Concepts about Print (end of K)</td>
<td>-.272</td>
<td>.010</td>
</tr>
<tr>
<td>Kindergarten Writing Score</td>
<td>-.280**</td>
<td>-.052</td>
</tr>
<tr>
<td><strong>G1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion wrong mismatch</td>
<td>.181*</td>
<td></td>
</tr>
<tr>
<td>Letter Identification (end of K)</td>
<td>-.274*</td>
<td>.146</td>
</tr>
<tr>
<td>Concepts about Print (end of K)</td>
<td>.123</td>
<td>.063</td>
</tr>
<tr>
<td>Kindergarten Writing Score</td>
<td>-.053</td>
<td>-.079</td>
</tr>
<tr>
<td>1st Grade Writing Score</td>
<td>-.052</td>
<td>-.234**</td>
</tr>
</tbody>
</table>

\( *p < .1, **p < .05, ***p < .01 \)
Breaking these results down further by frequency we see that the relationship between letter identification and spelling in the match condition is driven primarily by high frequency words for both kindergarteners \( p = .012 \) and first graders, \( p = .001 \).

**Table 7: Correlations between literacy tests and morphological suffix spelling by frequency**

<table>
<thead>
<tr>
<th></th>
<th>Proportion wrong match (High Frequency)</th>
<th>Proportion wrong match (Low Frequency)</th>
<th>Proportion wrong mismatch (High Frequency)</th>
<th>Proportion wrong mismatch (Low Frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Letter Identification (end of K)</td>
<td>-.517**</td>
<td>.137</td>
<td>-.197</td>
</tr>
<tr>
<td></td>
<td>Concepts about Print (end of K)</td>
<td>-.288</td>
<td>.308</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>Kindergarten Writing</td>
<td>-.245</td>
<td>-.111</td>
<td>-.122</td>
</tr>
<tr>
<td>G1</td>
<td>Letter Identification (end of K)</td>
<td>-.482***</td>
<td>-.217</td>
<td>.079</td>
</tr>
<tr>
<td></td>
<td>Concepts about Print (end of K)</td>
<td>-.259</td>
<td>-.052</td>
<td>-.137</td>
</tr>
<tr>
<td></td>
<td>Kindergarten Writing</td>
<td>-.035</td>
<td>.028</td>
<td>.135</td>
</tr>
<tr>
<td></td>
<td>1st Grade Writing Score</td>
<td>-.062</td>
<td>-.021</td>
<td>-.148</td>
</tr>
</tbody>
</table>
3.3 Discussion

This study presents two situations in which orthographic rules create a mismatch between orthography and phonology and examines children’s naturalistic spelling in each situation. First, I show that children are more likely to make errors of omission when the omission would improve the sound-to-letter phonological correspondence. This is particularly true in Kindergarten; the major difference we see between kindergarten and first grade spelling of these combinations is that 1st graders produce fewer omissions in the stop/vowel/liquid combinations, particularly for low frequency words. Next, I show that, when a morphological suffix causes a mismatch between phonology and orthography, children show more difficulty following the orthography than when there is no mismatch, but that this improves from kindergarten to first grade. Taken together, these analyses show a trend of transitioning from errors created by a focus on phonology to a better understanding of orthography and focusing more on that. As children gain experience with written language they increase their ability to focus on the orthography and are better able to make choices based on the orthography rather than exclusively on the phonology. We see an integration of phonological and orthographic information so that each may be used when it is required.
Interestingly, the initial hypotheses regarding the role of high and low frequency words were not substantiated. In fact, my data show that children spell these particular combinations better in low frequency words. The contrast between this finding and the findings in Caravolas et al. (2005) is likely due to the creative nature of these spellings, and the specific combinations I have chosen. First, most measures of frequency are based on the frequency with which people encounter the words in literature; however, the high frequency words used here are the words that children produce the most often. This measure is a more accurate portrayal of the words children have the most experience writing. Similarly, because I am focused on specific letter combinations, I limited the sample to words that have these combinations. Because this is a fairly narrow and limited sample of possible spellings, the high frequency words are unlikely to appear in any measure of overall high frequency. As stated earlier, for the stop/vowel/liquid combinations, the highest frequency word ranks 96th on overall measures of frequency in the entire corpus. This is a constraint that would equally apply to dictation-based experimental methods. Furthermore, because these are creative spellings, children are focused more on communication than the spelling process. This may mean that, for words they use once or twice, they are very careful to make sure they can be understood, but, for words they use frequently, they are focused more on the communication and use a phonologically based strategy to spell. This finding, in
particular, would benefit from experimental manipulation where children’s communicative intent and attention to form can be examined thoroughly.

Correlational data with children’s school administered literacy tasks shows that letter knowledge in kindergarten predicts spelling ability in stop/liquid combinations and stop/stop combinations in first grade. Interestingly, for kindergarteners a good knowledge of letters predicts a higher proportion of omissions in a stop/vowel/liquid combination. This may be because kindergarteners who know their alphabet very well may insist on a phonologically consistent spelling more than a child who doesn’t know the alphabet. Similarly, school literacy tasks predict children’s production of phonologically matched morphological suffixes in kindergarten and first grade. These relationships reinforce the importance of knowledge of the elements of words in knowing how they go together and how to put them together. On the other hand, a good knowledge of the alphabet does not seem to help with phonologically mismatched suffixes. This pattern of correlation suggests that knowledge of letters helps orthographic spelling in general, but when there is an active mismatch between the sound and letter, knowledge of letters does not help.

The naturalistic nature of this data set and these analyses mean that I can only draw correlational conclusions about children’s orthographic knowledge. I can indicate that, in situations where there is a phonological and orthographic conflict, kindergarteners tend to focus on the phonological more than the orthographic, but I am
unable to control for variables like the words children choose to spell and the amount they spell. As a trade-off for these limitations, however, I can measure children’s performance in the real-world situations where they use it. Despite this advantage, a natural follow-up to this work would be to isolate children’s spelling patterns and habits and manipulate what children spell and when, to determine how they learn to produce words that have such a conflict.

Overall this study shows that in naturalistic circumstances, when there is a conflict between phonological and orthographic rules, young children emphasize the phonological and grow to understand and give importance to the orthographic. This suggests that early in reading development children’s orthographic knowledge, particularly their generalized orthographic knowledge can be masked by phonological concerns. Once children have integrated the phonological and orthographic constituents they are better able to use orthographic knowledge. This replicates findings from previous examinations of creative spelling (Read, 1986), but expands on their findings by focusing on a detailed analysis of individual particular conflicts between phonology and orthography.
4. Orthographic knowledge in decoding

In addition to spelling, a main activity for elementary school aged children in literacy is decoding written words on the path to spelling. The purpose of this study is to determine whether children implicitly show conceptual orthographic knowledge in the process of sounding out written non-words. Share and colleagues have shown that the process of decoding can help learn lexicalized orthographic representations (Shahar-yames & Share, 2008; Share, 2004). Presumably, decoding increases the focus on letters and combinations, which helps children later remember the combination in question. Here I am asking whether children show generalized orthographic knowledge in the process of decoding. Specifically, I am examining the online processes that occur as children encounter novel, illegal, combinations. If children have generalized orthographic knowledge, and this knowledge appears in the process of decoding, they should be sensitive to illegal orthographic combinations.

