The Development of the Phonetic Similarity Effect and its Relationship to Early Reading Acquisition in Children

Debra Susan Lean

A Thesis

in

The Department

of

Psychology

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montréal, Québec, Canada

November 1986

ISBN 0-315-97689-6

ABSTRACT

The Development of the Phonetic Similarity Effect and its Relationship to Early Reading Acquisition in Children

Debra Lean, Ph.D.
Concordia University, 1986

The present study examined the development of the phonetic similarity effect and its relationship to reading skill in children over the normal period of early reading acquisition. The effect occurs when phonetically similar stimuli are not recalled as well as phonetically dissimilar stimuli. Children in three Grade level cohorts (28 in Kindergarten, 31 in Grade 1 and 29 in Grade 2) were given three memory tasks measuring the phonetic similarity effect with letters, words and sentences and were retested on the same tasks one year later. Within each cohort, the children represented a wide range of prereading and reading abilities as measured by standardized tests. The first hypothesis of the study, that there is a phonetic similarity effect throughout the early reading acquisition period, was supported. Developmental analyses for each memory task showed that phonetic similarity affected memory in the letter and sentence tasks in all cohorts in both years of testing. The phonetic similarity effect was less marked with the word task, and again there was no evidence of an increase in the effect with development. A second hypothesis, that phonetic coding ability is positively related to reading achievement, was also supported. The results of hierarchical multiple regression analyses showed that the letter task and to a lesser extent the word task were positively related to reading achievement in Kindergarten and Grade 1, and were predictive of reading achievement in Grade 2. The findings were discussed with respect to the processes involved in early reading acquisition and to their potential applicability to the identification of children at risk for reading disability.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Tannis Arbuckle-Maag, for her support, guidance and thoughtful supervision throughout my graduate career. I would also like to thank Dr. Anna Beth Doyle and Dr. Norman Segalowitz for their helpful comments and interest in my research. Dr. Gloria Waters' efforts and assistance were invaluable in the formulation of the research.

The author was supported by a Social Science and Humanities Research Council of Canada Doctoral Fellowhip. The research costs were supported by an external FCAC grant awarded to A.B. Doyle, T. Arbuckle, D. Gold, E. Jacobs, L. Sherman and D. White by the Quebec Ministry of Education entitled "An interdisciplinary study of preschool development".

I would also like to express my appreciation to the staff and students of Bedford, Coronation and Somerled Schools, who were most cooperative.

I wish to thank the staff at the McGill-Montreal Children's Hospital Learning Centre for their helpful support over the past three years. Special appreciation goes to Dr. Gerry Taylor for his interest in my research, his much appreciated insights into the reading literature, and his willingness to answer my endless questions. I would also like to express my gratitude to Dr. Margaret Bruck for her thoughtful comments and computer graphing expertise. A note of thanks also goes to Joan S[illegible], who helped me to unravel the mysteries of Appleworks, typed the tables, and was always available with a sympathetic ear.

My graduate studies at Concordia have been enriched by wonderful friends and fellow students. I would like to thank Denise Messmer and Sharna Olfman who have been extremely supportive and caring throughout my graduate years.

I would like to thank my parents, Stan and Lilyan Lean, for their constant support and for instilling in me a willingness to work hard and achieve my goals, my sister Shelley Shusterman and her family for their encouragement and my in-laws Gerry and Pearl Ginsherman for their understanding. A special thank you goes to Honey Agar for her inspiration.

The hardest person to thank is my husband, Martin. He took the time to gain a clear understanding of my work and was always willing to listen and contribute. His patience and unending emotional support made the rough parts bearable. Together we proved that it is possible for husband and wife to write a thesis at the same time and survive.

I dedicate this thesis to the memory of my Aunt, Frances Faye Ngai, who in many so ways helped me to achieve my academic milestones and took pride in sharing them with me.

TABLE OF CONTENTS

LIST OF TABLES

PAGE

LIST OF FIGURES

It has been estimated that 10 to 25% of English-speaking children fail to learn how to read (Helfgott, 1976). An even larger percentage of children encounter at least a passing amount of difficulty when they begin to learn how to read (Belmont & Belmont, 1978; 1980). Recent research on reading acquisition has begun to address these problems. Major areas of inquiry include determining why such a large proportion of children have problems in learning how to read, comparing different teaching methods in order to find out how to remediate these problems, and predicting at the prereading stage which children are at risk for problems so that early intervention may be carried out.

Research on reading disabilities has shown that the majority of reading problems are related to difficulties with the linguistic nature of reading, as opposed to the auditory, motor or visual components of the reading process (Vellutino, 1979). Consequently, much of the research on the skills involved in reading acquisition has focussed on those that are linguistically based.

One skill that has been assumed to be relevant to reading acquisition is the use of phonetic codes in short-term memory (Baddeley, 1978). A measure commonly used to differentiate good and poor readers is the children's sensitivity to the phonetic similarity effect, which reflects how recall is affected by rhyming confusion in a short-term memory task. In general, the effect occurs when phonetically similar stimuli are remembered less well than phonetically dissimilar stimuli. In such a task, poor readers generally are found to be less affected than good readers by rhyming confusion. This finding has been interpreted as evidence that poor readers are either delayed or deficient relative to normal readers in their acquisition of phonetic coding abilities and that their lack of phonetic coding skills over the normal reading acquisition period is one of the primary reasons for their difficulties in learning to read. However, interpretation of the findings with poor readers is unclear because the age boundaries at which normal readers show the phonetic similarity effect have not been established. This state of affairs is mainly due to the paucity of research concerned with the development of the phonetic similarity effect during the reading acquisition stage.

The purpose of the present project was to investigate the development of the phonetic similarity effect in children over the normal period of early reading acquisition. A secondary purpose was to examine the relationship between the phonetic similarity effect and reading ability within this early reading period. The literature review for this project is presented in the following manner. The first section deals with the historical development of the phonetic similarity effect, its role in short-term memory storage processes in reading and speculations on its involvement in reading acquisition in children. The second section deals with empirical evidence for the hypothesis that phonetic coding is related to reading ability.

The Phonetic Similarity Effect and its Role in Reading Acquisition

This section begins by examining the empirical development of the phonetic similarity effect and its hypothesized role in short-term memory storage processes in reading. Although reading involves many processes, short-term memory is thought to be one that is vital in beginning reading, when word identification processes are not highly automated (e.g., Perfetti, 1985). The role of short-term memory in reading appears to involve holding information gained from the visual input in a short-term memory store long enough to abstract it into phonemes, words or meaningful phrases. Research with adult subjects suggests that this short-term storage is accomplished by coding the information into a representation that is speech-based or phonetic. For example, Conrad (1964) found that adults coded unrelated letters phonetically during visual memory or auditory listening tasks. Errors in ordered recall tended to be phonetically similar to the correct letter (e.g., if the subjects forgot the letter "D", they would be more likely to recall the letter "B" than "H"). In an additional experiment (Corrad and Hull, 1964), adult subjects had more difficulty recalling strings of phonetically similar letters (e.g., B,C,D,P,T) than strings of letters that differed in sound or articulation (e.g., H,N,Q,R,W). Baddeley (1976) reported a series of experiments where adults were tested for recall of visually presented items that were phonetically, visually or semantically similar. The typical finding was that phonetically similar items were not recalled as easily as the other two types of items. This evidence suggests that the subjects were coding the items in a phonetic format and that the phonetic similarity interfered with memory. Impaired memory for phonetically similar stimuli in comparison to memory for phonetically dissimilar stimuli is termed the phonetic similarity effect. The effect suggests that perceivers have so strong a tendency to store information in short-term memory on a phonetic basis that they persist in using a phonetic coding strategy even when it creates interference and penalizes recall.

Baddeley and Hitch (1974) consolidated the findings from research on the phonetic similarity effect and developed a "working memory" approach to explain the role of storage systems in short-term memory. They hypothesized a division of working memory into at least two components. The first part of the system, called the "central executive", has information processing and decision-making roles, as well as being used for storage. The central executive is aided by a slave-like system, called the "secondary articulatory loop", which is used for sub-vocal rehearsal of material. The loop is hypothesized to operate like a tape loop; it is temporally based, and has a limited duration of approximately 2 seconds. As described by Baddeley, the presence of this loop in the working memory system enables any material that can be verbalized to be stored in a way that makes minimal demands on the central executive.

Baddeley (1979) suggested that these two components of working memory play an important part in beginning reading. He hypothesized that, in beginning reading, the central executive is involved in translating letters into phonemes, and in enabling the reader to progress from a series of unrelated sounds to words. The central executive can also be used to hold the decoded speech sounds in a temporary store until the reader can integrate the separate sounds and abstract the meaning of the word, phrase or sentence. However, because the central executive's space is limited, using it for essentially a memory storage role would decrease the processing space available for decoding and blending activities. Baddeley envisioned the secondary articulatory loop as a supplementary way of holding onto sequences of phonemes, syllables and words. As each phoneme is decoded by the central executive, it is stored in the secondary articulatory loop, thereby leaving the central executive free to decode the next phoneme and perhaps hypothesize about what the final word may be. When two or three phonemes have been loaded into the loop, they can be blended by the central executive into a syllable. This syllable is then returned to the loop, and is held there until further phonemes are decoded, which are then in turn blended with the initial syllable. This process continues until sufficient information has been processed for the beginning reader to decode the meaning of the word, phrase or clause. Baddeley contrasted this method with the strategy of using the central executive for both storage and general processing, where each decoded letter would use up space in the central executive leaving progressively less room for letter analysis, blending and word identification.

On the basis of his model, Baddeley hypothesized that children who have difficulty with, or are unable to use, the secondary articulatory loop may experience difficulty in learning to read. These children would rely on the central executive for both general processing and storage purposes. Baddeley did not attempt to test this hypothesis specifically, although he cited findings which suggested that beginning and disabled readers were not transferring the information to the secondary articulatory loop for short-term storage.

Several studies have demonstrated that, consistent with what would be predicted on the basis of Baddeley's model, poor readers do indeed experience difficulties with cognitive activities that deal with the phonetic basis of reading. For example, beginning and below average readers have been shown to have difficulty with recognizing, analyzing and segmenting phonemes and syllables (e.g., Calfee, Lindamood & Lindamood, 1973; Fox & Routh, 1983; McNinch & Richmond, 1972; Rosner & Simon, 1971; Zifcak, 1981). Liberman, Shankweiler, Liberman, Fowler and Fischer (1977) found that beginning and disabled readers made more reading errors on final consonants than on initial ones, which shows that they are not performing phonemic segmentation properly. An example would be a child presented with the word "bag" would read "butterfly", thus showing that he or she is not aware of the word's phonetic structure. Preschoolers trained in phonemic awareness skills have been found to perform reading-like tasks as well as children who already had phonemic awareness skills and better than untrained children (Treiman and Baron, 1981; Bradley and Bryant, 1983; Fox and Routh, 1983, 1984; Torneus, 1984). Torgesen (1978-1979, 1985) reviewed studies that show beginning and disabled readers have a great deal of difficulty with memory for ordered verbal information, for example, digit span.

In presenting his model, Baddeley (1979) cited some preliminary research which showed that good and poor readers are differentially sensitive to the phonetic similarity effect as measured by recall of rhyming and nonrhyming letters (Shankweiler & Liberman, 1976). Since 1976 there has been a great deal of research comparing good and poor readers on various rhyming confusion tasks that involve phonetic coding. The following section reviews the literature on relations between the phonetic similarity effect and reading ability. The review is divided into three areas which correspond to three ways in which phonetic coding in children has been examined in the literature: (a) comparisons between good and poor readers within a single age level, (b) comparisons between good and poor readers across age levels using cross-sectional and longitudinal analyses and (c) phonetic coding before reading acquisition occurs, that is, how it develops in prereaders.

The Phonetic Similarity Effect and Reading Ability

Comparison of Good and Poor Readers Within a Single Age Level

Baddeley (1978, 1979) hypothesized that some poor readers have difficulty with or cannot use phonetic coding in working memory, and speculated that poor readers would likely be less affected than the good readers by phonetic similarity. A group of researchers at Haskins Laboratories in Connecticut performed several studies that tested Baddeley's speculations. They examined the differences in phonetic coding ability between Grade

2 good and poor readers using the phonetic similarity effect paradigm. It was expected that the rhyming stimuli would generate confusion and thus penalize children who used a phonetic code. The researchers expected that good readers would behave much like the adult subjects in Conrad and Hull's (1964) and Baddeley's (1976) studies by showing a phonetic similarity effect, whereas the poor readers would not show such an effect.

The Haskins group initially examined serial recall for rhyming and nonrhyming letters that were presented auditorily and visually (Shankweiler, Liberman, Mark, Fowler & Fischer, 1979) and for rhyming and nonrhyming word strings and sentences that were presented auditorily (Mann, Liberman & Shankweiler, 1980). In both studies, the good readers generally surpassed poor readers in overall recall. However, the overall better performance of the good readers was attributable to their superior recall of nonrhyming stimuli. The good readers's recall of rhyming stimuli fell to the level of the poor readers. The good readers showed this phonetic similarity effect in their recall for visually and auditorily presented letters and auditorily presented words and sentences, while poor readers did not show any significant difference in recall between rhyming and nonrhyming stimuli under any task conditions.

The Haskins group of researchers speculated that the difference in the phonetic similarity effect between the good and poor readers could be due to more rehearsal of the materials by good readers than by poor readers. Mark, Shankweiler, Liberman and Fowler (1977) examined the phonetic similarity effect in good and poor readers with a task that minimized the opportunity for rehearsal. Good and poor readers were asked to read aloud words printed on flash cards. Then a surprise recognition test was presented that consisted of the original words and 14 foil words that were either phonetically similar (rhyming) or dissimilar (nonrhyming) to the original words. Children were asked to read the words aloud and to respond "yes" if they thought the word was from the original list and "no" if they thought the word was new.

Good readers made more false positive responses to the rhyming foils than to the nonrhyming foils. Poor readers made roughly equal numbers of false positive errors on the two types of foils. The good readers showed a phonetic similarity effect as defined by a greater number of rhyme confusion errors. Poor readers did not show any differential sensitivity to either type of foil, and therefore did not show evidence of a phonetic similarity effect. Mark and his colleagues concluded that the letter task findings of the Shankweiler et al. (1979) study could not be solely attributed to differences in rehearsal strategies between good and poor readers. The researchers concluded that good and poor readers differed in their use of phonetic coding in working memory, regardless of type of stimuli or modality of access.

Byrne and Shea (1979) extended the Haskins results to an auditory recognition memory task using a sample of Australian Grade 2 good and poor readers (in Australia formal reading instruction begins in Kindergarten). The children heard a list of words where later items were related either semantically (as synonyms or antonyms) or phonetically (as rhymes) to earlier ones. The children were required to indicate whether each word had previously appeared in the list. Byrne and Shea took false positive responses to foils as indices of semantic or phonetic coding.

The results were consistent with the Haskins groups' findings in showing a greater sensitivity to rhyme confusion in good readers. Good readers made equal numbers of false positive responses to semantic and rhyming foils. In contrast, poor readers made a large number of semantic false positive responses, but almost no rhyme-based false positive responses. Thus, the poor readers did not show evidence of using phonetic coding, while the good readers appeared to be using both phonetic and semantic coding.