Children have been shown to utilize different decoding strategies in the process of word recognition. Children who learn a highly regular language like Finnish or German, rely heavily on a letter-to-sound correspondence strategy, while English speaking children tend to rely on strategies like rime and whole word recognition (Goswami, Ziegler, Dalton, & Schneider, 2001). Skilled English readers tend to use rime and other large word segments when decoding non-words (Forster & Taft, 1994; Goswami, Gombert, & de Barrera, 1998). However, when very young children are
learning to read, they are encouraged to sound out one letter at a time, which encourages a process of direct letter-to-sound correspondence (Landerl & Wimmer, 2000). Indeed, Duncan, Seymour and Hill (2000) have shown that children progress from a process of single letter-to-sound sounding out, to taking larger units of rime into account. Studies of adults have shown that experienced readers tend to ‘sound out’ non-words using even larger units of information, using analogy or generalized orthographic information to pronounce and unfamiliar written word. For example, Glushko (1979) showed that experienced readers had an easier time decoding a non-word like ‘taze’ than a non-word like ‘tave’. Although each of these non-words is orthographically regular at the letter-to-sound level, ‘taze is analogous to several Standard English words with consistent spelling-sound correspondences (haze, gaze, etc.); on the other hand, ‘tave’ is analogous to Standard English words with inconsistent spelling-sound correspondence (gave, have). This finding indicates that experienced readers don’t sound words out on a single letter level; rather, they examine both the letters and the combinations in the word to determine the best pronunciation. Taken together these studies suggest that we may expect a progression in decoding strategy in the current study from a strategy by which children sound out each letter (for example, sounding out ‘late’ as /l/ /æi/ /t/ /i/), to a strategy by which children sound out the onset and rime (/l/ /eit/, until finally children use a strategy by which children sound out the word as a blended whole word /leit/. Ehri, reports several ways to identify the methods
individuals are using in the process of decoding, including differences in timing, and differences in pronunciation of phonologically inconsistent patterns (pronouncing ‘pednesday’ to rhyme with ‘Wednesday’ for example) (Ehri, 1991). In this way, I expect to see a difference in performance from children who are able to blend the entire word together as one and children who sound out one letter at a time. Decoding strategy may, in fact, be a better proxy for development than either age or standardized measures of reading ability. Because the current task is one of decoding, examining children’s real time behaviour and approach to the decoding task is an immediate and strong measure of their skill in this area, rather than a mere proxy for an abstract reading skill or maturational development. Thus, the current study measures children’s natural decoding strategy as a quasi-experimental categorical variable by which to compare implicit orthographic knowledge across decoding types.

Study 1 has shown that children show implicit orthographic knowledge before they learn to read, and study 2 shows that, when children are beginning to learn to write, they focus on the phonological side of written language before they balance the process with an equal focus on the orthographic side. I predict that a similar shift will happen in decoding, i.e., initially children focus mostly on the letter-sound correspondence to the exclusion of their pre-existing orthographic knowledge. This will be replaced with a more balanced strategy that takes into account of both sources of knowledge. I argue that children have implicit understanding of orthographic
constraints before it is reflected in their decoding performance, and a key to reading development is to incorporate both the immediate task requirement (sounding out) and a broader array of existing knowledge, such as orthography.

Using the spelling corpus as a guide, I’ve identified specific consonant clusters that reflect children’s understanding of the rules regarding the interplay of phonology and orthography. In order to directly test children’s explicit strategies and implicit perception of these clusters, I asked children to decode non-words that contain these letter and sound patterns. The 3 combinations selected are 1) word initial doublets; 2) word initial stop/stop clusters; and 3) Word final stop/liquid clusters. I expect children’s implicit orthographic knowledge to appear differently in different clusters. Previous studies have shown that children have a knowledge that word-initial doublets are illegal from very early in the reading development process (Cassar & Treiman, 1997; Wright & Ehri, 2007), thus I expect these to be the most recognizable to early readers, even in a phonology driven task, like non-word reading. Because stop/stop clusters are both phonologically and orthographically illegal I expect children to respond to these fairly early in development. Because stop/liquid clusters are phonologically legal and orthographically legal in some positions, I expect children to have the greatest difficulty recognizing their illegality in a decoding task.

The current study has 3 hypotheses. First, that children’s decoding strategy will progress from a strategy based on direct sound-to-letter correspondence to one that
incorporates orthographic information and blends sounds together. Second, I expect that, similar to chapter 2, children will show implicit orthographic knowledge of combinations, and this knowledge will be expressed by longer looking time at the illegal combinations in non-words. Finally, I expect that children’s implicit orthographic knowledge will reflect differently for different letter combinations. I expect children to show the most knowledge on easier and more noticeable combinations, like illegal doubles at the beginning of words, and unpronounceable combinations, like stop clusters at the beginning of words, and show less knowledge about more difficult combinations, like stop/liquid clusters at the end of a word.

4.1 Methods

4.1.1 Participants

Sixteen 5-year-olds (mean age 5 years, 4 months; 40% female) and Twenty-one 6-year-olds (mean age 6 years, 5 months; 33% female) tested in schools in the Durham area. The ethnic diversity was representative of the school population, though less diverse than the overall population, with 5% of the sample being Hispanic, and 2% of the sample being African American.

4.1.2 Materials

Children were presented with a series of non-words and on a computer monitor attached to a Tobii eye tracking system (see Study 1).
The combinations are (see appendix 2 for all stimuli):

1) Word initial stop/stop combinations (e.g. DTEST)– These clusters are both phonologically and orthographically illegal. They can only be pronounced by turning the combination into a syllable (by adding a vowel sound in between the stops, or rearranging the letters), and do not occur within syllable in typical spelling.

2) Word final stop/liquid combinations (e.g. BAMTR) – These combinations are phonologically legal, that is they can be pronounced as written, but orthographically illegal. Although a combination like ‘tr’ is legal at the beginning of a word, it doesn’t appear in the word final position. These combinations are orthographically legal and frequent.

3) Word initial consonant doubles – consonant doubles like ‘ff’ are common at the end and middle of words and fully pronounceable, but illegal at the beginning of words. Consonant doubles have been examined extensively in previous literature and knowledge of how children approach decoding this particular combination can be plugged into a broader background of orthographic knowledge literature and broaden the insight into children’s perception of orthographic rules.

Tests of children’s reading and decoding ability will be conducted using the Woodcock-Johnson letter/word identification and Word Attack subsets.
4.1.3 Procedure

Children performed the non-word reading task in an eye-tracking session with several other tasks. As with study 1, children were introduced to a cartoon bear named Fiffi and instructed that Fiffi only likes to eat words. In order to choose the best words for Fiffi, they had to sound out the non-words on screen. The non-words were presented one at a time on the monitor with 82 point font. Children were allowed to take as much time as necessary to read aloud. Once the children had completed decoding the word the experimenter moved on to the next stimuli. After the reading trials, children completed the Woodcock-Johnson Letter Word Identification tests and Word Attack tests. If the children have problems with any of the tasks they will be encouraged and prompted accordingly. If children continue to have problems with the non-word reading or spelling tasks, the experimenter will give minor coaching to the participant in terms of “What’s just the first part?”

4.1.4 Behavioural coding

Children’s non-word reading performance was categorized into 1 of 5 decoding strategies. The strategies were based on evidence (Duncan, et al., 2000; Glushko, 1979) that finds that children’s decoding progresses from phonologically based (sound out the letters of a word) to partial orthographic (break words into onset-rime portions), and finally identify sound patterns by sight (read a word without sounding out at all). These categories were adjusted based on pilot data to include children who refused to decode.
at all, and accounting for children’s strategies in sight reading unpronounceable words.

The final list of coded strategies is:

1) Letter-by-letter – these children sounded out every letter in the non-word individually. For example, DTAM would be pronounced ‘d-t-a-m’

2) onset-rime – These children pronounced the first letter individually then blended the remaining letters. For example, FFELP would be pronounced ‘f-elp’

3) Blended (insertion subtype) – These children would blend the letters together while inserting vowels into unpronounceable combinations. For example, DTAM would be pronounced ‘detam’.

4) Blended (re-arrange subtype) – These children would blend the letters together while re arranging unpronounceable combinations. For example, DTAM would be pronounced ‘damt’.

5) Refusals – These children refused to attempt to sound out the word, even after repeated encouragement.

4.2 Results

4.2.1 Behavioural results

Children’s non-word reading performance was recorded and categorized into 1 of 5 decoding strategies: Children were coded into one of these groups for each word they decoded. Children were highly consistent in their decoding strategies, and only 1
child used different strategies from word to word, and he was given an overall code based on the strategy used most often. Once they were coded, the groups were collapsed into ‘Blenders’, made up of groups 3 & 4, and ‘Non-blenders’ made up of the remaining groups. A second coder examined a subset of videos and inter-coder reliability was 100%. This strategy division contains significant overlap with the age division, with most 6-year-olds in the blending group, and most 5-year-olds in the non-blending group. In fact, only five 5-year-olds were in the blending subgroup and three 6-year-olds were in the non-blending group.

The blending vs. non-blending grouping dynamic is also significantly correlated with children’s reading ability. The letter word identification subtest of the Woodcock Johnson is correlated with strategy choice at .432, \( p = .01 \). Similarly the word attack subtest is correlated with strategy at .443, \( p = .009 \). This suggests that strategy choice is an appropriate and meaningful division by which to compare student’s performance as well as their implicit orthographic knowledge. Thus, for the rest of these analyses, the students are compared along the lines of strategy rather than the lines of age.

### 4.2.1 Word Initial Doublets

In order to test children’s recognition of the illegal aspects of the word, I compared children’s looking time at the first 2 letters of a word with a word initial doublet with the first 2 letters of a pseudo-word. The mean was calculated across letters, so looking time represents the average looking time of any one of the defined letters. A
2 by 2 repeated measures ANOVA shows a significant main effect of condition $F(1, 31) = 6.497$, $p = .016$, but no significant effect of strategy $F(1, 31) = .014$, $p = .91$ or significant condition by strategy interaction $F(1, 31) = .025$, $p = .88$. Follow up tests show that non-blenders looked longer at the legal combination than the doublet $t(10) = 2.43$, $p = .036$, and blenders also look marginally longer at the legal combination $t(21) = 1.85$, $p = .079$. Overall these results suggest that children recognize that word-initial doublets are inappropriate, but once they have identified one of the doubled letters, they do not spend much time examining them, considerably less time than they spend examining letter that vary in the word initial position.

![Figure 10: Looking time at initial letters by condition and decoding strategy](image)

In order to fully examine the process by which children interpret, I further broke down children’s eye-movement data by time, to examine where children are looking at individual time points from the appearance of the word on the screen.
Figure 11: Percent of non-blenders looking at each letter by time
Figure 12: Percent of blenders looking at each letter by time

These analyses show a trend in blenders towards looking from left to right. They move faster to the end of the word when that word contains a doublet at the beginning of the word. Non-blenders on the other hand, do not show the same overall trend of
moving left to right, though they still spend less time on the beginning of the word when that word contains a doublet at the beginning.

In order to more closely examine children’s focus on the illegal segment of these stimuli, I divided each word into segments, with the first segment containing the first two letters and the second segment containing the remaining letters. From this I calculated the proportion of subjects focusing on the first two letters at any given time point and compared the proportion of subjects looking at the first two letters when that letter was a doublet with the proportion of subjects looking at the first two letters when it was two different letters.
As figure 13 shows, the strategy difference of blending or not shows a fundamental difference in how children approach the process of decoding a written word. When dealing with a pseudo-word, the blenders show a fairly clear pattern of movement from the first 500ms to 800ms. The non-blenders on the other hand tend to
look at the beginning, and transition much more slowly to looking at the end of the word. When dealing with doublets blenders follow the same basic pattern they do with pseudo-words, until late in the decoding process, when they tend to ignore doublets. Pre-planned follow up tests show that for blenders, the significant difference in proportion looking at the initial segment in the doublet vs. phonology conditions comes towards the end of the decoding process, specifically the 2200ms $t(75) = 2.86$, $p = .006$, 2800ms $t(78) = 2.16$, $p = .03$, 3600ms $t(61) = 2.5$, $p = .015$, and 4400ms $t(34) = 2.42$, $p = .021$ time points. For non-blenders, the difference in looking times comes sporadically throughout the decoding process, specifically, the 200ms, $t(13) = 3.6$, $p = .003$, 400ms, $t(21) = 2.4$, $p = .031$, 1800ms, $t(19) = 2.51$, $p = .021$, 2000ms, $t(19) = 2.21$, $p = .040$, 3200ms, $t(23) = 3.39$, $p = .003$, 3400ms $t(12) = 2.74$, $p = .018$, and 5000ms $t(15) = 3.00$, $p = .009$ time points. Overall these results suggest that blenders generally spend more time looking at the legal parts of the word rather than the illegal parts of the word. Non-blenders however, seem to revisit the doublet pattern more often than blenders.

**4.2.2 Word Initial Stop/Stop Clusters**

Because stop/stop clusters are both phonologically and orthographically illegal I expect children to respond to these fairly early in development. Again, I compared looking times between the first 2 letters in a pseudoword and the first 2 letters in a stop/stop cluster. A 2 by 2 repeated measures ANOVA shows no significant main
effects for either condition, $F(1, 31) = 1.00$, $p = .32$ or strategy, $F(1, 31) = 1.68$, $p = .21$, but there is a significant condition by strategy interaction $F(1, 31) = 4.23$, $p = .048$. For non-blenders there was no significant difference in looking time between the illegal combination and pseudo-words $t(11) = .654$, $p = .527$. Blenders, did show a difference, looking significantly longer at the illegal combination than the pseudo-word letters in the same position $t(20) = 2.55$, $p = .019$.

![Figure 14: amount of time looking at the first 2 letters in a word by strategy and condition](image)

To further examine the online processes of decoding when an illegal stop cluster appears in the letter string, I broke down the proportion of children looking at each individual letter at each time point.
Figure 15: Percent of non-blenders looking at each letter by time.
Figure 16: Percent of blenders looking at each letter by time

As with the doublets, these analyses show non-blenders do not show a trend of moving left to right, though they still spend more time on the beginning of the word when that word contains a stop/stop cluster at the beginning. In blenders, however we
do see a trend towards looking from left to right. They move slower to the end of the word when that word contains a stop/stop cluster at the beginning of the word.

Again, in order to fully examine the decoding process with regards to the illegal combination, I split the word into the same segments outlined above, and again compared the proportion of subjects looking at the first 2 letters when that segment was a stop/stop cluster with the proportion of subjects looking at the first 2 letters when it was 2 different legal letters.
Figure 17: Proportion of participants looking at first 2 letters by decoding strategy, condition, and time

Similar to the data from the doublet condition, we see a large difference in general processes between the blenders and non-blenders. While for pseudowords, the non-blenders start with a high proportion looking at the beginning of the word and
slowly the proportion decreases, implying that participants are looking more at the end of the word, in stop/stop clusters, the proportion looking at the illegal combination, spikes in the middle of the decoding process. A series of t-tests shows that for non-blenders the difference between pseudo-words and stop/stop clusters is only significant at the 1200ms $t(25) = 3.95, p = .001$, and 3000ms, $t(20) = 3.6, p = .002$, time points. Blenders, on the other hand, show a clear trend towards looking longer at the illegal portion of the stop/stop cluster throughout much of the decoding process. This difference is significant or marginally significant from the 600ms time point through the 2000ms time point, 600ms $t(70) = 1.82, p = .072$, 800ms $t(70) = 1.99, p = .051$, 1000ms $t(69) = 1.84, p = .070$, 1200ms $t(103) = 1.93, p = .056$, 1400ms $t(100) = 2.29, p = .024$, 1600ms $t(107) = 1.8, p = .038$, 1800ms $t(113) = 2.06, p = .042$, 2000ms $t(106) = 2.65, p = .009$. This suggests a pattern wherein non-blenders attend less to illegal combination, and blenders look more often at the illegal segment at any given time point.

### 4.2.4 Word Final Stop/Liquid Clusters

Because stop/liquid clusters are phonologically legal and orthographically legal in some positions, I expect children to have the greatest difficulty recognizing their illegality in a decoding task. This time, I compared looking times between the last 2 letters in a pseudo word and the last 2 letters in a stop/liquid cluster. A 2 by 2 repeated measures ANOVA showed no significant main effects for either condition $F(1, 31) = 1.35,$
$p = .25$ or age, $F(1, 31) = .006$, $p = .94$ or a condition by strategy interaction $F(1, 31) = .027$, $p = .871$.

Figure 18: Amount of time looking at last two letters by strategy and condition

To further examine the online processes of decoding when an illegal stop/liquid cluster appears at the end of the letter string, I broke down the proportion of children looking at each individual letter at each time point.
Figure 19: Percent of non-blenders looking at each letter by time
As with the previous combinations, these analyses show non-blenders do not show a trend of moving left to right, though they still spend slightly more time on the end of the word when that word contains a stop/liquid cluster at the end. In blenders, however we do see a trend towards looking from left to right. We also see that they
move faster to the end of the word when that word contains a stop/liquid cluster at the end of the word.