Byrne and Shea conducted another experiment with the same children in order to see what would happen when phonetic coding was the only possible choice available. The children were tested with the same procedure using pronounecable nonsense words with rhyming and nonrhyming foils (e.g., "jome" as the target, "vome" as the rhyming foil and "fove" as the nonrhyming foil) . The researchers found that poor readers made a small but significant number of phonetically-based false positive responses, although the good readers made more. Byrne and Shea concluded that the poor readers in their sample were relatively insensitive to phonetic coding, and that this weakness was most marked when sound and meaning were both available as memory codes.

The studies discussed above show that poor readers have difficulty with the use of phonetic memory codes in recall of visually and auditorily presented letters, auditorily presented words and sentences and recognition of visually and auditorily presented words. These results suggest that poor readers have a general problem with the use of a phonetic code in short-term memory. This difficulty occurs with several memory paradigms, and in both the auditory and visual modalities. Poor readers appear therefore to be relatively insensitive to phonetic characteristics of certain linguistic stimuli.

Brady, Shankweiler and Mann (1983), working at Haskins Laboratories, specifically examined the source of poor readers' difficulties with the phonetic similarity effect. These researchers hypothesized that poor readers' phonetic coding problems may be due to a disorder in perceptual processing and may be associated with problems in speech perception. The authors conducted three experiments using Grade 3 good and poor readers. The first experiment basically replicated Mann et al.'s (1980) experiment comparing the good and poor readers' performance on recall of phonetically similar and dissimilar word strings. Brady et al's results were similar to Mann et al's; poor readers were less affected by the phonetic characteristics of the stimuli than were the good readers. In addition, poor readers made more transposition errors in their recall of the words, which the authors took as a further indication of the poor readers' problems with memory for order. The second and third experiments consisted of testing the same children on two auditory perception tasks.

Each task involved the presentation of stimuli under two conditions: one with a favourable signal to noise ratio and one with masking noise. Experiment 2 utilized spoken words that were chosen to control for syllable pattern, phonetic composition, and word frequency (e.g., door, bale, list, rasp). The poor readers made significantly more errors than the good readers when listening to speech sounds in noise. Experiment 3 utilized recorded environmental sounds (e.g., dog barking, thunder). The two reading groups did not differ in the perception of nonspeech environmental sounds, whether noise-masked or not. The results showed that the poor readers could process the speech sound adequately, but that they required a higher quality signal, or more complete stimulus information.

Brady and her colleagues found deficits in serial recall and speech perception in the same group of poor readers. It appears that poor readers may have difficulty in the perception as well as the recall of linguistic items. This may help to explain the pervasiveness of reading level differences with the phonetic similarity effect in auditorily as well as visually presented stimulus items.

The finding that poor readers have difficulty in short-term memory for phonetically codable stimuli has not been replicated in all studies and hence questions have been raised about the validity and generality of the effect. Two research groups have reported a series of experiments which, in general, fail to support the Haskins findings. The first group is Hall and his colleagues at Northwestern University. Hall, Ewing, Tinzman and

Wilson (1981) argued that the poor readers' low performance on both rhyming and nonrhyming lists found by Shankweiler et al. (1979) was due to their procedures in which the children were tested in groups and required to write down the letters. Hall and his colleagues stated that poor readers tend to be slower and less accurate in writing letters. For that reason they tested their subjects individually and asked for oral responses. A second issue, noted by Hall's group of researchers was that the Haskins work had not tested severely disabled readers. In the Hall et al. study the poor readers represented more extreme cases of reading disability. They consisted of six adolescents from special education classes who were delayed in reading by at least five grades and two adults who were described as being very deficient in reading. The normal reading group consisted of 15 children in Grades 3 and 4 who were reading at or above grade level.

The subjects were presented 16 rhyming and nonrhyming letter strings auditorily. Both normal and poor readers showed large decrements in oral recall of rhyming letters in comparison with nonrhyming letters. Hall et al. concluded that, contrary to previous findings, the poor reading group was as highly sensitive to phonetic similarity as were the normal readers.

Shankweiler, Liberman and Mark (1982) responded to Hall et al.'s article with several objections. Hall and his colleagues' findings were based on only eight poor readers who varied in age from 15 to 40 years. Shankweiler et al. questioned (a) the appropriateness of grouping together such a diversity of ages in a memory test and (b) the validity of comparing those group results with those obtained with third and fourth graders. They also objected to the fact that neither the reading levels nor the nature of the disabled group's reading problems were properly reported.

In a further examination of the generalizability and validity of the Shankweiler et al. (1979) study, Hall, Wilson, Humphreys, Tinzman and Bowyer (1983) noted that in the Shankweiler et al. experiment the overall poorer letter recall by the poor readers on both types of materials indicated that for them the task was relatively difficult. If there were floor effects for the poor readers on the nonrhyming materials, these would make it difficult, if not impossible, for them to show a phonetic similarity effect. Thus, the observed differences in the phonetic similarity effect between the two reading groups would simply be an artifact of floor effects in the poor overall performance of the poor readers on the memory tasks.

Hall et al. (1983) performed a series of experiments using poor reading groups that were not likely to be generally low in achievement, as they suspected Shankweiler et al.'s (1979) poor readers had been, and that were not markedly different from the good readers in terms of age, as had been the case in the Hall et al. (1981) study. The children were given standardized tests measuring general intelligence, reading and math skills. The poor readers performed at least at an average level on tests measuring general intelligence and math skills, and performed at a below average level in reading skills only. The good readers performed at a level that was at least average on all three measures. The first two experiments involved a replication of Shankweiler et al.'s (1979) visual and auditory experiments, except that there were two letter string lengths, recall was oral, children in the visual experiment were in grades 2 and 3, and children in the auditory experiment were in grades 3 and 4. Hall and his colleagues found that the poor readers in these two experiments showed a phonetic similarity effect, with both 4- and 5-letter string lengths.

A third experiment involved testing similarly selected grade 3 and 4 good and poor readers using lists of rhyming and nonrhyming auditorily presented word strings. Good readers had a slightly, but not significantly, larger phonetic similarity effect than the poor readers. Hall and his colleagues noted that the poor readers also had very low scores on the nonrhyming lists and suggested that the relative absence of a more pronounced phonetic similarity effect in the poor readers group was attributable to a floor effect for that group on the task.

The subjects in the fourth experiment consisted of Grade 3 and 4 good and poor readers, plus a group of generally low achievers who scored below average on all three scales of the screening battery. The researchers expected that the low achieving poor readers would be less likely to show a phonetic similarity effect than poor readers whose problems were solely related to reading. All three groups were presented 4- and 5-letter strings auditorily. The good and poor reading groups showed a phonetic similarity effect with both letter lengths. The low achievers showed a phonetic similarity effect only with the shorter, 4-letter lists. This finding suggests that these children were susceptible to phonetic coding when the length of the stimulus string was within their capacity and, by implication, that other studies may have been testing children beyond their limits.

A fifth experiment was conducted to examine the possibility that phonetic similarity effects in letter recall tasks may be diminished when task difficulty is high. Twelve college students heard 7-letter rhyming and nonrhyming strings. Task difficulty was varied by the insertion of a brief auditory counting task between presentation and recall on half of the lists, in order to generate interference in auditory storage. The subjects showed a large phonetic similarity effect in the normal condition but the effect was completely eliminated under the interference condition. The authors concluded that the elimination of the phonetic similarity effect in poor readers could be explained analogously by the fact that the task is more difficult for them.

Hall and his colleagues summarized the conclusions from the five studies by stating that their subjects' sensitivity to the phonetic similarity effect was highly dependent on the task difficulty when the nonrhyming stimuli were presented. They cautioned that when different subject populations are being compared, task difficulty must be at an appropriate level for both populations. If this control is not incorporated, then there will be the possibility that observed interactions between reading group and task conditions are simply artifacts of ceiling or floor effects differentially affecting the two groups on the different levels of task.

Hall et al.'s addition of the experiment with college students is of questionable relevance to studies with beginning readers. These subjects were most likely skilled readers, and, according to research with adults, (see e.g., Waters, Komoda & Arbuckle, 1985), would be capable of switching to nonphonetic codes when interference tasks made the use of phonetic coding less helpful.

The second research group to challenge the Haskins laboratory findings was Brown, Sanocki and Schrot (1983). These researchers attempted to address a different methodological problem in the Haskins studies than that addressed by Hall and his group. They pointed out that only one of the Haskins studies had involved a comparison of visual and auditory presentation of stimuli, and that this comparison was done in separate experiments (Shankweiler et al., 1979, Experiments 2 and 3). Brown et al. found this puzzling because the study involved the same subjects and counterbalancing procedures, which would have qualified the two experiments as a single study with modality of presentation as a within subjects variable. The fact that the studies were reported as separate experiments raises the possibility that factors other than modality may have varied between them.

Brown and his colleagues used phonetic similarity and presentation modality as independent variables. The poor readers attended a special class for reading improvement and were identified by teacher evaluations and reading comprehension scores. The good readers were identified by their teachers' evaluations only. Sets of rhyming and nonrhyming letters were presented both auditorily and visually. Within each modality, there were equal numbers of rhyming and nonrhyming trials.

Brown and his colleagues found a modality effect, but only in interaction with reading ability. The poor readers showed a significant phonetic similarity effect in the auditory modality only. Brown et al. stated that good readers were more involved in phonetic memory coding than poor readers. However, both types of readers showed more evidence of phonetic coding when letters were presented auditorily than when they were presented visually. The authors theorized that auditory presentation may make the phonetic characteristics of items more salient and thus more likely to be used as a basis for short-term memory coding. Brown and his colleagues mentioned that the visual presentation may also have offered alternative ways of coding items, and therefore the children were not as likely to use the phonetic features as a basis for coding. The authors concluded that phonetic coding is a potentially important source of variance in reading ability, but cautioned that phonetic coding deficiency in poor readers may not be as pervasive as suggested previously.

The studies reviewed above generally show that some groups of poor readers are less sensitive to the phonetic similarity effect than good readers, thus supporting Baddeley's theory that poor readers may have difficulty with phonetic coding in working memory. However, the evidence is far from clear. For example, Hall et al.'s (1983) main premise that floor effects were responsible for the interactions found in Shankweiler et al.'s (1979) study does not account for the fact that similar interactions were found in other studies using different modalities and recall tasks where floor effects were not evident. There were no floor effects in Mann et al.'s (1980) study, where the poor readers performed as well as good readers. This confusion may be due to methodological differences from study to study. As can be seen in the first two pages of Table 1, five standardized reading tests were used, along with four standardized intelligence tests. The scores used to differentiate good and poor readers ranged from grade equivalents to stanines, and in two cases, the groups were reported merely as differing significantly. Although the fact that many of these studies were done independently by different experimenters could explain the variety in reading measures, even the Haskins researchers used different reading measures across their different experiments. Furthermore, the cut-off points for stanines or grade equivalents also varied among studies, thus making it difficult to determine whether the same types of poor readers were being compared. The effects of differences in determination of reading groups was shown in the Hall et al. (1983) study where they attempted to address this problem with their determination of reading groups. Nevertheless, the interpretation of Hall's results is far from clear. Both Hall et al.'s (1983) and Brown et al.'s (1983) research used a wide and varying combination of grades for the individual experiments. Hall et al. combined two grades in each of the experimental analyses and summed over three different grade levels (2-4) in

Table 1

Comparison of Reading and Intelligence Measures

and Developmental Levels Across Studies

Study	Reading Test[a]	Reading Test Score	Intelligence Test[b]	Developmental Level
Single Age Level Studies				
Shankweiler et al. (1979)	WRAT Reading	Grade Equivalents	PPVT	Grade 2
Mann et al. (1980)	WRMT Word Recognition and Word Attack	Error Scores	SIT	Grade 2
Mark et al. (1977)	WRAT Reading	Grade Equivalents	WISC-R	Grade 2
Byrne & Shea (1979)	SLWRT	Reading Age	Not tested	Grade 2
Brady et al. (1983)	WRMT Word Recognition and Word Attack	Error Scores	PPVT	Grade 3

Table 1 (Cont'd)

Study	Reading Test[a]	Reading Test Score	Intelligence Test[b]	Developmental Level
Hall et al. (1981)	WJPTB Reading (Good readers only)	Grade Equivalents	Not tested	Good readers - 8 & 9 years; "Dyslexics" - 15-40 years
Hall et al. (1983)	WJPTB Reading and Math	Standard Scores	WJPTB Brief Cognitive	Grades 2-4
Brown et al. (1983)	SDRT (Poor readers only)	Unknown	PPVT (Poor readers only)	Mean ages- Poor readers - 8.5 years; Normals - 9 years

Table 1 (Cont'd)

Study	Reading Test[a]	Reading Test Score	Intelligence Test[b]	Developmental Level
Multi-Age Level Studies				
Siegel & Linder (1984)	WRAT Reading	Percentiles	PPVT	7-13 years
Siegel & Ryan (1984)	WRAT Reading	Percentiles	PPVT	7-14 years
Olson et al. (1984)	PIAT Reading	Grade Levels	WISC-R	Mean ages 7.8-16.8 years
Johnston (1982)	BAS Reading	Reading Ages	EPVT	9-14 years
Bisanz et al. (1984)	SGRVT	Reading Ages	CCAT	Grades 2-6
Beggs & Howarth (1985)	GAP Reading Test	Reading Ages	Not tested	7-14 years
Mann & Liberman (1984)	WRMT Word Recognition and Word Attack	Raw Scores	PPVT	Grades K-1
Share et al. (1984)	NARA	Standard Scores and Informal Tests	PPVT	Grades K-1

Table 1 (Cont'd)

a

WRAT - Wide Range Achievement Test (Jastak, Bijou, & Jastak, 1965)
WRMT - Woodcock Reading Mastery Test (Woodcock, 1973)
SLWRT - St. Lucia Word Reading Test
WJPTB - Woodcock-Johnson Psychoeducational Test Battery (Woodcock & Johnson, 1973)
SDRT - Stanford Diagnostic Reading Test (Karlsen, Madden, & Gardner, 1977)
PIAT - Peabody Individual Achievement Test (Dunn & Markwordt, 1970)
BAS - British Abilities Scale
SCRVT - Schonell Graded Reading Vocabulary Test (Schonell, 1986)
GAP - GAP Reading Test (McClecd, 1965)
NARA - Neale Analysis of Reading Ability (Neale, 1966)

b

PPVT - Peabody Picture Vocabulary Test (Dunn, 1966)
SIT - Slosson Intelligence Test (Slosson, 1963)
WISC-R - Wechsler Intelligence Scale for Children - Revised (Wechsler, 1974)
WJPTB - Woodcock-Johnson Psychoeducational Test Battery (Woodcock & Johnson, 1973)
EPVT - English Picture Vocabulary Test (British Standards of PPVT)
CCAT - Canadian Cognitive Abilities Test

their conclusions. Brown et al. mentioned the mean ages of their subjects, but they did not record any age ranges. Therefore, it is unclear from the published information whether their data came from a broad or narrow age range. The subskills important to reading are assumed to change markedly during the ages tested in these studies (Chall, 1979; 1983). One wonders about the extent to which developmental changes over the age range studied were relevant to the results.