Again, in order to fully examine the focus on the illegal combination, I split the word into segments, because this analysis focuses on the end of a word, the segments consisted of the last two letters and the initial letters. I compared the proportion of subjects looking at the last two letters when that segment was a stop/liquid cluster with the proportion of subjects looking at the last two letters when it was two different legal letters.
Similar to the word initial conditions, the patterns between blenders and non-blenders differ slightly for pseudo-words, with the blenders showing a clear left to right trend, and the non-blenders moving more slowly towards the right side of the stimuli.
Unlike the word initial conditions, however, children spend very little time examining the illegal portion, likely because it is at the end of the word. In a series comparison tests, for non blenders we see the same sporadic pattern of significant differences in looking time between the pseudo-word and stop/liquid combinations, with significant differences at the 2200ms $t(23) = 2.24$, $p = .035$, 3200ms $t(23) = 5.21$, $p < .001$, and 3600ms $t(10) = 2.42$, $p = .036$ time points. Whereas for the blenders, the only significant difference occurs towards the beginning at the 400ms time point $t(65) = 2.31$, $p = .024$. This suggests a similar pattern to those above, with non-blenders looking at the illegal segments of the illegal word less often than blenders at most points in the timeline. This distinction is more subtle for the word final condition however, possibly due to the position at the end of the word.

4.3 Discussion

This study shows children’s implicit generalized orthographic knowledge when they engage in a traditional early literacy task, decoding novel words. Each of our hypotheses was supported. First, children vary in their strategies for decoding and children who use a sound-to-letter correspondence based strategy show less orthographic knowledge than children who use a blended strategy. This reinforces findings that an increase in knowledge about combinations and patterns in letters is related to how one decodes written words (Duncan, et al., 2000; Glushko, 1979). This
also suggests that strategy choice may be a more informative and meaningful way to examine children’s behaviour than age or abstract reading knowledge.

We also see that some children show implicit orthographic knowledge in the form of longer looking times at the illegal segment of the non-word. Furthermore, for the easiest combination, identifying word-initial doublets as illegal, I find that children who use either the blending or non-blending strategies recognize the illegal orthographic combination, and this recognition reveals itself early in the orthographic process. A slightly harder combination to identify, stop/stop clusters, shows that children using the blending strategy can recognize the illegal combination, and this recognition also reveals itself towards the end of the decoding process. Finally, the most difficult task, identifying word final stop/liquid clusters as illegal, shows that children with the blending strategy type fail to recognize the illegality of the combination except briefly at the beginning of the decoding process.

For non-blenders, we see a slightly different pattern in which their implicit knowledge reveals itself sporadically throughout the decoding process. They show the most knowledge in the easiest task to recognize, the word-initial doublets, it also frequently happens in the more difficult stop/stop cluster task, and finally happens the fewest in the most difficult task, word-final stop/liquid combinations.

Taken together these results suggest that children with different strategy types have fundamentally different reactions when they encounter an orthographically illegal
letter combination. Both groups move away and are drawn back to the illegal combination, but where blenders are roughly equally likely to look at the illegal combination at any point in the timeline, non-blenders tend to look back less often and only at certain points. This suggests that non-blenders recognize the illegal combination less consistently than blenders, but that they do show some implicit recognition.

This work provides an interesting combination of the self-teaching hypothesis and children’s decoding strategies. Children show implicit knowledge of orthography during the process of decoding, but only if they have reached a point in their reading development that they can decode non-words based on more than just simple sound-to-letter correspondence. Furthermore, self-teaching suggests that decoding is the process by which children learn about orthographic rules. I have shown here that in the process of decoding children who have some sense of orthography can implicitly show previously learned knowledge. In this way decoding can be seen as both a learning opportunity and test of previously learned knowledge. Every encounter with a new word provides information about orthographic combinations, but we also see previously learned knowledge affecting how children approach and process a novel word.

We also see a trend of looking times that reflects a progression of the stimuli from easy to difficult. Unsurprisingly, given the results from chapter 3, I find that children recognize unpronounceable combinations before they recognize subtle pronounceable combinations. At first glance, this finding may not seem to reflect
orthographic knowledge as much as it reflects phonological blending knowledge. That is, once children are able to blend sounds together, they should clearly get hung up on a combination that cannot be pronounced or blended without the addition of a vowel. However, we also see that children show awareness of doublets, which are illegal but pronounceable. This suggests that children are not attending to the stop/stop patterns solely because they are unpronounceable, but at least partly because they recognize that the combination is unusual or illegal.

Limitations to this study include the low sample size, particularly in the non-blending strategy type. The robust p-values obtained suggest that this is less of an issue than would initially seem, and, because children cannot be randomly assigned to decoding strategy, the imbalance between the two groups is expected. I also did not directly ask children about their decoding strategy, I categorized them based on their naturalistic performance in the task. Although this method has some problems, it has the advantage of capturing children’s naturalistic tendencies in a decoding task. Furthermore, reliability data indicate that this is a highly consistent and easily identified behaviour in children. By not drawing attention to the strategy they used, I can allow them to choose the style they feel most conductive to the task. Finally, I tested a fairly small subset of illegal orthographic combinations. Ideally I would have been able to test children’s awareness of several orthographically illegal combinations in different word positions and using different letter frequencies. I limited our study to these
combinations because of the findings related to these combinations in the spelling corpus, and I did not want to overtax our participants.

Overall this study expands on the findings in chapter 2, to show that children do have a perception of illegal letter combinations, although they only show this knowledge if they have a ‘blending’ decoding strategy. I also expand on chapter 3 by showing that children demonstrate implicit generalized orthographic knowledge in a typical literacy task, where children are not asked to make any judgments about word quality. Finally, I demonstrate the online processing that occurs as children encounter and decode novel words with illegal or unusual letter combinations.
5. Orthographic knowledge in decoding and judgment

If generalized orthographic knowledge develops before lexicalized orthographic knowledge, but is difficult to detect due to phonological task requirements in early literacy, then we would expect children to show orthographic knowledge in a task that does not require phonological knowledge, and then fail to show generalized orthographic knowledge in a similar task that does require phonological knowledge.

The process of balancing orthographic and phonological knowledge in learning to read requires children to attend to both aspects of a word as they learn it. The self-teaching hypothesis, as described above (Jorm & Share, 1983), indicates that decoding or spelling a word requires children to attend to both the orthography and phonology and thus, is the process by which children learn and develop orthographic representations. In fact, a single encounter with a novel word or non-word is often enough for children to encode orthographic detail (Share, 2004). This theory is supported by evidence from chapter 4 of this dissertation, where I showed that children attend to violations orthography (particularly if it is salient or unpronounceable) during the process of decoding.

Chapter 3 of this dissertation, however, presents contrary evidence. Although children's free spellings show that they frequently do attend to orthography, they often focus on the phonology rather than the orthography. Similarly, where there is a conflict between orthography and phonology children tend to emphasize the phonological aspect over the orthographic one. This emphasis on phonology in spelling has been
shown in previous studies (Treiman, 1993), which suggest that an emphasis on phonology could mask or impair children’s pre-existing orthographic knowledge.

These two lines of research have created somewhat of a conflict in the understanding of the processes of the development of orthographic knowledge. On the one hand, there is evidence that as children attend to phonology in decoding, they learn orthography. On the other hand, when children attend to phonology they make errors in the production of orthography. The purpose of the present study is to determine compare children’s orthographic knowledge in situations with and without a focus on sound-to-letter correspondence.

With study 1, I showed that even pre-reading children have implicit – and sometimes explicit – orthographic knowledge. One limitation of that study was the lack of direct comparison between legal and illegal combinations. In order to replicate and expand on this finding, I tested children in a choice paradigm where they were asked to choose which of the two non-words presented was a “better word”. In order to further examine the interaction between phonological and orthographic knowledge, children performed this task with and without a phonological component. I either ask them to simply judge the quality or I ask them which of the 2 non-words is better for a certain pronunciation of a non-word. Because I am testing older children, I focused on illegal letter combinations (like stop/stop clusters), rather than non-letter symbols (like numbers). I used the same categories of illegal combinations as in Chapter 4.
This study attempts to directly test children’s perceptions and implicit understanding of orthographic rules as they run in contrast to phonological rules. In order to test this, I have children choose between two words on a screen, in which I either asked them to simply judge the quality or I asked them which of the two non-words looks better for a certain pronunciation of a non-word. I track their gaze as they perform this task to determine where they focus when they make their decision. There are two competing hypotheses for this study. If Share’s self teaching hypothesis is applicable, children should be more aware of orthographic violations when attention is called to the phonological make-up of a letter string. That is, they should look longer at the illegal combination (or select the legal combination) more often when they are asked which word is best for a specific pronunciation. On the other hand, if children approach decoding the same way they approach spelling, the focus on phonology may make children less likely to attend to the combination of letters and more attentive to the sounds of individual letters, therefore leading to a decreased sensitivity to illegal combinations. That is, they should look longer at the illegal combination (or select the legal combination) more often when they are just asked which non-word is better. In addition to the general effect of focus on phonology on orthographic knowledge I expect different levels of response for different combinations.
5.1 Method

5.1.1 Participants

Sixteen 5-year-olds (mean age 5 years, 4 months; 40% female) and twenty-one 6-year-olds (mean age 6 years, 5 months; 33% female) tested in schools in the Durham area. The ethnic diversity was representative of the school population, though less diverse than the overall population, with 5% of the sample being Hispanic, and 2% of the sample being African American. This was the same sample used in Chapter 4.

5.1.2 Materials

Children performed 2 lexical decision tasks in this study each involving their perception and understanding of 3 different orthographically illegal letter combinations. The first lexical decision task is similar to study 1. In this condition children are presented with 2 words at the same time. One of which is orthographically legal and one of which is orthographically illegal. Children will then be asked which of the two words is a “better” word. The 2nd lexical decision task is similar to the first, but it involves the phonology of the non-words. In this condition children are presented with 2 non-words at the same time. One of which is orthographically legal and one of which is orthographically illegal, but both of which would be pronounced the same way. Children will then be asked which of the two words is a “better” word for the spoken pronunciation of the word. Appendix 3 shows the exact words used in this study.

Combinations:
1) Word initial stop/stop clusters – These clusters are both phonologically and orthographically illegal. They can only be pronounced by turning the combination into a syllable (by adding a vowel sound in between the stops). I showed in study 2 that 5-year-olds produce this combination at the end of word about 16% of the time, and I am interested in how children perceive this combination.

2) Word final stop/liquid clusters – as in study 2 these combinations are phonologically legal but orthographically illegal. I showed in study 2 that children produce this combination about 30% of the time, suggesting that it’s a problem they frequently have.

3) Word initial consonant doubles – consonant doubles like ‘ff’ are common at the end of words and fully pronounceable, but illegal at the beginning of words. Consonant doubles have been examined extensively in previous literature, and knowledge of how children approach decoding this particular combination gives a great deal of insight into their perception of orthographic rules. Children rarely produce this combination in their spelling, which suggests that the orthographic understanding is well established.

5.1.3 Procedure

Children performed these lexical decision tasks in an eye-tracking session with the non-word reading task of study 3, the apparatus was the same as for studies 1 and 3,
but because 2 words were on the page, they were presented in 72 point font. Because I did not want to influence children’s strategies for evaluating non-word quality in the no-phonology condition, children completed this task first. The phonology condition and non-word reading task were counterbalanced. As with Study 1 and 3 children were told about a bear named Fiffi who only liked to eat words, they were instructed to choose the ‘best’ word for Fiffi, of the two non-words presented on the screen. Children were instructed to place their fingers on the z or / keys for the entire study, and to press down on the button to indicate the word on which side of the screen they thought was ‘better’. If at any time they have difficulty performing the task they will be encouraged and prompted accordingly.

5.2 Results

5.2.1 Explicit orthographic knowledge

Figure 16 shows that for the word initial stop and word final stop liquid combinations, children performed at chance when selecting which stimuli on the screen was a better word. This was true for both the no-phonology \( t(37) = .47 \ p = .64, \ t(37) = 1.62 \ p = .112 \) and phonology conditions \( t(36) = .62 \ p = .54, \ t(36) = .33 \ p = .74 \). There was also no significant age effect in these stimuli for either the no phonology \( t(36) = .62 \ p = .54, \ t(36) = .33 \ p = .74 \) or phonology conditions \( t(36) = 1.13 \ p = .27, \ t(36) = .91 \ p = .367 \).
Using children’s decoding strategies from chapter 4 as a divider, rather than their ages, I get similar results with blenders and non-blenders performing similarly in the no-phonology conditions for stop/stop combinations $t(36) = 1.58 \ p = .12$, but blenders perform marginally significantly better in stop/liquid combinations $(36) = 1.9 \ p = .054$. In the phonology condition blenders and non-blenders performed similarly in both of these conditions $t(35) = .41 \ p = .68$, $t(35) = .71 \ p = .49$. This suggests that both blenders and non-blenders show no explicit knowledge that word initial stop and word final stop liquid clusters are illegal letter combinations.
On the other hand children performed significantly above chance in the word-initial doubles condition for both the no-phonology $t(37) = 2.26 \ p = .03$, and phonology conditions $t(36) = 3.88 \ p < .001$. The decoding strategies from study 3 show a significant difference between blenders and non-blenders in the no-phonology condition $t(36) = 2.14 \ p = .039$, and a marginally significant difference in the phonology condition, $t(35) = 1.78 \ p = .08$ This suggests that children learn a letter combination is illegal when it is salient, like a doubled letter at the beginning of a word and thus, they can successfully select a better option.

![Figure 23: Performance on behavioural task by condition, phonology and decoding strategy](image.png)
Because of the similarity between the age based division and the strategy based division, as well as the more meaningful nature of the strategy division, for the remainder of these analyses, I will use decoding strategy as the contrast of interest.

5.2.2 Eye-tracking results

Based on children’s behavioural responses I’ve collapsed the three combinations into two main groups. Word initial stop clusters and word final stop/liquid clusters are categorized as subtle illegal letter combinations, whereas word initial doublets, because of children’s explicit recognition, are categorized as salient illegal letter combinations.

5.2.2.1 Subtle illegal letter combinations

A repeated measure ANOVA was conducted on children’s looking time at the letter strings on the screen. A marginally significant main effect of strategy was found $F(1, 34) = 4.02, p = .053$, as well as a significant main effect of task (phonology vs. no-phonology) $F(1, 34) = 4.45, p = .042$, and legality $F(1, 34) = 5.35, p = .027$. Furthermore, there was a significant strategy by task interaction, $F(1, 34) = 5.34, p = .027$, as well as a significant phonology by legality interaction $F(1, 34) = 4.85, p = .035$. Figure 18 shows these main effects of age and phonology, and the age by phonology interaction.
Figure 24: Looking time at non-word by legality, condition, and strategy

Figure 24 also shows the main driving force behind the phonology by legality interaction is the no-phonology condition. Follow up t-tests show that non-blenders look marginally longer at the illegal letter combination than the legal one in the no-phonology condition, $t(13) = 2.12 \ p = .054$ and blenders look significantly longer at the illegal letter combination in the no-phonology condition $t(22) = 2.17 \ p = .041$. Both non-blending $t(12) = .596 \ p = .563$ and blending $t(22) = .613 \ p = .546$ strategy types look at both the illegal and legal non-words for an approximately equal amount of time in the phonology condition. This suggests that when early readers select their own strategies for examining and evaluating words, they show an implicit knowledge of the orthographic rules that govern writing. When they must use decoding, or at least focus on decoding, we do not see an implicit orthographic awareness.
In order to fully examine the decoding process I examined children’s focus on each of words presented on screen over time. This gives us the ability to determine how children make judgments about word quality. I examined the proportion of subjects looking at the illegal non-word side of the screen at different time points as they make their decision. An ANOVA showed, in addition to the analyses above, a significant phonology by strategy by time interaction $F(50, 8869) = 1.4, p = .033$ Because blenders took longer to make the decision overall, the time course of their looking pattern is longer than the time course of the non-blenders looking pattern. Similarly, children took more time overall to make the decision in the no-phonology condition, so the time course of the looking pattern for the no-phonology condition is longer than that of the phonology condition.
In the no-phonology condition, non-blenders spend significantly more time on the illegal non-word at the 1200, \( t(62) = 2.07, p = .03 \), 1400, \( t(55) = 2.21, p = .03 \) and 1600, \( t(89) = 1.97, p = .05 \), time points. Blenders look significantly longer at the illegal non-word from the 2400ms time point through the 4200ms time point \( t(117) = 2.05, p = .04 \), \( t(114) = \).
In the phonology condition there is no point at which the proportion of non-blenders looking at the illegal non-word is significantly different from chance. For blenders the only significant difference from chance is at the very beginning, $t(42) = 2.41$, $p = .02$, $t(83) = 2.47$, $p = .02$, and around the three second time point $t(88) = 2.75$, $p = .01$, $t(89) = 2.15$, $p = .03$. Both of these time sections identify points at which students are less likely than chance to be looking at the illegal non-word.