Comparison of Good and Poor Readers Across Age Levels

Some recent studies have specifically examined the developmental aspects of the relation between reading ability and the phonetic similarity effect. Two hypotheses have been proposed to account for the apparent difficulty of poor readers with phonetic coding. The developmental deficit hypothesis states that poor readers cannot code phonetically and will not develop this ability on their own. This hypothesis assumes that the poor readers will not "catch up" with good readers in their phonetic coding skills. In contrast, the developmental delay hypothesis states that poor readers do develop phonetic coding skills, but at a slower rate than good readers and perhaps at a later stage in reading development. Poor readers of equivalent chronological age are assumed to be performing at a less mature level than good readers but could potentially "catch up" in phonetic coding with the good readers. Determining whether poor readers are delayed or deficient in their phonetic coding abilities has important implications for remediation techniques. If poor readers are delayed, remediation would likely involve more practice in phonetic coding and exposure to print so that they would catch up with same aged good readers. On the other hand, if poor readers are deficient in phonetic coding skills, they would have to be "taught" phonetic coding or guided towards using an alternative type of coding for short-term storage.

Cross-sectional Studies. The following studies (Siegel & Linder, 1984; Siegel & Ryan, 1984; and Olson, Davidson, Kliegl & Davies, 1984) use cross-sectional designs to examine the development of the phonetic similarity effect in children.

Siegel and Linder (1984) examined the development of the phonetic similarity effect in Canadian children aged 7 to 13 who were good or poor readers. They also added a subject group that had arithmetic disabilities in order to see if deficits in the phonetic similarity effect were limited to reading-disabled children or were characteristic of learning-disabled children in general. The poor readers were defined by their below average scores on a standardized word recognition test but scored within normal limits on a standardized math test. The arithmetic-disabled children scored well below average on the math test but were within normal limits on a reading test. The control children were in Grades 1 to 8 and scored well above average on both the arithmetic and reading tests.

The stimuli consisted of 5-letter and 6-letter rhyming and nonrhyming strings presented simultaneously. There were three conditions: visual presentation with oral recall, visual presentation with written recall and auditory presentation with written recall.

Siegel and Linder at first compared reading-disabled children with the control group. The youngest poor readers (children aged 7 to 8) did not show evidence of a phonetic similarity effect. The youngest control group and all children in the older age groups (children aged 9 to 10 and 11 to 13) did show evidence of a phonetic similarity effect. This pattern of results was essentially the same for both list lengths, for auditory and visual presentations and for oral and written recall.

The results were less consistent across conditions when children with arithmetic disabilities were compared with the control group. In these comparisons, all groups showed a phonetic similarity effect under all conditions except for the youngest group of arithmetic disabled children with a visual-written presentation. The authors suggested that the use of a phonetic code in short-term memory seems to develop more slowly in learning-disabled children, but it does develop. They go on to say that their data support a developmental delay hypothesis in which reading and arithmetic disabilities represent a delay in maturation as opposed to a deficit.

In an unpublished report, Siegel and Ryan (1984) replicated the study with groups of good and poor readers aged 7 to 14 years, as defined by performance on a standardized word recognition test. The exposure time of the letter stimuli was extended to 6 seconds from the 3 seconds employed by Siegel and Linder (1984). The poor readers still had significantly lower overall scores than the control group, reinforcing the suggestion that poor readers have difficulty with short-term memory. However, all the poor readers showed evidence of a phonetic similarity effect, as did the control group. The authors concluded that the poor readers in their study have some phonetic skills, but their development is significantly slower and less efficient than that of the control group.

Olson, Davidson, Kliegl and Davies (1984) matched 141 pairs of good and poor readers between the ages of 7.8 and 16.8 years of age on sex, age and socioeconomic status. Reading ability was defined by performance on a standardized reading recognition test.

Olson and his colleagues utilized the Mark et al. (1977) false recognition paradigm where good and poor readers were asked to read a list of words and later were given a recognition test that included the original words and rhyming and nonrhyming foils. Children were assumed to be coding phonetically when they falsely recognized the rhyming foils more than the nonrhyming foils.

In an analysis that grouped together all ages, both good and poor readers showed a phonetic similarity effect. The differences between rhyming and nonrhyming letters were slightly larger for the good readers, but the Reading Group by Word Type (rhyming or nonrhyming) interaction was not significant.

Olson and his colleagues found that the younger poor readers (about age 7 to 8) showed little evidence of phonetic coding, whereas the younger good readers did, thus replicating Mark et al.'s (1977) study. Olson et al. therefore examined how the phonetic similarity effect develops with age. The phonetic similarity effect was found to increase with age for disabled readers, a finding consistent with the developmental delay hypothesis. For normal readers, the phonetic similarity effect decreased significantly with age. The authors attributed this decrease to an increase with age for normal readers in the precision of phonetic codes. The increase in precision provides a better discrimination of rhyming foils as opposed to greater confusion.

The cross-sectional studies described above tend to support the hypothesis that some poor readers are delayed in their phonetic coding abilities, because poor readers began to show evidence of a phonetic similarity effect at about the ages of 8 or 9 years. However, the results could be attributed to the good and poor readers' different levels of reading achievement. Thus, the poor readers' insensitivity to the phonetic similarity effect could be a consequence of a more limited experience in written language rather than poor reading ability. Much of the most recent research in this area has attempted to avoid this problem by utilizing reading level matching, comparing the performance of children who differ in age but who are matched in reading skill. Reading level matching also allows researchers to distinguish between developmental delay or deficiency in phonetic coding as a cause of reading difficulties in cases where poor readers are found to perform more poorly than same-aged good readers. A finding that older poor readers show a level of phonetic coding that is similar to younger good readers matched in reading level would be consistent with the idea that poor readers are delayed in their development of phonetic coding ability. Conversely, a finding that older poor readers perform more poorly than younger good readers on phonetic coding tasks would be consistent with the idea that poor readers suffer from a deficiency in phonetic coding.

Reading level matching studies. The following studies (Johnston, 1982; Bisanz, Das & Mancini, 1984; Beggs & Howarth, 1985) used reading level matching in order to compare the development of phonetic coding in good and poor readers.

Johnston (1982) used an auditory presentation paradigm similar to Shankweiler et al.,'s (1979) study to examine the development of the phonetic similarity effect in British children. Johnston tested 3 groups of poor readers who were 9, 12 and 14 years of age and matched each group with chronological and reading-level matched normal readers. The children were auditorily presented 14 alternating rhyming and nonrhyming letter strings. Children who were nine years old were presented with strings that were five letters in length, while 12-year-olds and 14-year-olds were presented with six and seven letter strings, respectively. The children were asked to recall the letters either immediately or after a 15-second delay. Recall was measured by having the children write down the letters in the order they were presented.

Johnston found that all the children showed evidence of a phonetic similarity effect in both immediate and delayed conditions. The magnitude of the effect was the same for all subject groups. The only difference between the good and poor readers was that the poor readers' overall recall levels were much poorer than that of the age-matched controls, but they were very similar to the reading-level-matched controls. Johnston concluded that although the poor readers do not have any difficulty with phonetic coding, they did have a developmental delay in short-term memory skills. However, Johnston tempered her conclusions by stressing the need for replication of her results with visually presented stimuli.

Bisanz, Das and Mancini (1984) conducted a study with Canadian children that utilized both grade and reading level matching. The stimuli consisted of letters and the paradigm was similar to Shankweiler et al.'s (1979) visual successive experiment. The children were divided into 6 groups, average readers in Grade 2, poor, average and good readers in Grade 4, and average and poor readers in Grade 6. This combination of groups enabled examination of the phonetic similarity effect developmentally among average Grade 2, 4, and 6 average readers; between different reading levels in Grade 4 and 6; and by matching children at the Grade 2 and 4 reading levels. The children were presented with five rhyming and five nonrhyming six-letter strings. There was a 15-second delay following presentation. Children were told to recall the letters in the order they had seen them. There were two other presentation conditions that were designed to prevent rehearsal of phonetic codes. The first was a partial suppression condition, where the children had to repeat the word "cola" over and over as quickly as possible during the delay interval. The second condition was a total suppression condition where the children were required to say "cola" during stimulus presentation and the delay interval.

The researchers first examined age trends among average readers in Grades 2, 4 and 6. They found that errors decreased with age in all three presentation conditions, although performance was always better in the nonsuppression condition. Memory was always better for nonrhyming letters than for rhyming letters, that is, all children showed evidence of a phonetic similarity effect.

The second analysis compared the same-grade children at different reading levels. Superior, average and poor readers in Grade 4 used phonetic codes differentially, but the differences were smaller than those found for Grade 2 children in the Shankweiler et al. (1979) study. Poor readers showed a minimal phonetic similarity effect and were not affected by the suppression conditions as much as were the average and superior readers. There were no differences in the phonetic similarity effect between average and poor Grade 6 readers.

The third analysis involved two reading-level matched groups: (a) Grade 2 average readers matched with Grade 4 poor readers, and (b) Grade 4 average readers matched with Grade 6 poor readers. The developmental delay hypothesis was expected to be confirmed if the performance of the matched groups was similar. However, in both matched groups, the older poor readers had fewer errors overall (for both rhyming and nonrhyming letters combined) than the younger average readers. The authors stated that their results were not clear enough to reject the developmental delay hypothesis. Although the two groups of average and poor readers scored the same on the standardized reading test, the authors suggest that the strategies that the older poor readers used to achieve their reading scores may have differed from the younger average readers, along with their performance on the experimental task. The poor readers, though matched with the average readers on cognitive measures, were found to have a higher mental age as measured by the intelligence tests. Bisanz and her colleagues speculated that if there are any processes common to memory and mental age, then older poor readers will tend to perform better on the memory task than their younger average reading level counterparts.

Beggs and Howarth (1985) examined phonetic coding in British children aged 7 to 11, comparing (a) normal with above average readers and (b) normal with below average readers. The phonetic coding task was a word reading test using sets of 5 rhyming and nonrhyming words presented visually to the children. Recall was written and the children were asked to remember the serial order. The first experiment was designed to examine the development of phonetic coding in normal and above average readers. The normal readers ranged in age from 8 to 11 years, while the above average readers were reading from 1 to 2 years above their age level (one group reading at the 9 year level and one reading at the 10 year level). In the examination of normal readers, Beggs and Howarth found that the proportion of rhyming errors increased significantly across both the chronological and reading ages tested. The two groups of above average readers did not differ in the proportion of rhyming errors made. The authors also compared 12 matched pairs of normal and above average readers, who had a matched reading age of about 9 years of age. There were no significant differences between the mean proportion of rhyming errors. Beggs and Howarth concluded that normal readers start to develop the ability to code written words phonetically somewhere between the reading ages of 8 and 9, as evidenced from the proportion of rhyming errors on a short-term memory task. This proportion begins to asymptote after 10 years of age. The authors believe that the change in the pattern of errors probably represents the acquisition of a new skill, as evidenced by the analysis in terms of reading ability. Here they found that above average readers conform to the same pattern, reaching an asymptote that was not significantly different from that reached 2 years later by normally developing readers.

The second experiment was designed to compare the development of phonetic coding in children reading at average and below average (at least 1 year) levels for their chronological ages. The authors expected that slow readers would show no evidence of phonetic coding up to a reading age of about 8, and that coding would increase at this point. The procedure was the same as in the first experiment. The slow readers did not show evidence of phonetic coding until they were at an 8-year old reading level. Ten pairs of slow and normal reading children were matched for reading age which averaged at the 8 year level. The researchers found no significant differences between the two groups.

The authors conclude that normal readers improve their phonetic coding skills up until a reading age of about 10 years. Slow readers do not show evidence of phonetic coding until they reach a reading age of about 8. There is a rapid increase in phonetic coding between the reading ages of 8 and 10, after which it stabilizes. Beggs and Howarth suggested that the same phonetic coding skill acquisition process occurs in all beginning readers, but that the skill may be accelerated or delayed in individual cases. They believe that this skill is acquired by all readers over a critical period.

The evidence from the cross-sectional reading level studies discussed above tends to support the hypothesis that poor readers are delayed in their development of phonetic coding skills and appear to "catch up" at around the reading age of 8 or 9. The reading level matching studies were inconclusive in providing support for a developmental delay or deficit hypothesis. Reading-level matching is problematic because equating good and poor readers by reading age does not necessarily mean that they have been exposed to the same experience with written language. Stanovich (1986) states that poor readers are exposed to much less text than same-aged good readers and that this difference can be found very early in the reading acquisition process. Bisanz et al. (1984) and Backman, Mamen and Ferguson (1984) both suggest that adequate tests of the developmental delay hypothesis require longitudinal studies that trace the relationship between the acquisition of component skills and the developmental and individual differences in reading ability.

Longitudinal studies. Despite the problems with cross-sectional and reading level designs and the call for longitudinal studies, there are only a few such studies in the literature. Mann and Liberman (1984) and Share, Jorm, Maclean and Matthews (1984) both examined the development of phonetic coding in Kindergarten children longitudinally, following up the children one year later when they were in Grade 1.

Mann and Liberman (1984), working at Haskins Laboratories, conducted a longitudinal study of the type advocated by Backman et al. (1934) and Bisanz et al. (1984). They were interested in how childrens' performance on phonetic coding in short-term memory and phonemic awareness tasks predicted later reading skills. The authors used a syllable counting task to test for phonemic awareness. Memory for rhyming and nonrhyming word strings was used to test phonetic coding, and the Corsi Blocks were used as a control to test nonverbal memory. The children were retested the following year on the three experimental tasks mentioned above, and were also given standardized word recognition and nonsense word reading tests. In the second testing year, the children were divided into three groups of good, average and poor readers according to their teachers' recommendations. There were 26 good readers, 19 average readers and 17 poor readers. The teacher ratings were corroborated by the children's scores on a standardized reading test. The mean score for the good readers was significantly higher than the average readers, which was in turn significantly higher than the poor readers.

The word string test consisted of 16 auditorily presented 4-word strings, eight of which were rhyming strings and eight of which were nonrhyming strings. The same words appeared as both rhyming and nonrhyming words, thus eliminating the possibility that one of the lists could be harder to remember than the other.

Mann and Liberman found that the good readers showed a phonetic similarity effect on the verbal materials in both Kindergarten and Grade 1, whereas the poor readers did not show a phonetic similarity effect at either grade level. There was no difference in performance between good and poor readers with nonverbal material. The results were not influenced by any floor or ceiling effects, that is, the children performed well below ceiling levels and well above floor levels. Regression analysis examined relationships between total scores on the reading test and the experimental task scores. In Kindergarten testing, error scores for both rhyming and nonrhyming words were each related to reading. The relationships were in a positive direction and were significant. Performance on the syllable counting test was similarly related to reading ability. Taken together, error scores on syllable counting and nonrhyming word memory tasks accounted for 24 percent of the variance in reading scores, and each one uniquely accounted for 9 percent of the variance. The analogous regression on the Grade 1 scores resulted in similar relationships.

Mann and Liberman concluded that phonemic awareness and verbal short-term memory not only correlate with early reading ability, but may presage future reading ability in the first grade. In their study, the children whom they identified as future good readers showed evidence of relying on a phonetic code whereas the future poor readers did not. Mann and Liberman discussed how useful their simple phonemic awareness and phonetic coding tasks could be in identifying extremes of reading success in the first grade.

Share et al. (1984) examined a variety of oral language, motor, personality, home background and early literacy skills in relation to performance of reading in the first grade. They tested 543 Australian children (in Australia formal reading education begins in Kindergarten). Share and his colleagues did not partition their subjects into different reading level groups, but instead formed a single composite reading score from measures of informal word and nonsense word reading and a standardized reading test given at the end of Kindergarten. These tests were repeated at the end of Grade One.

One literacy skill measured the children's susceptibility to the phonetic similarity effect. Children were tested for memory of 12 rhyming and 12 nonrhyming sentences which were presented and recalled orally. Rhyming and nonrhyming sentences correlated significantly with Kindergarten and Grade 1 reading measured at the end of each year. Children who showed a phonetic similarity effect had higher reading scores. Memory for rhyming sentences was the fifth variable to enter in a stepwise multiple regression which in total accounted for 59 percent of the variance of reading. The four other variables which entered prior to sentence memory were phoneme segmentation ability, letter copying ability, sex, and letter naming ability. The ability variables were positively related to reading, while sex appeared to function as a moderator variable indicating differential predictability of males and females.

In summary, Mann and Liberman (1984) and Share et al. (1984) examined the development of phonetic coding and its relationship to reading achievement. These researchers used regression analyses to examine the relationship between the phonetic similarity effect and reading. They found that prereaders who have the ability to code phonetically are more likely to become good readers. An important implication of their research is that phonetic coding can develop before reading instruction begins. This finding contrasts with the findings of the studies which matched reading levels, particularly those of Beggs and Howarth (1985), where phonetic coding was not supposed to appear until the child was reading at an 8-year old level. The remaining issue is whether or not the development of phonetic coding is dependent on reading acquisition. This issue has been dealt with in research with preschool children. These studies with preschoolers have examined whether or not phonetic coding is present before school reading instruction begins.

The Phonetic Similarity Effect Before Reading Acquisition

Preschool children's susceptibility to the phonetic similarity effect has been studied by Conrad (1971), Alegria & Pignot (1979), Brown (1977) and Lean and Arbuckle (1984). The idea that phonetic coding develops along with reading acquisition, and hence would not be found in prereaders, was based on a study that examined early development of phonetic coding. Conrad (1971) showed that nonphonetic strategies appear to be typical of younger children at a prereading stage, and that the frequency of use increased with age.

Conrad (1971) found that British children less than about 5 years mental age (as measured by the English Picture Vocabulary Test) did not show evidence of a phonetic similarity effect; their recall of rhyming pictures was equal to recall of nonrhyming pictures. Children with mental ages of 5 to 6 years (the age when most children begin reading instruction) appeared to make some use of phonetic coding, since their recall of rhyming pictures was somewhat worse than their recall of nonrhyming pictures. The remaining three mental age groups (children with mental ages ranging from 6 to 11), preferred the phonetic strategy, showing a systematically increasing disadvantage in recall of rhyming pictures as mental age increased. Conrad concluded that preschool children typically employ a nonphonetic strategy to hold information in their working memory, and suggested that phonetic coding may not be available as a memory strategy for children until they master reading.

Conrad's (1971) study was influential in formulating hypotheses about the development of phonetic coding abilities; however, there are certain methodological problems in the study. Although Baddeley (1978; 1979) suggested that the phonetic similarity effect is related to retention of ordered information, Conrad's recall task consisted of a visual spatial location task for items. Furthermore, Conrad used pictorial stimuli. Pictures offer more visual or semantic coding options than do words, and hence may not have been the best stimulus choice for tapping the phonetic aspects of working memory. The use of mental age as a grouping factor is also suspect, particularly since the children in the youngest group had mental ages lower than their chronological ages. In addition, Conrad did not know exactly which children were reading, he just assumed those older than six years mental age were reading. His argument would have been strengthened by using chronological ages, determining reading levels, using nonpictorial stimuli and testing for ordered recall.

Subsequent studies have refuted Conrad's claims by showing that preschool children are capable of coding phonetically. Alegria and Pignot (1979) attempted to replicate Conrad's study with 16 Belgian preschoolers aged 3.7 to 5 years. The researchers presented the children with rhyming and nonrhyming pictures using Conrad's recall procedure. They reported that a phonetic similarity effect was present by the age of 4 and the magnitude of the effect increased with age.

Additional evidence for phonetic coding abilities in prereaders was found by Brown (1977). He attempted to eliminate the problems inherent with picture stimuli by presenting letters to his sample of preschool children. He reasoned that the use of letters, as opposed to pictorial stimuli, would more likely predispose preschoolers who knew their letters to code them phonetically.

Brown presented a group of 4 and 5 year old children with 16 sets of upper case letters that varied in phonetic as well as visual similarity. The children labelled each letter during presentation, and serial recall was probed either visually or verbally. Brown suspected that visual probes may bias young children towards visual as opposed to phonetic coding. Brown found a significant phonetic similarity effect for preschool children only when recall was probed verbally.

Lean and Arbuckle (1984) showed that the magnitude of the phonetic similarity effect did not change during the preschool years in a sample of Canadian children. They utilized Shankweiler et al.'s (1979) phonetic similarity effect paradigm, but incorporated several modifications. Letter stimuli were used and item memory and order memory were examined separately. Two groups of children (aged 42 to 59 months and 60 to 68 months) were shown 12 sets of four-letter long rhyming and nonrhyming strings in a sequential presentation. A large phonetic similarity effect was present in both age groups in both item and order recall. Since the children were selected for the experiment only if they knew their letters, it was possible that the failure to find age differences in the phonetic similarity effect was because the younger group was particularly precocious with respect to reading. A follow-up with 27 members of the original younger group 14 months later indicated, however, that the magnitude of the phonetic similarity effect shown by the children had not changed over the interval. Thus, neither cross-sectional nor longitudinal comparisons of the two age groups showed any evidence of a developmental change in the magnitude of the phonetic similarity effect. The authors suggested that preschoolers can use a phonetic code at an early stage and that both item and order memory are sensitive to the use of a phonetic strategy.

In summary, the more recent research using picture and letter stimuli has shown that prereaders as young as 42 months of age are capable of coding phonetically. Thus, phonetic coding may antedate the acquisition of reading skills.

Summary of the Literature

The relationship between phonetic coding in working memory and reading has been examined by several methods. Early studies compared the phonetic similarity effect in good and poor readers within a single age level. The researchers at Haskins found that at the Grade 2 level (ages 7 to 8), good readers showed a phonetic similarity effect while poor readers did not. This finding held for both auditory and visual presentation of letters (Shankweiler et al., 1979), auditory presentation of words and sentences (Mann et al., 1980), and visual presentation of words (Mark et al., 1977). Byrne and Shea (1979) extended these findings to a continuous auditory presentation of words and nonsense words.

Several studies questioned the validity of these results. Hall et al. (1983) suggested that poor readers' relative insensitivity to the phonetic similarity effect may have been overgeneralized and may apply only to academically less competent poor readers and to more difficult tasks. Brown et al. (1983) questioned the influence of presentation modality effects and found small but significant phonetic similarity effects in poor readers when recall for auditorily presented letter strings was tested. However, these two studies grouped together several developmental levels in their analyses. Developmental differences could have contributed to the discrepancy between their findings and those reported earlier.

A second group of studies examined the development of the phonetic similarity effect over the early school years in good and poor readers. One goal of these studies was to determine whether poor readers' failure to show a phonetic similarity effect was due to a developmental delay or a deficit in the ability to use phonetic codes in working memory. Cross-sectional studies, where good and poor readers were matched on age levels, indicated that younger poor readers did not show a phonetic similarity effect until the age of 8 or 9 years (Siegel & Linder, 1984; Siegel & Ryan, 1984; Olson et al., 1984). These findings support the hypothesis that, during the reading acquisition stage, some poor readers are delayed in their ability to code phonetically. Several studies paired poor readers with both age-matched and reading-level-matched good readers (Johnston, 1982; Bisanz et al., 1984; Beggs and Howarth, 1985). The results were inconclusive in supporting a developmental delay or deficit hypothesis. However, the poor readers in Johnston's and Bisanz et al.'s studies were all over 8 years of age and, considering the findings from the cross-sectional studies mentioned above, were more likely to show phonetic similarity effects. Beggs and Howarth used younger poor readers and found that poor readers around the age of 8 or 9 began to use phonetic codes. Methodological problems with reading level matching, (e.g., equating children on reading level does not necessarily mean that they have the same experience with reading) were noted as posing difficulty in interpreting the findings of these studies.

A third set of studies avoided some of the methodological problems inherent in cross-sectional studies of reading development by using longitudinal designs. Two longitudinal studies of the development of phonetic coding and its relationship to reading achievement were discussed. Mann and Liberman (1984) found that children who showed a phonetic similarity effect with auditorily presented words in Kindergarten were more likely to be classified as good readers in the first grade than were readers who had not shown a phonetic similarity effect in Kindergarten. Share et al. (1984) did not use reading groups, but instead used a composite reading score as a dependent variable in regression analyses. They found that memory for rhyming sentences in Kindergarten (where formal reading instruction begins) was a significant predictor of Grade 1 reading ability.

The last group of studies examined the development of phonetic coding at the prereading stage. Conrad (1971) reported that phonetic coding (based on spatial recall for rhyming and nonrhyming pictures) develops at the time when reading instruction is introduced. The validity of his findings was questioned on methodological grounds. Alegria and Pignot (1979) found evidence of phonetic coding using stimuli and recall methods that were similar to Conrad's, but defining prereaders in terms of chronological rather than mental age. The use of pictorial stimuli still posed a problem, so Brown (1977) introduced the use of letter stimuli, reasoning that letters would be more likely to predispose preschoolers who knew their letters to code them phonetically. He found that 4-year-olds showed a phonetic similarity effect under verbally probed conditions. Lean and Arbuckle (1984) examined the phonetic similarity effect with letter stimuli using a temporal recall task. They added a developmental aspect to see whether the magnitude of the phonetic similarity effect changes over the preschool period. They tested children under 5 years of age and over 5 years of age and found a large phonetic similarity effect in both groups. The phonetic similarity effect did not change in magnitude over the preschool period in both cross-sectional and longitudinal comparisons.

A developmental progression for phonetic coding can be abstracted from the review of the literature. Phonetic coding is used as a memory code at the prereading stage for children who will become or are likely to become good readers. It does not seem to be used by future poor readers, at least when it is measured at the Kindergarten stage. Good readers continue to be sensitive to the phonetic similarity effect at least through most of the elementary school years, whereas poor readers appear not to be sensitive to it until they reach the age of 8 or 9 years.

There are several methodological problems that make it difficult to draw any general conclusions either concerning how good and poor readers perform on phonetic similarity tasks, or concerning the relationship between the use of phonetic coding and reading achievement during reading acquisition.

One of the most basic methodological problems is that the studies have varied widely in the level or levels of development that were being assessed. Since Baddeley (1979) stated that phonetic codes are most likely to be used in the early stages of learning to read, it seems appropriate to concentrate research at this developmental level. Nevertheless, as seen in Table 1, the majority of the studies reviewed have concentrated on cross-sectional examinations of children in Grades 2 to 6. There is a paucity of research examining the development of phonetic codes during the early reading acquisition period. The longitudinal research, which did investigate early development of phonetic coding, examined children in Kindergarten and Grade 1 only.

Another reason for focusing on early stages of reading acquisition has been advanced by Stanovich (1986). He stressed the importance of examining the determinants of reading differences at early stages of reading acquisition. He developed the "specificity hypothesis" which states that reading difficulties are at first a specific problem but may begin to have more generalized effects on a broader range of tasks and achievement than reading only. Stanovich suggested that the period of specificity in disabled readers may be quite short and occur at the very earliest stages of reading acquisition. At later stages, disabled readers' more generalized problems may make it difficult to uncover specific correlates of their reading difficulty. Therefore, it seems reasonable to concentrate research effort on determining correlates of reading difficulty at early stages of reading acquisition.

A further methodological problem concerns sample selection, specifically, the identification of good and poor readers. As seen in Table 1, and mentioned previously, many of the studies reviewed used different standardized tests (both reading and IQ tests), and definitions of the reading levels varied widely both within and across studies. For example, an examination of the age and reading score ranges of Johnston's (1982) good readers' showed that some were reading below age level. However, these good readers could be classified as marginal readers according to the classification system used in Shankweiler et al.'s (1979) study. Furthermore, Johnston's younger poor readers were not as disabled as her older ones in terms of delay in reading age. Conversely, Bisanz et al.'s (1984) poor readers were all reading 2 years below Grade level. It is difficult to draw general conclusions about the phonetic coding abilities of "poor" readers at different age groups when the poor readers in one study are essentially reading at the same level as the average readers in another study. Furthermore, matching reading groups on IQ levels or selecting groups that do not differ significantly in mean IQ levels may not be representative of the population that is being tested (see Backman et al., 1984; Torgesen, 1985). One solution to this problem is to use reading scores as a dependent variable in regression analyses with phonetic coding ability as an independent variable, after IQ is partialled out as the first stepped independent variable. This method avoids the problems inherent in partitioning children into "good" and "poor" reading groups, while still controlling for IQ differences.

Statement of Purpose

The present study addresses two issues that emerge from the review of the literature. The first issue is that most of the research in the area has neglected the developmental level where phonetic coding is assumed to be most influential in reading achievement. The research generally has concentrated on older (Grade 2 to Grade 6) children. It therefore seems important to examine differences in phonetic coding ability at early stages of reading acquisition.

The second research issue is that the majority of the researchers have been concerned with whether or not good and poor readers can be differentiated by their ability to code phonetically. The use of dichotomous groups of good and poor readers raises problems regarding group definition. As mentioned in the review of the literature, it is very difficult to control for reading levels. There are no clear-cut or established lines of demarcation between good and poor readers. In addition, the examination of early reading acquisition skills assumes the inclusion of prereading Kindergarten children, which renders the separation of the sample into "good" and "poor" ability groups questionable.

Both of these issues were taken into account in the present study. The phonetic similarity effect was investigated in children who represented a range of reading achievement levels during the early reading acquisition period. A cross-sequential design, shown in Table 2, was used to examine the phonetic similarity effect over two testing sessions separated by a 1 year

Table 2

Cross-Sequential Design

Grade Levels

Cohort	n	Year of Testing 1	2
1	28	K	1
2	31	1	2
3	29	2	3

interval in children at three grade levels or cohorts which correspond to three consecutive levels of reading instruction: Kindergarten to Grade 1, Grade 1 to Grade 2 and Grade 2 to Grade 3. Performance scores on each of three memory tasks were entered into ANOVAs where Cohort was the factor used to measure cross-sectional development and year of testing was the factor used to measure longitudinal development within cohorts. Then the relationship between the phonetic similarity effect and reading achievement was examined. In order to avoid previously discussed methodological problems in reading group assignment, the present study investigated whether phonetic coding was positively related to reading achievement instead of examining whether or not good and poor readers differed in their ability to code phonetically after the effects of IQ had been accounted for. Reading scores for this purpose were used as a dependent variable in hierarchical multiple regression analyses where IQ scores were forced first into the equation. The regressions were performed for each Cohort and for each year of testing. The cross-sequential design allowed the examination of the development of relationships between phonetic coding and reading achievement both concurrently and one year later.