Taken together, these results reinforce the previous findings that blenders, and to a certain extent, non-blenders in the no phonology condition are implicitly aware of the more subtle orthographic violations, whereas in the phonology condition, neither group shows a strong recognition of the illegal combination.

### 5.2.2.2 Salient illegal letter combinations

Again, a repeated measures ANOVA was conducted on children’s looking time at the letter strings on the screen. A significant main effect of task (phonology vs. no-phonology) was found $F(1, 30) = 5.23$, $p = .03$, but no other significant main effects or interaction effects were found.
In order to fully examine the decoding process I examined children’s focus on each of words presented on screen over time. This gives us the ability to determine how children make judgments about word quality. I examined the proportion of subjects looking at the illegal non-word side of the screen at different time points as they make their decision. Similar to the previous section, non-blenders make the decision significantly quicker than the blenders, so their overall looking pattern is significantly shorter. An Anova showed that in addition to the effects outlined above, there was a significant main effect of time, $F(50, 2421) = 1.42, p = .03$ and a significant time by strategy interaction, $F(48, 2421) = 1.69, p = .002$. 

Figure 26: Time spent looking at non-words by task, strategy and legality
Figure 27: proportion of subjects looking at illegal non-word across trial time

In the no-phonology condition, follow-up tests show that there is no point at which the proportion of either non-blenders or blenders looking at the illegal non-word is significantly different from chance. Similarly, in the phonology condition there is no point at which the proportion of non-blenders looking at the illegal non-word is significantly different from chance. For blenders the only significant difference from
chance is at the very beginning, $t(8) = 3.5, p = .01$, $t(12) = 3.32, p = .01$, where blenders are significantly more likely to look at the illegal non-word, and around the 1800ms time point $t(43) = 2.20, p = .03$, $t(89) = 2.15, p = .03$, when blenders are significantly less likely to look at the illegal non-word.

Taken together, these results suggest that when children have an explicit knowledge about the illegality of a letter combination, they don’t spend more time inspecting the source of errors.

5.2.2.3 Implicit knowledge and reading skill

Children’s reading and decoding skills were measured using the Woodcock Johnson test and correlated with their looking time at the pseudo-words and illegal non-words. Unsurprisingly, because of the high correlation between reading scores and decoding strategy (outlined in chapter 4), once strategy was partialed out there were few significant correlations between reading and looking time. For the doublet combinations, there were no significant correlations between the reading tests and looking time once decoding strategy was taken into account. For the subtle violations of orthography, in the phonology condition, the word attack subtest was significantly correlated with the amount of time looking at the illegal combinations ($p = .036$) and marginally significantly with the amount of time looking at the legal combinations ($p = .081$). There were no other significant correlations once decoding strategy had been
taking into account. This suggests that the variability in reading skill that is not already accounted for by decoding strategy significantly affects children’s ability to detect illegal combinations in the phonology condition.

Table 8: Looking time correlated with reading subtests, partialed for decoding strategy

<table>
<thead>
<tr>
<th></th>
<th>Letter Word</th>
<th>Word Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Salient</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No phonology</td>
<td>.305</td>
<td>.246</td>
</tr>
<tr>
<td>Pseudoword</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salient*</td>
<td>.246</td>
<td>.237</td>
</tr>
<tr>
<td>Illegal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonology</td>
<td>.110</td>
<td>.284</td>
</tr>
<tr>
<td>Pseudoword</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illegal</td>
<td>.211</td>
<td>.236</td>
</tr>
<tr>
<td><strong>Subtle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No phonology</td>
<td>.171</td>
<td>.211</td>
</tr>
<tr>
<td>Pseudoword</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtle violations</td>
<td>.122</td>
<td>.142</td>
</tr>
<tr>
<td>Illegal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonology</td>
<td>.312</td>
<td>.335*</td>
</tr>
<tr>
<td>Pseudoword</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illegal</td>
<td>.253</td>
<td>.398**</td>
</tr>
</tbody>
</table>

* p < .1, ** p < .05

5.3 Discussion

This study directly compares children’s orthographic knowledge in identical situations where children are either required to use phonology or not. I find that for
salient violations of orthography, children have explicit awareness and can successfully choose the legal option over the illegal one. I find this in both the phonology and no phonology condition. When it comes to subtle violations of orthography, however, I find that neither age group shows explicit knowledge, but each group shows some implicit knowledge in the no-phonology condition.

These results further the exploration of the role of phonology in the development of implicit and explicit orthographic knowledge. These findings support a theory of orthographic development in which generalized knowledge develops before lexicalized knowledge, but is masked by phonological concerns. When children are focusing on sound to letter correspondence they tend to show less implicit orthographic knowledge. This seems to run in contrast to theories of orthographic development, like the self-teaching hypothesis, that argue that the process of decoding leads to better attentiveness for letter patterns and increased orthographic knowledge.

These findings give further clarification to the interactive process between phonological decoding and orthographic knowledge. The fact that we do not see a boost in attention to orthographic detail when children are asked to select the best word for a certain sound, suggests that self-teaching does not apply here. There are a few potential explanations for this finding. First, the self-teaching work has primarily shown that decoding helps later orthographic recall, it’s possible that real-time online processes show no difference, and that decoding only helps in the long-term. This explanation is
contradicted by the results from chapter 4, however, in which I see that children do
attend to orthographic detail during the decoding process, and show an implicit
sensitivity to illegal combinations.

Second, self-teaching studies deal exclusively with explicit knowledge. In this
study I show that when phonology is introduced children’s implicit knowledge no
longer shows, although it has no effect on their explicit knowledge. This suggests that in
the early stages orthographic knowledge is fragile and susceptible to being masked.
Once this knowledge is more explicit, it becomes robust to interference from
phonological tasks. Once again, however, because we see in Chapter 4, that children do
attend to illegal combinations during the process of decoding, this explanation doesn’t
hold.

Third, in both the self-teaching studies and chapter 4 children were presented
with words one at a time. In this study they were asked to compare 2 non-words. This
comparison and choice procedure is used frequently to measure orthographic
knowledge in young children (Cassar & Treiman, 1997; Zivian & Samuels, 1986). This
comparing and contrasting non-words, may require a more advanced orthographic
conceptualization and the added pressure of decoding while comparing and contrasting
may be too much for beginning readers.

Finally, the self-teaching work has focused mostly on individual orthographic
representations rather than the generalized knowledge I am testing here. It is possible
that children learn specific representations through decoding, but when they are tested about general knowledge acquired indirectly, they don’t access previously existing orthographic knowledge when they are decoding. In Chapter 4, children weren’t asked to make any judgments about the words in question; they only had to decode the word. This makes it more of a learning task than a performance task. That is, children could be using the opportunity to build a conceptual orthographic understanding, rather than using a pre-existing orthographic understanding to make decisions.

A combination of these final two explanations seem to have the best chance for adequately explaining the nuance found in the interaction between phonology and orthography in the development of orthographic knowledge. Children show implicit orthographic knowledge for some illegal letter combinations when they simply decode them (Chapter 4) or when they make judgments about them with no phonological component. Children have difficulty showing orthographic knowledge when they make judgments on words they must first decode, and when they are spelling difficult or phonologically inconsistent patterns. The next section goes into more detail tying these results to a larger picture of orthography and phonology in reading development.