The early reading acquisition period was defined as the primary school grades where formal reading instruction ranges from initial prereading activities to the beginnings of fluency in normal, achieving readers. The children tested were in regular classes and were known to vary in reading achievement, as expected in a normal classroom. Children referred for special services such as reading remediation were excluded so that all subjects were receiving their formal reading instruction in the normal classroom situation. Reading achievement was measured initially as a means of getting a reasonably-sized range of reading or prereading achievement in each of the cohorts. A range of reading levels was examined, instead of simply classifying the children as good or poor readers.

Two hypotheses were advanced. The first hypothesis was that there is a phonetic similarity effect throughout the early reading acquisition period. The second hypothesis was that phonetic coding ability, as measured by the phonetic similarity effect, is positively related to reading achievement during the early reading acquisition period.

Most of the research in this area has used one memory coding task to examine phonological similarity effect differences. In the present study, each child was tested on three memory coding tasks that represented three levels of written expression: letters, words and sentences. The tasks used were adapted from tasks that had produced phonetic similarity effects in good readers at the Grade 2 level in the phonetic coding research. The choice of the tasks was constrained by the fact that prereaders would be tested and therefore the tasks could not require the reading of words. The first coding task was a visually presented letter memory task that was a modified version of the one used by Lean and Arbuckle (1984) which was originally adapted from Shankweiler et al. (1979). One modification was that only order memory was measured. Lean and Arbuckle had found memory for order to be somewhat more sensitive to phonetic similarity effects in young children than was memory for items and, according to Baddeley (1979), the phonetic similarity effect is based on an overload in memory for ordered information. Another modification to the letter task was the addition of visual similarity as an alternative coding choice. In order to increase the number of visually similar letters available, lower case letter stimuli were used. This modification also served to make the task more analagous to reading since connected text consists mainly of lower case letters. The second coding task was a word memory task which was adapted from Byrne and Shea (1979, Experiment 1). This task tapped both phonetic and semantic coding through measures of false recognition of phonetically related and semantically related foils. The third coding task was a memory task for sentences composed of rhyming or of nonrhyming words, adapted from Mann et al.'s (1980) study. Only meaningful sentences were used because Mann and her colleagues had not found any major effects of the meaningfulness manipulation.

In summary, the present study investigated the effect of phonetically similar letter, word and sentence stimuli on children's memory during the early reading acquisition period. Children in Kindergarten to Grade 2 who represented a range of prereading and reading abilities were tested on memory tasks and reading and were retested on the equivalent measures one year later. Two hypotheses were developed. The first was that there is a phonetic similarity effect throughout the early reading acquisition period. The scores on each memory task were entered into an ANOVA where both cross-sectional and longitudinal developmental effects were examined. The second hypothesis was that phonetic coding ability is positively related to reading achievement. The scores for each memory task competed equally for entry after IQ was forced first in a stepwise multiple regression equation to predict reading scores for each cohort concurrently and one year later.

Method

Subjects

Sample Selection

The initial sample of 108 children from Kindergarten, Grade 1 and Grade 2 classes was recruited as follows. In November and early December of the school year, group reading tests were administered to all children in those grade levels in three elementary schools in predominately black, lower class areas of Montreal. Kindergarten and Grade 1 children were given the level of the Metropolitan Achievement Test Reading Survey-Form JS (Balow, Farr, Hogan & Prescott, 1978) that was appropriate for their grade; Preprimer for Kindergarten and Primer for Grade 1. Grade 2 children were given the Reading Comprehension subtest of the Stanford Diagnostic Reading Test, Red Level (Karlsen, Madden & Gardner, 1977). When the raw scores were compared with the standardized norms, it was found that nearly all the children were reading at average or above average levels for their grades, perhaps reflecting the fact that Canadian children tend to do well on American standardized tests (see Backman et al., 1984). Because the American norms were inappropriate for our sample, it was decided to characterize reading levels relative to the performance of all children in these grades in these schools. In order to do so, z-scores were computed for all the children at each grade level across the three schools using the means and standard deviations for that grade. Out of the population tested at each grade level, 12 children were randomly chosen from among those having z-scores at or above 0.50, 12 children were randomly chosen from those having z-scores between 0.49 and -0.49, and 12 children were randomly chosen from those having z-scores at or below -0.50. These children all spoke English as their first language and were not receiving any special educational assistance outside of the classroom. The 108 children selected were then administered the full test battery between the months of February and May of that school year.

Approximately 15 months after the initial reading tests were given, in February and March of the following year, all children in the Grade 1 and 2 classes at the three schools were administered the same reading tests that had been given at those grade levels the previous year, while the Grade 3 classes were administered the Stanford Diagnostic Reading Test Reading Comprehension Subtest Green Level (Karlsen et al., 1977).

<u>Final Sample</u>

Of the initial 108 subjects, 90 were available for individual retesting in the second year. This retesting was carried out between the months of March and June of the school year. Two of these subjects were dropped because they were repeating their grade. The remaining 88 children constituted the final sample and, unless otherwise indicated, all analyses are based on their data only.

Cohort 1 (Kindergarten in the first testing year) consisted of 28 children, 15 males and 13 females. In the first year of testing, they ranged in age from 60 to 78 months with a mean age of 71.9 months. Cohort 2 (Grade 1 in the first testing year) consisted of 31 children, 20 males and 11 females. In the first year of testing they ranged in age from 75 to 99 months, with a mean age of 85.0 months. Cohort 3 (Grade 2 in the first year of testing) consisted of 29 children, 18 males and 11 females. In the first year of testing they ranged in age from 89 to 114 months with a mean age of 100.4 months.

Materials, Procedures and Scoring

Letter Task

The stimuli in this task consisted of 32 sets of four 7.6 x 12.7 cm cards, each card showing a lowercase letter, 5.5 cm in height, of the English alphabet. The composition of the sets are shown in Appendix 1.

There were eight sets for each of the four stimulus conditions: (a) visually similar and phonetically similar (b,d,p,g), (b) visually similar and phonetically dissimilar (m,u,h,r), (c) visually dissimilar and phonetically similar (d,e,t,v), and (d) visually dissimilar and phonetically dissimilar (w,k,l,s). In each set, each letter was allowed to appear only once. Within each stimulus condition, each letter appeared equally often in each serial position.

There was a set of practice stimuli that consisted of three 7.6 x 12.7 cm cards with hand drawn pictures of a star, a fork and a hand. Pictorial practice stimuli were used to minimize potential interference with the subsequent presentation of the experimental letter stimuli.

The first session began with a practice trial that used the pictorial stimuli. Once it was clear that the child understood the task, the first 16 trials (visually similar or visually dissimilar) were given. The remaining 16 trials were given at the following testing session without a practice trial. The trials were presented in alternating blocks of four rhyming and four nonrhyming trials. There were four possible orderings of block presentation, which were counterbalanced across grades and reading levels.

The procedure used for the practice and experimental trials was as follows. The first card of the first set was presented face up on a desk for 3 seconds; then the card was turned over. The remaining cards in the set were presented one at a time in succession in the same manner, each one turned over and placed on top of the preceding card. After a 5 second delay, during which the cards were randomly shuffled, the children were given the cards. They were told to look through the cards and asked to pick out the card that had come first and to place it on the desk. Then the children were asked to pick out the card that had come next, and so on, until they had placed the last card in the last position.

The score for each set was the number of letters correctly placed. A correct placement was a letter placed in the proper serial position. The score for each set could thus range from 0 to 4. Since there were eight sets per stimulus condition, the total possible correct for each stimulus condition was 32.

Word Task

The word list for this task was derived from the list used by Byrne and Shea (1979) and is presented in Appendix 2. The word list consisted of 12 groups of related words. Each group consisted of five words, one of which was always the first one of the set to occur in the list. This antecedent word will be referred to as the target. The remaining four words were respectively: (a) a word that was semantically related to the target (semantic foil); (b) a word that was unrelated to the target or the semantic foil and which was arbitrarily designated as a semantic control word (semantic control); (c) a word that was phonetically related to the target (rhyme foil); and (d) a word that was unrelated to the target or the rhyme foil and was arbitrarily designated as a phonetic control word (rhyme control). One half of the semantic foils were antonyms of the target and the remaining half were synonyms of the target. The rhyme foils were not orthographically similar to the targets in order to avoid any possible confounding of orthographic interference with phonetic interference in the case of subjects who could spell. Within the list the target word of each set was presented twice, the foils and control words each were presented once. Thus each of the 12 sets took up 6 list positions, for a total of 72 positions in the list. The whole list contained 121 words, the remaining words being filler words which were not related to the other words in the list. Ten filler words appeared once, 15 appeared twice and 3 appeared three times. Filler words predominated in the first part of the list, occupying Positions 1 to 5, 7, 10 to 13, 15, 16, 18 and 20. The remaining 34 filler words were scattered throughout the remaining 101 positions of the list. Actual placement of experimental and filler words was determined so that the word sequence would fulfill the following conditions.

In the list, each target appeared twice, separated by 7 positions. Each foil and its control word appeared only once and were separated by one position. The foils preceded the control words in one half of the cases; this arrangement being reversed for the remaining half. Six semantic and six rhyming sets of foils and control words were placed nine positions after the second appearance of the relevant target. The remaining six sets were placed 13 positions after the second appearance of the target.

The words chosen for the word list were mostly from the list used by Byrne and Shea (1979). The remaining words replaced certain words in Byrne and Shea's list that appeared inappropriate because they fit into both rhyming and semantic foil categories. These words (and the filler words, which were not reported in Byrne and Shea's article) were chosen from lists in Felzen and Anisfeld (1970) and Freund and Johnson (1972) who also used false recognition paradigms with children the same age as those used in Byrne and Shea's study.

The list was recorded by the experimenter and played back on a cassette tape recorder through headphones. The list was recorded at the rate of one word every 5.5 seconds. An alerting signal was presented 1 second before each word to ensure that the subjects attended to it. A seven word practice trial was produced in the same manner, using words that were unrelated to the experimental list.

The children were told that they were to listen carefully to a long list of words, and that they would hear some of the words more than once. After they heard each word, they were to tell the experimenter "new" if they were hearing the word for the first time and "old" if they had heard the word previously in the list. The children were given a practice trial which was repeated until they had performed perfectly. Each child heard the list through the earphones to minimize outside distractions. The experimenter also had earphones to know where the child was in the list, but she could hear the child's response directly.

The number of false positives (saying "old" at the occurrence of a new word) was recorded for the semantic and rhyme foils and their corresponding control words. The number of false negatives (saying "new" for the occurrence of the second target and the repeats of the filler words) was also recorded.

Sentence Task

Materials and procedures for the sentence task were adapted from Mann, Liberman and Shankweiler (1980). The sentences used are shown in Appendix 3. Pilot testing on Kindergarten and Grade 1 children of the 13-word sentences used by Mann et al. produced marked floor effects, suggesting that sentences of that length were too difficult for the younger children. Therefore, the 13-word sentences used by Mann et al. were given only to the children in Grades 2 and 3 while the children in Kindergarten and Grade 1 were given 9-word sentences which were shortened versions of the 13-word ones. For each sentence length, seven were phonetically confusable (rhyming) and seven phonetically nonconfusable (nonrhyming) sentences. Syntactic structure was the same for both types of sentences. Phonetically confusable sentences contained from seven to nine rhyming words in the 13-word length sentences and from five to seven rhyming words in the 9-word length sentences. Phonetically nonconfusable sentences contained no rhyming words.

The 14 sentences were arranged in a random sequence with the restriction that no more than two phonetically confusable or phonetically nonconfusable sentences appeared in sequence.

The experimental sentences were recorded and played back on a cassette tape recorder through headphones. The children were first given a practice trial using the same procedure as would be used for the experimental trials. This trial consisted of one rhyming and one nonrhyming sentence that was the same length as the sentences the child received in the experimental trials. On both practice and experimental trials, the children were asked to listen very carefully to the sentences that they were going to hear. They were told that some of the sentences were going to sound very funny and some of them were going to sound like regular sentences. They were asked to tell the experimenter as much of the sentence as they could remember, just like they had heard. The practice trials were repeated until the child understood the directions. Each sentence was preceded by the prompt "ready" to ensure that the child was attending.

The sentence t.sk was scored in the following way. A score of zero was given for correct repetition of the sentence. Each omission, substitution and reversal made on each sentence constituted one error. When a child did not recall any of the words in the sentence, a score of 13 was given for 13-word sentences and 9 was given for 9-word sentences.

Intelligence Testing

In the first year of testing, the children who were less than 72 months of age were administered the Vocabulary and Block Design subtests of the Wechsler Preschool and Primary Scales of Intelligence (WPPSI) (Wechsler, 1967). Children 72 months of age or older were given the same subtests of the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974). Full scale IQ scores were calculated from Sattler's (1982) estimation tables. Although the WPPSI and the WISC-R are different tests, the format and demands of the subtests used and the normative samples are similar, the factor loading is the same, and the subtests in each test correlate similarly with Full Scale IQ scores.

General Procedure and Design

Each child was administered the experimental tasks individually over two testing sessions. The child was brought from the classroom into a quiet separate room and seated at a desk or table. The experimenter chatted with the child in order to establish rapport and to put the child at ease.

All children were tested on all the measures that were given in each year, with the tasks being administered in two different orders, counterbalanced across children within grade and reading levels. In Year 1 there were six task components administered over two sessions, each session about 30 minutes in duration. Half of the children in each grade and reading level were given the following sequence of tasks in the first experimental session: (a) Visually Similar Letter Task, (b) Vocabulary Subtest and (c) Word Task. In the second experimental session, held approximately 2 days later, these children were given the remaining tasks in the following order: (d) Visually Dissimilar Letter Task, (e) Block Design Subtest and (f) Sentence Task. The remaining children in each grade had the reverse order, with Tasks (f), (e) and (d) being presented in that order in the first session and Tasks (c), (b) and (a) in that order in the second session. For the second year of testing, the order was exactly the same as in the first year for each child, except that the intelligence tests, (tasks (b) and (e) above), were not administered.

Three experimenters worked at all schools for Year 1 of testing. In Year 2, one experimenter was not available, but the remaining two experimenters worked at all three schools.

Results

The results are presented in two sections. The first section examines the evidence relevant to the major hypothesis, that is, that there is a phonetic similarity effect throughout the reading acquisition period. The three memory tasks represented three levels of written expression, letters, words and sentences. Each task was analyzed separately within the cross-sequential design, which assessed both cross-sectional developmental differences among three cohorts, representing three successive age and grade levels and longitudinal developmental differences within each Cohort over a one-year period. The analyses were performed with ANOVAs. The second section of the results examines the secondary hypothesis, that is, that phonetic coding ability is related to concurrent and future reading achievement. After IQ scores were forced in first, the ratios of rhyming error scores over total error scores for each of the memory tasks were entered into hierarchical multiple regressions. The regressions were performed separately for each Cohort and each year of testing in order to enable the inclusion of all the data from the Sentence Task. Relationships between sensitivity to the phonetic similarity effect and reading achievement with the effects of IQ partialled out were thus determined both concurrently and predictively.

The Phonetic Similarity Effect During Reading Acquisition

Each of the three memory tasks was analyzed with an ANOVA. For the cross-sectional age comparison, cohort was used as a between subjects factor, while for the longitudinal comparison, Year of Testing was used as a within subjects factor. Originally, it was planned to do a third possible set of comparisons, between cohorts who were the same age at testing, for example, (a) the two Grade 1 cohorts, that is, Cohort 2 at Year 1 and Cohort 1 at Year 2, and (b) the two Grade 2 cohorts, that is, Cohort 3 at Year 1 and Cohort 2 at Year 2. When the data had all been collected, however, it was found that the comparisons would be of doubtful validity since problems with beginning the testing schedule in the second year led to a 15 month average interval between tests and hence second year samples of Grade 1 and Grade 2 readers that were somewhat older on average than the corresponding first year samples had been. For that reason, the "same age" comparisons were not made.

Letter Task

A 3 X 2 X 2 X 2 split-plot ANOVA was performed on the total number of correct reconstructions for the eight sets in each stimuli condition. The between-subjects factor was Cohort (1 -- Kindergarten-Grade 1; 2 -- Grade 1-Grade 2; 3 -- Grade 2-Grade 3). The three within factors were (a) Year of Testing (Year 1, Year 2), (b) Phonetic Similarity (rhyming, nonrhyming) and (c) Visual Similarity (high, low). The ANOVA summary table for this analysis appears in Appendix 4. The means and standard deviations for the task are given in Table 3.

Total number of letters correctly reconstructed differed by cohort, $F(2, 85) = 7.03$, $p < .01$. Post hoc testing (all post hoc testing reported utilized Scheffe´s method) showed that Cohort 1 reconstructed fewer letters in the correct order than Cohort 2 at the .05 level of significance, and Cohort 3 at the .01 level of significance. Cohorts 2 and 3 did not differ significantly. Year of Testing also had an effect, with children reconstructing more letters in the correct order in the second year of testing than in the first year of testing, $F(1, 85) = 46.88$, $p < .001$.

Table 3

Means and Standard Deviations for Correct Reconstructions on Letter Task in Relation to Cohort, Year of Testing and Phonetic Similarity (Maximum = 32)

Cohort	n	Year (Grade) of Testing	Phonetic Similarity: Rhyming M	Rhyming SD	Nonrhyming M	Nonrhyming SD	Mean
1	28	1(K)	12.5	4.66	14.6	6.18	13.7
		2(1)	14.3	4.52	18.4	6.07	16.4
2	31	1(1)	14.0	4.73	18.7	5.36	16.4
		2(2)	15.7	4.81	21.7	7.13	18.7
3	29	1(2)	14.8	4.11	16.0	5.26	15.4
		2(3)	17.7	5.45	23.4	6.83	20.6

The main interest in the data lay in the results of the phonetic similarity manipulation. As expected, the overall effect of Phonetic Similarity was significant. Children recalled the order of more nonrhyming than rhyming letters, $F(1, 85) = 111.08$, $p < .001$. Phonetic similarity also entered into three significant two-way interactions, Phonetic Similarity X Cohort, $F(2, 85) = 3.32$, $p < .05$, Phonetic Similarity X Year of Testing, $F(1, 85) = 17.30$, $p < .001$, and Phonetic Similarity X Visual Similarity, $F(1, 85) = 8.27$, $p < .01$. Post hoc tests on each of these three interactions showed that none of them qualified the main effect of phonetic similarity. Order recall of nonrhyming letters was better than order recall of rhyming letters for each of the three cohorts, at each year of testing and for both visually similar and visually dissimilar letters (all p values $< .01$). As can be seen from Figure 1, the Phonetic Similarity X Cohort interaction reflected a linear upward trend in recall of rhyming letters and a nonlinear trend in recall of nonrhyming letters across successive cohorts. Specifically, as the post hoc tests showed, order recall of rhyming letters improved systematically across cohorts with Cohort 1 reconstructing significantly ($p < .01$) fewer letters correctly than Cohort 3. In contrast, order recall of nonrhyming letters improved significantly ($p < .01$) from Cohort 1 to Cohort 2 but then showed a nonsignificant decline to Cohort 3.

The Phonetic Similarity X Year of Testing interaction is seen in Figure 2. The interaction is attributable to differential improvement in order recall of rhyming versus nonrhyming letters

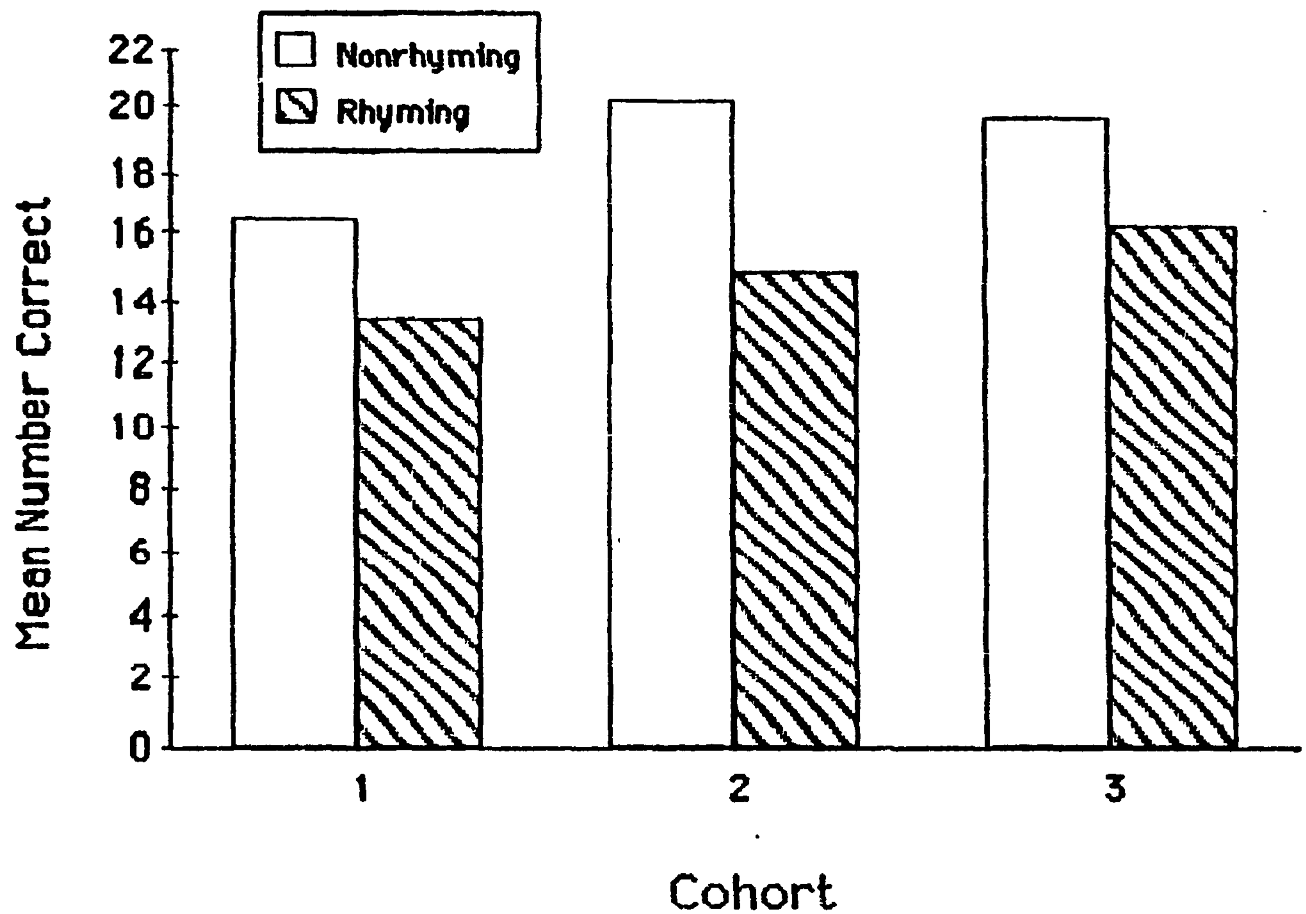

Figure 1. Mean Number of Correct Reconstructions of Letter Strings as a Function of Phonetic Similarity and Cohort

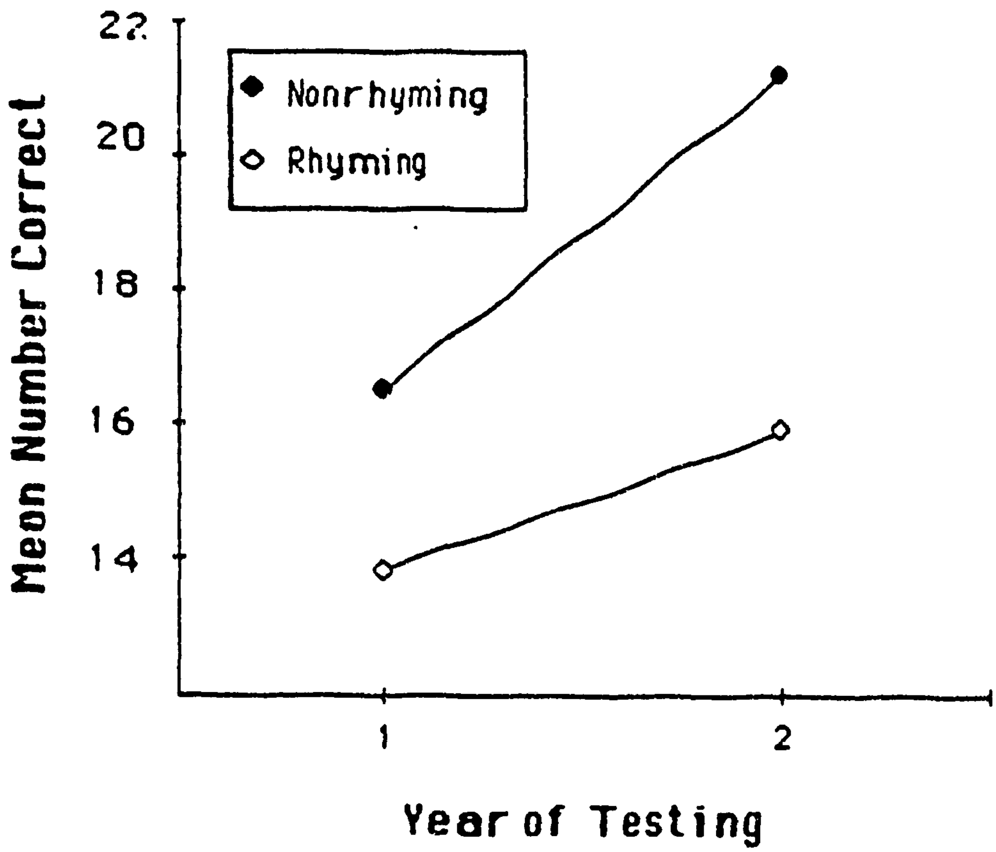

Figure 2. Mean Number of Correct Reconstructions of Letter Strings as a Function of Phonetic Smilarity and Year of Testing

across the two years of testing. Figure 2 shows that order recall for nonrhyming letters improved over the two testing times more than order recall for rhyming letters, although both pairs of means differed at the accepted alpha level of $p < .05$.

The interaction between Phonetic and Visual Similarity is shown in Figure 3. Post hoc testing showed that the interaction was attributable to the differential effect of visual similarity on rhyming and nonrhyming letters. Specifically, visual similarity did not affect order recall for nonrhyming letters, $p > .10$, but it did affect order recall for rhyming letters. Order recall for these letters was lower under visually similar conditions than visually dissimilar conditions, $p < .05$.

The interaction between Cohort and Year of testing approached significance, $F(2, 85) = 3.08$, $p < .06$. Closer examination of the data show that the near-interaction is mainly due to the performance of Cohort 3, and in particular, to what appears to be an abnormally poor performance of Cohort 3 when they were in Grade 2 (Year 1 of testing). Table 3 shows that this poor performance was primarily with respect to recall of the order for nonrhyming letters.

Word Task

This memory task was analyzed in two ways, each by a 3 X 2 X 2 ANOVA. The between-subjects factor was Cohort and the two within-subjects factors were (a) Year of Testing (Year 1, Year 2), and (b) Foil Type (semantically similar, phonologically similar and control foils). In the first analysis, the raw number of false positives for the semantic, rhyming and control foils was

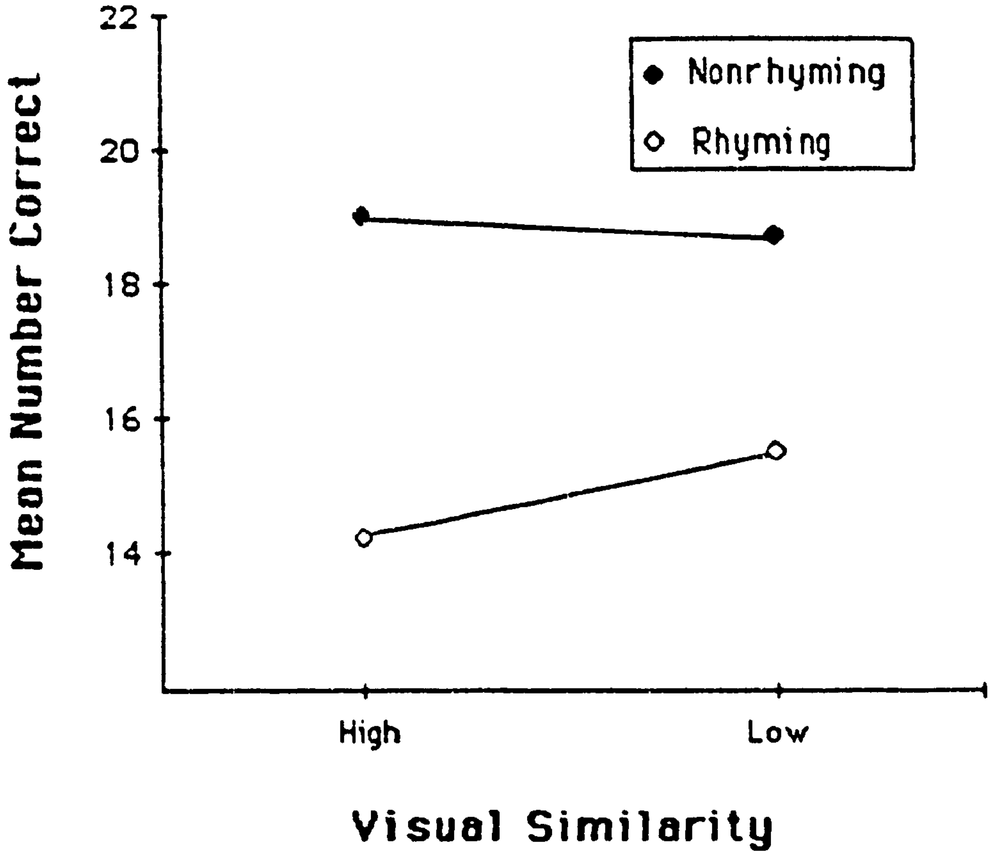

Figure 3. Mean Number of Correct Reconstructions of Letter Strings as a Function of Phonetic Similarity and Visual Similarity

analyzed. For the second analysis, the results were transformed into *d'* scores. The *d'* measure combines the ˈit rate on the target words with the false alarm rate (false positives) to a particular type of foil, thus taking into account more of the data and minimizing the potential effects of response biases. The higher the *d'* score the better the child is at discriminating between the target and the foil. Low *d'* scores indicate greater confusion between target and foil. In the case of semantic and rhyme foils respectively, higher false positives and lower *d'* scores for these conditions relative to control foils suggests that the child is encoding information on these dimensions. Low *d'* scores for rhyme foils in comparison to controls is thus another version of the phonetic similarity effect, while low *d'* scores for semantic foils in comparison to controls is suggestive of an analagous semantic similarity effect. The *d'* scores were computed using the HR - FAR equation and were calculated using Hochhaus' (1972) tables. Hit rate (HR) was calculated by recording the number of times the children recognized the second presentation of the target word. When all 12 targets were recognized, the HR proportion was transformed from 1.00 to .99 in order to use the Hochhaus tables. The false alarm rate (FAR) was calculated separately for the semantic, rhyming and control foils by recording the number of times each foil was falsely recognized. When there were no false positives recorded, the FAR proportion was transformed from 0 to .01 in order to use the Hochhaus tables. The pattern of results for the analyses using the raw false positive scores and the *d'* scores was very similar, and hence only the d' measures are reported. The ANOVA summary table for this analysis appears in Appendix 5. The means and standard deviations for the task are given in Table 4.