Limitations to this study include the relatively small sample size, although the robust p-values suggest that power may not be an issue. Also, because children were not given direct instructions as to strategy choice in the no-phonology condition there is no guarantee they did not use a decoding strategy to determine which word was best.
However, the fact that the looking time differed between the no-phonology and phonology tasks suggests that children were using different strategies in the two conditions.

Overall this study provides strong evidence that a focus on sound-to-spelling correspondence in a judgment task can mask children’s implicit orthographic knowledge. The reasons for this masking may have to do with children failing to access pre-existing orthographic knowledge when they are focusing on the dual task of judgment and decoding. Nonetheless, this supports a theory in which generalized orthographic knowledge develops early but is masked by phonology early in reading development.
6. General discussion

This dissertation presents a model of the development of orthographic knowledge that isolates and examines children’s knowledge of generalized orthographic patterns. Taken together, the current studies show evidence for a model of development of orthographic knowledge that begins before children are able to read, develops concurrently with children’s phonological knowledge, and is well established before children demonstrate lexical orthographic knowledge. Furthermore, these studies show evidence that children’s generalized orthographic knowledge can be over-shadowed by the demands of several literacy tasks, making it difficult to detect early in development. Finally, these studies demonstrate that generalized orthographic knowledge develops from implicit to explicit awareness.

6.1 Conceptual Model of Generalized Orthographic Knowledge

Replicating and expanding on previous findings, study 1 showed that children have generalized orthographic knowledge before they learn to read. In pre-reading children this knowledge takes the form of understanding the appropriate symbols within words. Study 3 showed that as children begin to learn to read, this knowledge develops to encompass patterns of letters in words. In both of these studies, children were tested on non-words, so lexicalized orthographic knowledge cannot explain their recognition and judgment of these combinations. The fact that none of the children tested in study 1 were able to read, strongly supports the hypothesis that generalized
knowledge develops before lexicalized knowledge. The significant relationship between letter knowledge and orthographic knowledge in this study also supports the overall model, which predicts at least some letter knowledge as a pre-requisite for generalized orthographic knowledge.

Future research is needed to directly test a model like this and elucidate the relationship between letter knowledge, generalized orthographic knowledge, and lexicalized orthographic knowledge. The current model proposes that letter knowledge develops before generalized knowledge which develops before lexicalized knowledge, but each of these domains is expected to overlap. For example, experience with specific words, like their name, may affect what kinds of letter combinations children perceive as illegal. Similarly, the progression of generalized knowledge from knowledge about the symbols within words to the pattern of those symbols, is a transition that requires more exploration.

These findings also provide support for statistical learning as the mechanism for the development of orthographic knowledge. Unlike other theories of development, statistical learning does not require either phonological or lexicalized knowledge to develop before generalized knowledge. Theoretically, children encounter and encode words in their environment, and process the probability of certain patterns within words. When I show them an illegal pattern, they can implicitly recognize it as unusual, but it is not until they have more experience that they are able to explicitly say that this
pattern is wrong or illegal. Our studies indicate that once children have sufficient exposure to written language, and have some knowledge of letters, they are able to recognize symbolic violations like that numbers or symbols do not make up words. As they continue to gain exposure to the probability of letter combinations, they are better able to distinguish that only certain letter patterns go together. Although these studies suggest that statistical learning is the most promising current theory of development, it should be noted that the statistical learning of orthographic patterns is relatively untested, and requires much more empirical support.

6.2 Generalized Orthographic Knowledge and Phonological Knowledge

Based on evidence from previous studies (Cassar & Treiman, 1997) we would expect the bulk of generalized orthographic knowledge to develop at the same time as phonological knowledge. These, and other studies (Treiman, 1993; Read, 1975), show, however, that in several literacy tasks children use phonology over orthography. The current studies demonstrate the particular tasks and circumstances under which children show generalized orthographic knowledge and phonological knowledge.

Although participants and methodologies in studies 2 and 3 are different, because they fall in the same age range, I am comparing the two groups. It should be noted, however, that without direct experimental manipulation this discussion should be considered speculative only. In study 2, I show that children tend to spell stop/stop clusters, which are unpronounceable as written, less frequently than stop/liquid clusters,
which are pronounceable. In study 3, we see that children show more implicit recognition of the violations that result in stop/stop clusters than stop/liquid clusters. Finally, study 4 shows, that when children are asked to make orthographic judgments and decode non-words, they fail to show implicit orthographic knowledge. Taken together these results suggest that children have orthographic knowledge, but this knowledge is tenuous, and when there is a phonological and orthographic conflict, children focus on phonology over orthography.

These results support a theory in which generalized orthographic knowledge is masked by phonological knowledge. When we fail to see recognition of illegal letter combinations in judgment task that requires decoding, it suggests that children aren’t accessing that orthographic knowledge. It seems as though children are relying on phonology to make their judgments and, as such, no orthographic knowledge shows. This indicates that while both types of knowledge are available the two sources are not yet integrated fully. Children have difficulty using both types of knowledge to make decisions because they can’t access both kinds of knowledge at the same time. Research on second language learners supports this hypothesis, in that new readers of alphabetic languages integrate orthography and phonology faster in shallower orthographies than deep orthographies (Brysbaert, et al., 1999; Van Wijnendaele & Brysbaert, 2002). With regards to why children default to using phonology rather than orthography, the sound of words is something children are familiar with. Children learn to speak before they
learn to read, and the spoken version of most words is what children know best. Again, evidence that second language learners supports this finding by showing that when early readers learn a language that has a different phonology-orthography link (i.e. speakers of an alphabetic language learning a logographic language) they tend to rely on orthographic patterns rather than phonological ones (Liu, Wang, & Perfetti, 2007). This suggests that readers rely on the strategy they think will serve them best to solve the reading ‘problem’. For children learning to read in English, the best strategy they have is a phonological one. Until orthographic and phonological information is integrated, children will rely on phonology over orthography.

This finding raises several potentially interesting points of future research regarding when and how children learn to incorporate phonological and generalized orthographic knowledge. Since some tasks seem to lend children to use orthographic knowledge more than others, it may be possible to devise a task that encourages children to use both of these streams of information. Training studies could be particularly helpful for determining children’s nascent ability to integrate this information. Similarly, while this research provides evidence for statistical learning theory as a mechanism for the development of generalized orthographic knowledge, investigating the mechanisms for integration of generalized orthographic and phonological knowledge is relatively unexplored and important for developing a full
representation of how generalized orthographic knowledge develops as readers gain more experience.

**6.3 Implicit and Explicit Knowledge**

These studies also provide evidence children demonstrate knowledge implicitly before they can demonstrate it explicitly. We also find that that implicit orthographic knowledge does not demonstrate once children have progressed to explicit knowledge. This suggests that implicit orthographic knowledge, in the form of looking time, is a temporary manifestation. Once children are able to make explicit judgments about what is a good word and what is a bad word, they no longer need to attend longer to the illegal combination.

If we apply the findings on children’s explicit judgment from study 4 to study 3, we can examine how children assess combinations that they do and do not explicitly understand. For the word-initial doublings, which study 4 shows that children explicitly recognize as illegal we see a difference in looking time such that children look less at the illegal combination. Children register the illegality immediately and only have to look at one of the letters to pronounce the sound, so, compared to a legal letter combination in the same position the looking time is significantly less. For word initial stop/stop clusters and word final stop/liquid clusters, which children do not show explicit knowledge of in Study 4, we see that when decoding children show implicit knowledge for the stop/stop clusters but not the stop/liquid clusters. In this way the level of
difficulty discussed in study 3 is supported by children’s implicit and explicit knowledge in study 4. Taking the two studies together we see a progression from implicit to explicit knowledge. Children demonstrate implicit knowledge, as measured through eye-tracking, only as long as they are unable to express their knowledge explicitly. Once they have the means to express themselves, they no longer need to look longer at an illegal combination.

In this way we can see implicit knowledge, as measured through looking time, as the state shortly before children show explicit knowledge. The illegal combination is outside their realm of experience and therefore more deserving of attention. The illegality of an orthographic combination is really just a measure of its rarity. Patterns are orthographically illegal for a number of reasons, but in order to learn the combinations it is sufficient to realize that they do not occur often in conventional English spelling. Because it is uncommon, it requires more attention to parse and make a decision about. Once children gain more experience with different combinations the unusual combinations can be more easily labeled as ‘illegal’, and quickly rejected as ‘good word’.

This data is analogous to research in other cognitive domains. For example Siegler & Stern (1998) show that children recognize the difference before they are able to make an explicit judgment about it, but closer to the research from Church and Goldin Meadow (1986), such that the implicit behaviour disappears once the explicit behaviour
has manifested. Once the knowledge is explicit it is no longer necessary to carefully
examine the combination, just like once knowledge can be verbalized, there is no need to
communicate the knowledge via gesture. The implicit knowledge doesn’t vanish once
explicit knowledge is attained; the material is just known well enough that it doesn’t
need to be examined in depth. Children’s implicit knowledge in the decoding process,
though it becomes masked in a judgment task, indicates that they are attending to the
pattern, and potentially gathering information on the frequency with which they
encounter a pattern. Again, this provides evidence for statistical learning as a
mechanism of development for orthographic knowledge. We can speculate that in every
encounter with writing, children are gathering data about patterns and once they have
enough information that it becomes explicit, the data ceases to be masked in any kind of
task. As children continue to decode and spell words, they go through the self-teaching
process and solidify and build up their generalized orthographic knowledge; which
develops into a strong explicit knowledge, and is no longer vulnerable to individual task
masking.

Learning to read is one of the major cognitive challenges of childhood. In order
to read successfully, children must go from knowledge of spoken language to
knowledge of all the different aspects and iterations of phonological, semantic, and
orthographic knowledge. This dissertation has examined a fairly small, but important,
aspect of this knowledge. By determining when and how children gain knowledge that
branches across words we can begin to understand how children make sense of the massive amount of stimuli they are presented with, in order to develop a procedure by which they decipher and understand written words.
### Table 9: Stimuli list for Chapter 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbols</td>
<td>&amp;*%$#</td>
</tr>
<tr>
<td>Numbers</td>
<td>89341</td>
</tr>
<tr>
<td>Repeated letter</td>
<td>AAAAAA</td>
</tr>
<tr>
<td></td>
<td>KKKKKK</td>
</tr>
<tr>
<td>symbol/letter</td>
<td>TR#MP</td>
</tr>
<tr>
<td>Number/letter</td>
<td>A8PER</td>
</tr>
<tr>
<td>Illegal letter combo</td>
<td>YLRKS</td>
</tr>
<tr>
<td></td>
<td>DLACK</td>
</tr>
<tr>
<td></td>
<td>NAOTR</td>
</tr>
<tr>
<td></td>
<td>ATDNM</td>
</tr>
<tr>
<td>Pseudoword</td>
<td>WURGS</td>
</tr>
<tr>
<td></td>
<td>BLUPY</td>
</tr>
<tr>
<td></td>
<td>MOART</td>
</tr>
<tr>
<td></td>
<td>EDRAP</td>
</tr>
</tbody>
</table>
### Appendix 2

#### Table 10: Stimuli list for Chapter 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>Variable</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Initial stop/stop</td>
<td>DPALE</td>
<td>DLAPE</td>
</tr>
<tr>
<td></td>
<td>BTELP</td>
<td>BLEPT</td>
</tr>
<tr>
<td></td>
<td>PGERN</td>
<td>PRENG</td>
</tr>
<tr>
<td></td>
<td>KDENS</td>
<td>KENDS</td>
</tr>
<tr>
<td>Word Final stop/liquid</td>
<td>PEATR</td>
<td>PEART</td>
</tr>
<tr>
<td></td>
<td>NOIDL</td>
<td>NOILD</td>
</tr>
<tr>
<td></td>
<td>TEAPL</td>
<td>TEALP</td>
</tr>
<tr>
<td></td>
<td>MOOGR</td>
<td>MOORG</td>
</tr>
<tr>
<td>Word Initial doubles</td>
<td>LLOSH</td>
<td>LOSH</td>
</tr>
<tr>
<td></td>
<td>LLEAT</td>
<td>LEAT</td>
</tr>
<tr>
<td></td>
<td>FFELP</td>
<td>FELP</td>
</tr>
<tr>
<td></td>
<td>SSALM</td>
<td>SALM</td>
</tr>
</tbody>
</table>
Appendix 3

Table 11: Stimuli list for Chapter 5

<table>
<thead>
<tr>
<th>Condition</th>
<th>No phonology</th>
<th>Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word initial stop/stop</td>
<td>DTEST/STEST</td>
<td>DTAM/DETAM</td>
</tr>
<tr>
<td></td>
<td>BTOMP/BROMP</td>
<td>BKON/BEKON</td>
</tr>
<tr>
<td></td>
<td>BKROF/SHROF</td>
<td>KGUMP/KEGUMP</td>
</tr>
<tr>
<td></td>
<td>PGANS/GRANS</td>
<td>PGAVE/PEGAVE</td>
</tr>
<tr>
<td>Word final stop/liquid</td>
<td>TWATR/TWART</td>
<td>TESPR/TESPER</td>
</tr>
<tr>
<td></td>
<td>BRUKL/BRULK</td>
<td>STEGL/STEGLE</td>
</tr>
<tr>
<td></td>
<td>RHUDR/RHURD</td>
<td>BAKL/BAKLE</td>
</tr>
<tr>
<td></td>
<td>SNOKL/SNOLK</td>
<td>FRITL/FRITLE</td>
</tr>
<tr>
<td></td>
<td>FROTL/FROLT</td>
<td>SCEPR/SCEPER</td>
</tr>
<tr>
<td></td>
<td>FLAPR/FLARP</td>
<td>BAMTR/BAMTER</td>
</tr>
<tr>
<td></td>
<td>DAWTL/DAWLT</td>
<td>DAIPL/DAIPLE</td>
</tr>
<tr>
<td></td>
<td>MOATR/MOART</td>
<td>HIBR/HIBER</td>
</tr>
<tr>
<td>Word initial doubles</td>
<td>LLIEP/PIELL</td>
<td>FFULP/FULP</td>
</tr>
<tr>
<td></td>
<td>LLIRN/NIRLL</td>
<td>LLORT/LORT</td>
</tr>
<tr>
<td></td>
<td>FFURC/CRUFF</td>
<td>LUPS/LLUPS</td>
</tr>
<tr>
<td></td>
<td>SSAYT/TAYSS</td>
<td>SSANT/SANT</td>
</tr>
<tr>
<td>Length control</td>
<td></td>
<td>BLARK/B#LARK</td>
</tr>
<tr>
<td>Term</td>
<td>Term</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>KESTS/KES&amp;T</td>
<td>DWALL/D*WALL</td>
<td></td>
</tr>
<tr>
<td>FRIMB/F2FRIMB</td>
<td>RUNCH/RU%NCH</td>
<td></td>
</tr>
<tr>
<td>KELPS/KELP4S</td>
<td>STEET/STEE8T</td>
<td></td>
</tr>
<tr>
<td>KONTS/KO9NTS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Nation, K., & Hulme, C. (2001). The limitations of orthographic analogy in early reading development: Performance on the clue-word task depends on phonological
priming and elementary decoding skill, not the use of orthographic analogy. *Journal of Experimental Child Psychology, 80*(1), 75-94.


Biography

Tanya Kaefer was born in Vancouver, British Columbia on October 15th, 1980. Tanya received her Bachelor's Degree from University of Waterloo in Waterloo, Ontario in 2004, where she completed her Honors thesis in Psychology (her major). Tanya began her studies in Psychology at Duke in the fall of 2004 and received her M.A. in psychology from Duke in 2006. Over the course of her graduate education, Tanya was awarded the following academic honors:

- Duke Graduate School Summer Research Fellowship (2009)
- Sulzberger Social Policy Fellow (2008)
- Spencer Foundation: Discipline-Based Graduate Training Program in Education Science and Policy (2005-2006)

Tanya currently resides in Durham with her dog Rudy.