The overall effect of foil type was significant, $F(2, 85) = 36.3$, $p < .001$. Post hoc tests showed that the mean d' scores for discriminability of targets from control foils were significantly higher than that of rhyming foils, which in turn were significantly higher than that of semantic foils, all p values $< .01$. The overall effect of Year of Testing was also significant, $F(1, 85) = 4.03$, $p < .05$, with the d' scores being higher in the second year of testing. The relative order of the three conditions of foil type was found at each year of testing; however, there was a significant Foil Type by Year of Testing interaction, $F(2, 170) = 10.65$, $p < .001$. As seen in Figure 4, post hoc tests showed that the interaction was attributable to the fact that the children's ability to discriminate rhyme foils from targets did not improve from Year 1 to Year 2, $p > .10$, whereas their ability to discriminate semantic foils and control foils from targets did improve significantly, $p < .05$. In the first year of testing, the d' scores for the rhyme condition were significantly higher than those for the semantic condition, $p < .01$, and not significantly lower than the control condition, $p > .10$, while in the second year scores for the rhyme condition were significantly lower than the control condition, $p < .01$, and not significantly better than the semantic condition, $p > .10$.

In general, the results showed that in comparison to the control condition, there was no evidence of poorer ability to

Table 4

Means and Standard Deviations for d′ Scores on Word Task in Relation to Cohort, Year of Testing and Type of Foil

Cohort	n	Year (Grade) of Testing	Type of Foil: Rhyming M	Rhyming SD	Semantic M	Semantic SD	Control M	Control SD
1	28	1(K)	2.3	1.4	2.0	1.0	2.5	1.0
		2(1)	2.6	1.2	2.6	1.3	2.9	1.1
2	31	1(1)	2.7	0.9	2.2	0.9	2.8	1.0
		2(2)	2.6	1.0	2.5	1.0	3.1	1.0
3	29	1(2)	2.9	0.9	2.4	0.9	2.9	0.8
		2(3)	2.7	0.8	2.8	0.7	3.1	0.7

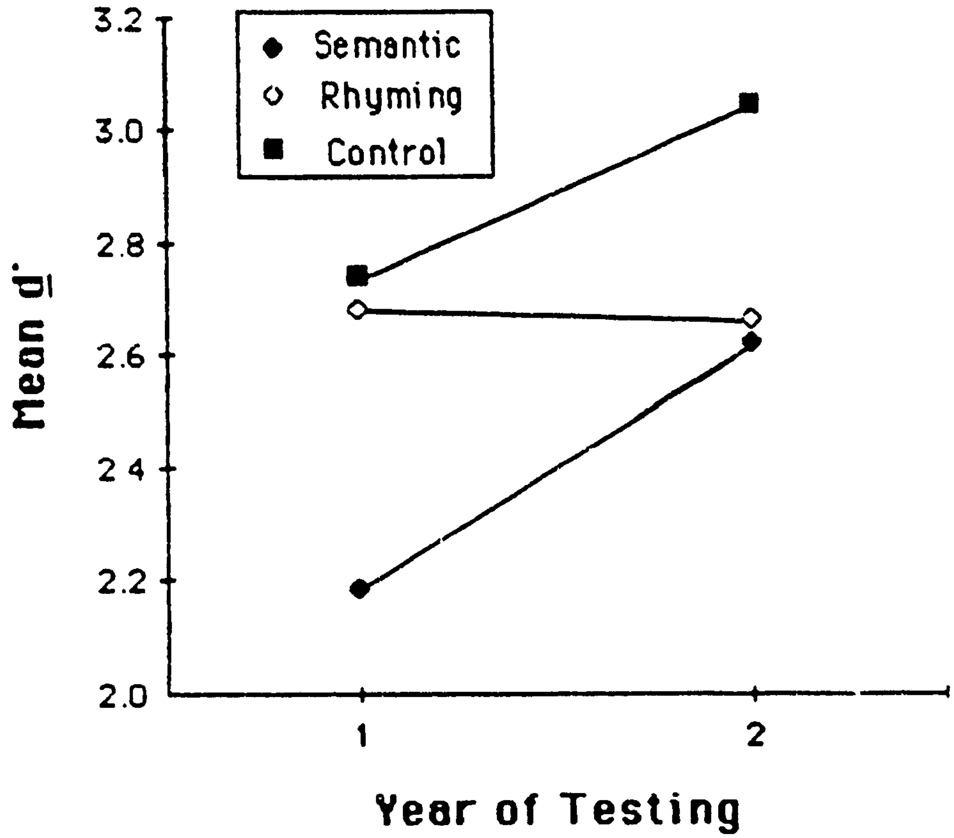

Figure 4. Mean d' for Word Recognition as a Function of Word Type and Year of Testing

discrimination between rhyming foils and the targets in the first year, but there was evidence in the second year that rhyme foils were more difficult to discriminate than were control foils. However, in comparison to the control condition, the children were subject to semantic interference in both the first and second years. There were no significant cohort differences in this task, although, as seen in Table 4, d' scores for control, semantic and rhyming conditions did tend to increase over successive cohorts.

Sentence Task

The means and standard deviations for this task are given in Table 5. Preliminary examination of the data for the sentence task showed that the data for the two sentence lengths could not meaningfully be analyzed in the same analysis. Therefore, separate ANOVAs were performed for Cohorts 1 and 3 on this test. Cohort 1 had been tested twice on 9 word sentences and Cohort 3 had been tested twice on 13 word sentences. The data from Cohort 2 were not analyzed in this manner because for that Cohort the sentence length had changed from Year 1 to Year 2.

Two 2 X 2 ANOVAs were performed. For both analyses, the two within-subjects factors were (a) Year of Testing (Year 1, Year 2), and (b) Phonetic Similarity (rhyming or nonrhyming sentences). The ANOVA summary table for the Cohort 1 analysis appears in Appendix 6 and the summary table for the Cohort 3 analysis appears in Appendix 7. For Cohort 1, the effect of phonetic similarity was significant, $F(1, 27) = 291.34$, $p < .001$. As expected, children recalled more nonrhyming sentences than rhyming sentences. For Cohort 3, the effect of Phonetic Similarity was

Table 5

Means and Standard Deviations of Recall Errors on Sentence Task in Relation to Cohort, Year of Testing and Sentence Type

Cohort	n	Year (Grade) of Testing	Sentence Length (in words)	Sentence Type: Rhyming M	Sentence Type: Rhyming SD	Sentence Type: Nonrhyming M	Sentence Type: Nonrhyming SD
1	28	1(K)	9	29.1	8.26	14.3	8.29
		2(1)	9	26.1	9.17	11.4	8.56
2	31	1(1)	9	24.5	8.91	9.6	7.44
		2(2)	13	42.7	16.2	27.1	11.34
3	29	1(2)	13	45.0	19.12	35.2	18.06
		2(3)	13	35.0	17.82	29.2	15.18

Maximum for 9-word sentences = 63

Maximum for 13-word sentences = 97

also significant, $F(1, 28) = 26.03$, $p < .001$. The effect of Year of Testing was also significant, $F(1, 28) = 9.18$, $p < .01$, with children making fewer errors overall in the second year of testing (Grade 3) than in the first year of testing (Grade 2). Phonetic Similarity did not enter into any significant interactions with Cohort or Year of Testing in either of the analyses. Although the data from Cohort 2 were not entered into the ANOVA, there still was a large phonetic similarity effect during both years of testing, as seen in Table 5. The data of Cohort 2 for each year of testing were analyzed by t-tests. Children made significantly more errors on rhyming sentences than nonrhyming sentences both in the first year of testing, $t(30) = 13.9$, $p < .001$, and in the second year of testing, $t(30) = 8.84$, $p < .001$.

Summary of the Developmental Findings

The results partially supported the major hypothesis, in that a phonetic similarity effect was found to be present throughout the period tested in the letter and sentence tasks. However, the cross-sectional and longitudinal developmental effects varied with the tasks used to measure the effect.

The letter task appeared to be the most sensitive to developmental changes, showing a phonetic similarity effect with all cohorts, at both years of testing, and for both visually similar and visually dissimilar letters. Although Phonetic Similarity entered into interactions with Cohort, Year of Testing and Visual Similarity, none of the interactions qualified the phonetic similarity effect. Therefore, the presence of the phonetic similarity effect was established at the Kindergarten level, and remained at the Grade 3 level. In the sentence task, there was evidence of a phonetic similarity effect in both cohorts analyzed by the ANOVA (1 and 3), and this effect was not modified in either cohort by interactions with Year of Testing. Cohort 2 also showed a significant phonetic similarity effect at each year of testing. In the word task, children showed a greater sensitivity to semantic interference, that is, more false recognitions of semantic foils than either rhyme or control foils, but only during the first year of testing. Sensitivity to semantic interference decreased over the two years of testing, as indicated by the increase in semantic $\underline{d'}$ scores, but no changes were found in sensitivity to rhyming interference. In the second year of testing, sensitivity to rhyming and semantic interference did not differ, and both produced lower $\underline{d'}$ scores than did control foils.

The main finding of the developmental analyses was that phonetic similarity adversely affected memory in the letter and sentence tasks in all cohorts in both years of testing. Phonetic and semantic similarity affected memory in the word task in comparison to control foils, although the effect was significant only in the second year of testing. Phonetic similarity produced less interference in the word task than did semantic similarity, although the difference was significant only in the first year of testing.

Phonetic Coding in Relation to Reading Achievement

This section examines the relations between phonetic coding ability and concurrent and future reading achievement in each cohort with the effects of IQ partialled out. The raw data for the letter, word and sentence tasks were transformed into a ratio score in order to have a single score for each task that would be comparable across tasks. The score used was Conrad's (1979) Internal Speech Ratio (ISR), which is the ratio of the number of errors on phonetically similar stimuli to the total number of errors. Conrad used the ISR as a measure of phonetic coding. If all errors were made on phonologically similar stimuli, the ISR would equal 1.0. If all errors were on phonologically dissimilar stimuli, the ISR would equal 0. Higher ISR scores indicate more phonetic coding is occurring in the memory task.

The ISR for each year of testing was calculated in the following manner. For the letter task, total correct responses for each subject under each condition were changed to error scores by subtracting the number correct from the total possible correct (32). Since the ANOVAs on the letter task had shown that the visual similarity manipulation did not independently affect serial reconstruction, the error scores for each of the visually similar and visually dissimilar rhyming letters were combined. Thus, two scores based on phonetic similarity were calculated: (a) a phonologically similar score (visually similar and visually dissimilar rhyming letters) and (b) a phonologically dissimilar score (visually similar and visually dissimilar nonrhyming letters). The phonologically similar error score was divided by the total error score to produce the letter task ISR. The ISR for the word task was based on the false positive raw scores for rhyming foils and their control foils. The ISR was computed by dividing the number of false positive responses to rhyming foils by the total of the number of false positive responses for the rhyming foils and their control foils. Because there were several cases where the number of false positive responses was 0, .01 was added to each false positive score used in constructing the ratio. The ISR for the sentence task was computed by dividing the number of errors made on rhyming sentences by the total number of errors made on the rhyming and nonrhyming sentences.

Three hierarchical multiple regression analyses were performed on the results from each cohort in order to examine three possible relationships between phonetic coding ability and reading z-scores. In each regression, IQ entered at the first stage, and the other three ISR scores were entered together at the second stage and competed with one another. This procedure is equivalent to treating IQ as a covariate, where reading achievement can be predicted from the ISR scores of the memory tasks after holding constant individual differences in IQ (Tabachnick & Fidell, 1983). It is noted that the alternate procedure of letting IQ compete equally with the other variables was done with standard multiple regressions. The results of the two types of analyses were similar; however, the effect was not as clear with the standard regression and it appeared from the comparison of the outputs that IQ was acting as a suppressor variable.

The first set of regressions examined first year memory and reading relationships, where the independent variables were IQ scores and the ISRs from each first year memory task and the dependent variable was the first year reading z-scores. The second set of regressions examined how first year memory scores predicted second year reading scores, where the independent variables were IQ scores and the ISRs from each first year memory task and the dependent variable was the second year reading z-scores. The third set of regressions examined second year memory and reading relationships, where the independent variables were IQ scores and the ISRs from the second year memory tasks and the dependent variable was the second year reading z-scores. The intercorrelations for these regressions are shown in Appendix 8. Table 6 displays the means and standard deviations of the three ISR scores for each cohort and each year of testing. There was some evidence of phonetic coding (Mean ISR > .5) in all the memory tasks for all cohorts at both years of testing, except for the word task in Cohort 3 at Year 1 of testing.

Tables 7, 8 and 9 display the beta weights, $\underline{R}^2$, the increase in $\underline{R}^2$ and the $\underline{F}$ to enter values for each cohort and for each type of regression analysis. Generally, the results showed that after the effects of IQ had been accounted for, the letter task accounted for the major portion of the variance of reading scores. However, these results depended on the cohort and year in which both the memory tasks and reading tests were measured.

Table 7 shows the results of the regression analyses for the concurrent relationships of IQ and the Year 1 phonetic coding ISR

Table 6

Means and Standard Deviations of ISR Scores in Relation to Cohort, Year of Testing and Task

Cohort	n	Year (Grade) of Testing	Task: Letter M	Letter SD	Word M	Word SD	Sentence M	Sentence SD
1	28	1(K)	.54	.06	.66	.28	.69	.10
		2(1)	.58	.10	.61	.24	.73	.11
2	31	1(1)	.58	.08	.50	.31	.75	.11
		2(2)	.64	.14	.75	.27	.61	.06
3	29	1(2)	.52	.07	.44	.32	.57	.09
		2(3)	.66	.15	.60	.35	.54	.11

Table 7

Hierarchical Multiple Regression of IQ, First Year Letter, Word and Sentence IRS's Predicting First Year Reading Scores

Cohort	n	Variable	Beta	R^2	Increase in R^2	F to Enter Value
1	28	IQ	.11	.02	.02	.47
		Letter	.40	.16	.14	4.31 *
		Word	.33	.28	.12	3.94 *
		Sentence	.09	.29	.01	.26
2	31	IQ	.26	.07	.07	2.07
		Letter	.43	.27	.20	7.78 **
		Sentence	.08	.30	.03	1.06
		Word	.17	.30	.00	.22
3	29	IQ	.10	.02	.02	.42
		Sentence	.11	.03	.01	.33
		Letter	.00	.04	.01	.33
		Word	.13	.04	.00	.00

* $p < .05$

** $p < .01$

Table 8

Hierarchical Multiple Regression of IQ, First Year Letter, Word and Sentence ISR's Predicting Second Year Reading Scores

Cohort	n	Variable	Beta	R^2	Increase in R^2	F to Enter Value
1	28	IQ	.44	.19	.19	6.10 *
		Letter	.39	.23	.14	5.15 *
		Word	.24	.39	.06	2.47
		Sentence	.07	.40	.00	.18
2	31	IQ	.29	.06	.06	1.80
		Letter	.69	.46	.40	20.68 ***
		Sentence	-.02	.46	.00	.13
		Word	-.05	.46	.00	.01
3	29	IQ	.40	.13	.13	4.15
		Letter	-.30	.23	.09	3.09
		Word	.05	.23	.00	.08
		Sentence	.01	.23	.00	.01

* $p < .05$

*** $p < .001$

Table 9

Hierarchical Multiple Regression of IQ, Second Year Letter, Word, and Sentence ISR's Predicting Second Year Reading Scores

Cohort	n	Variable	Beta	R^2	Increase in R^2	F to Enter Value
1	28	IQ	.42	.19	.19	6.10 *
		Letter	.23	.30	.11	4.02 *
		Word	-.30	.39	.08	3.28 *
		Sentence	.05	.39	.00	.08
2	31	IQ	.28	.06	.06	1.80
		Sentence	.10	.10	.04	1.11
		Letter	-.09	.10	.01	.23
		Word	.17	.11	.01	.24
3	29	IQ	.45	.13	.13	4.15
		Word	.13	.16	.03	.85
		Letter	-.24	.17	.01	.25
		Sentence	-.11	.18	.01	.34

* $p < .05$

scores with Year 1 reading scores. For each Cohort, IQ was forced to enter first, but in none of the three cases did it account for a significant percentage of the variance in the Year 1 reading scores. Once the variance accounted for by IQ had been removed, the three ISR scores were forced to compete for entry as potential predictors. For Cohort 1 the letter task entered first, accounting for an additional 14% of the variance in Kindergarten reading scores, $F(2, 25) = 4.31$, $p < .05$. The word task entered next, accounting for a further 12% of the variance, $F(3, 24) = 3.94$, $p < .05$, while the sentence task entered last and contributed negligibly to the explained variance. The multiple R of .54 was not statistically significant, $p > .08$. For Cohort 2, the letter task entered first among the ISR scores, accounting for a further 20% of the variance in the Grade 1 reading scores, $F(2, 28) = 7.78$, $p < .01$, with the sentence and word tasks entering second and third and accounting for a negligible percentage of the variance. The multiple R of .55 was statistically significant ($p < .05$). For Cohort 3, none of the three ISR scores accounted for a significant percentage of the variance in the Grade 2 reading scores with the multiple R being only .20, $p > .90$.

Table 8 shows the results of the multiple regressions for the predictive relationships, IQ plus Year 1 ISR scores as predictors of Year 2 reading scores. The IQ measure was again forced to enter first and, for Cohort 1 only, it was found to be predictive of a significant 19% of the variance in Year 2 (Grade 1) reading scores, $F(1, 26) = 6.10$, $p < .05$. With IQ partialled out, the three ISR scores were then allowed to compete for entry. For all three cohorts the letter task entered first, explaining an additional 14% of the variance in Year 2 reading scores of Cohort 1, $F(2, 25) = 5.15$, $p < .05$, and an additional 40% of the variance in the Year 2 reading scores of Cohort 2, $F(2, 28) = 20.68$, $p < .001$, but contributing insignificantly to the explained variance in the Year 2 reading scores of Cohort 3. Neither the word nor the sentence task contributed significantly to the explained variance of any of the three cohorts. The multiple *R*s were statistically significant in the case of Cohort 1, $R = .63$, $p < .05$, and Cohort 2, $R = .68$, $p < .01$, but not in the case of Cohort 3, $R = .48$, $p > .10$.

Table 9 shows the results of the multiple regressions for the second year concurrent relationships with the IQ scores measured at Year 1 and the second year ISR scores as predictors and second year reading scores as the dependent measure. In these regressions, the variance of the second year reading scores that is explained by IQ is that already described above and in Table 8. For Cohort 1, Year 2 letter task scores accounted for a further 11% of the variance of Grade 1 reading scores, $F(2, 25) = 4.02$, $p < .05$, and the addition of the word task explained 8% more of the variance, $F(3, 24) = 3.28$, $p < .05$. The sentence task did not significantly add to the explained variance. For Cohorts 2 and 3, where the children were in Grades 2 and 3, respectively, none of the Year 2 phonetic coding tasks significantly accounted for the variance in Year 2 reading scores. The multiple *R* for Cohort 1 was .62, $p < .05$, while the multiple Rs for Cohorts 2 (.33) and 3 (.43) were not statistically significant.

Discussion

The purpose of this research was to examine the development of phonetic coding during the early reading acquisition period and the relation between phonetic coding ability and reading acheivement. Children in regular Kindergarten to Grade 2 classes who represented a broad range of reading achievement were tested twice at an interval of approximately one year. Phonetic coding ability was assessed by performance on three tasks that had previously been shown to produce phonetic similarity effects in groups of normal readers. The three tasks represented three levels of written expression: letters, words and sentences. The second hypothesis was that phonetic coding ability, as measured by the phonetic similarity effect, is related to concurrent and future reading achievement.

The Phonetic Similarity Effect During Reading Acquisition

The first hypothesis was that there is evidence of a phonetic similarity effect throughout the observed period of reading development. There was strong evidence of a phonetic similarity effect throughout the developmental period tested on the letter and sentence tasks. The effect was not as strong for the word task in the first year of testing.

Letter Task

There was a large and significant phonetic similarity effect when memory for letters was tested. Although the magnitude of the effect varied with cohort, year of testing and visual similarity, none of these interactions reversed the phonetic similarity effect. In particular, the developmental factors of cohort and year of testing affected order recall for nonrhyming letters more than order recall for rhyming letters.

The interactions of Phonetic Similarity with Cohort (Figure 1) and Year of Testing (Figure 2) both show developmental increments in overall performance on both types of materials, with the interactions coming from the nonrhyming conditions. Figure 1 shows cross-sectional development and Figure 2 shows longitudinal development. This development is shown in another way in Figure 5. In Figure 5 the three factors of age, cohort and year of testing are all represented in order to display a kind of overall developmental pattern. For rhyming letters there was regular development of recall of order and a relative similarity of performance of the different cohorts when tested at similar ages. Nonrhyming letters were consistently remembered better, but the upward trend in performance over age and time was less regular, with the greatest deviance being associated with Cohort 3 at Year 1 of testing. The phonetic similarity effect is present from the prereading stage (Kindergarten) to a level where reading skill is more established (Grade 3). Perhaps the most important finding is the presence of the effect at the Kindergarten stage, that is, before formal reading instruction begins. Thus, the results from the letter task support the hypothesis that there is evidence of a phonetic similarity effect throughout the observed period of reading development.

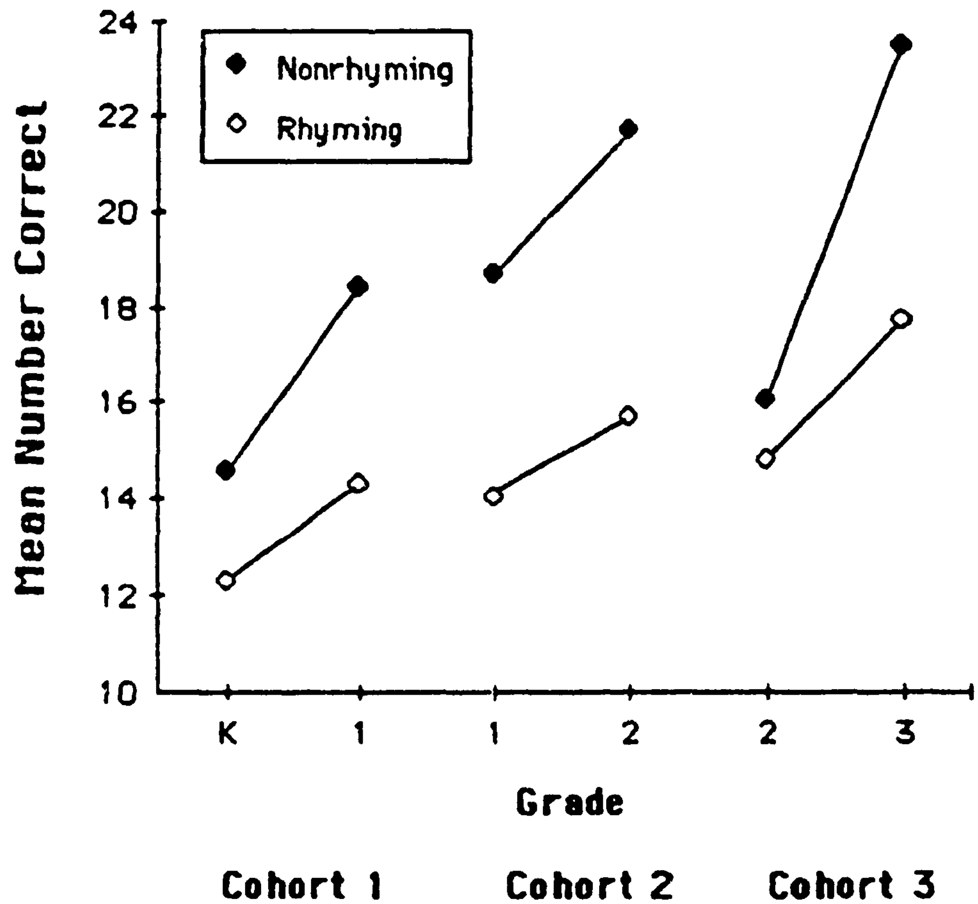

Figure 5. Mean Number of Correct Reconstructions of Letter Strings as a Function of Phonetic Similarity, Cohort and Grade Level

The interaction between Phonetic Similarity and Visual Similarity, as seen in Figure 3, shows that visual similarity affected recall only for rhyming letters. The fact that there was no main effect of visual similarity or that it did not affect recall of nonrhyming letters suggests that these children were not coding letters visually but only phonetically, and that the similar visual properties of the letters interfered with order memory only when working memory was already overburdened with phonetically similar stimuli. The absence of a visual similarity effect in the letter task is inconsistent with some views of early reading. Theorists such as Goodman and Goodman (1979) suggest that reading is a "psycholinguistic guessing game" and that children should learn how to read by guessing what the word is from the shape of the letters and the context. In contrast, Liberman (1984) states that comprehension of written or spoken language involves dealing in linguistic ways with the units of language, that is, the phonemes, syllables and words. The results of the present study support Liberman's position in that children did not show evidence of coding the letters purely by their visual properties.

Word Task

As seen in Figure 4, in the first year of testing, there was no significant difference between the children's ability to discriminate rhyming or control foils from the targets. However, in the second year of testing, rhyme foils were significantly more difficult to discriminate than control foils. Thus, there was evidence of phonetic coding only in the second year of testing.

Semantic foils were harder to discriminate than control foils in both years of testing. There were no cohort differences in recognition of any foil type. However, as seen in Table 5, the childrens' ability to discriminate foils from targets increased across cohorts.

Thus, the results from the word task offer less support for the hypothesis that there is evidence of a phonetic similarity effect than do the letter and sentence tasks. The finding that there was evidence of semantic confusion in both years of testing suggests that the children were primarily using semantic codes. Byrne (1986) recently wrote that continuous recognition tasks may not be as sensitive to phonetic coding as he believed when the original experiment was performed. Byrne mentioned that the reason many poor readers make very few recognition errors based on phonetic properties of speech may reside in the low reliability of the continuous recognition task. Another possible explanation of why the phonetic similarity effect was less marked for the word task was that it was the only one that used stimuli that could easily be coded in units. The stimuli for the other tasks consisted of sequences of items, letters for the letter task and words for the sentence task. Since Baddeley's (1978; 1979) model implies that the phonetic code is particularly useful for retaining sequential information, the nonsequential nature of the word task could have reduced the tendency to code phonetically. The words used were all familiar to the children and therefore could predispose them to code them as a single semantic unit rather than as a sequence of sounds. At early stages of reading acquisition, children asked to read single words that they already know are likely to code them semantically because they already have the meaning of the word stored in their internal lexicon, (see, e.g., Ehri & Wilce, 1985; Gough & Tumner, 1986; Perfetti, 1985; Stanovich, 1986; Tunmer, 1986). Thus, it is likely that in a task where children hear single words, they will code them in the same manner.

Another reason for the low reliability and sensitivity of this task in terms of producing a phonetic similarity effect may be that a simple measure of false recognitions is not sensitive enough to the process of phonetic coding. A possible solution to this problem would be the use of reaction time measures which would record the amount of time it took to decide whether or not they had heard the word before. It seems reasonable to assume that children would take longer to decide if a word was heard before if they were experiencing interference from the use of a phonetic or semantic code.

A final reason why phonetic coding was less evident in the present study than in that of Byrne and Shea (1979) could be the changes that were made in the word list used. Some of the words in the original list of Byrne and Shea belonged to both the rhyming and semantic categories. Specifically, two of Byrne and Shea's target words rhymed ("slow" and "go", which had, respectively, "toe" and "row" as rhyming foils). One of the semantic foils for a target also rhymed with these words ("low" as the semantic foil for the target "high"). Thus, there were four rhyming foils for one particular target, where there should only have been one. These extra rhyming foils may well have inflated the rhyming confusion found in Byrne and Shea's study. In the present study, there was only one rhyming foil for each target.

Sentence Task

In this task, a significant phonetic similarity effect was found in all cohorts. All children tested found it more difficult to recall rhyming sentences than nonrhyming sentences. In Cohorts 1 and 3, where constant sentence length for the two years of testing made it possible to do two-way analyses, phonetic similarity did not interact with year of testing for either cohort. There is thus evidence that the phonetic similarity is present in sentence recall from Kindergarten to Grade 2 and that it remains stable over a one year period, at least for children tested initially in Kindergarten and Grade 2. However, it is difficult to accept or reject the major hypothesis because full analyses were not performed. In order to see how the sentence task would be affected by cohort or age effects, a sentence length must be developed that is within the memory span of the children tested. One possible method of accomplishing this would be to give children two or three different sentence lengths and compare the results in order to find the length that has the fewest floor and ceiling effects.

Summary of the Developmental Findings

There was a very strong phonetic similarity effect in all cohorts at all times of testing in both the letter and sentence tasks. In the word task, phonetic coding was evident only in the second year of testing.

The Kindergarten results are generally consistent with the findings in the literature on prereaders. There was evidence of a phonetic similarity effect in these prereaders. Alegria and Pignot, (1979), Brown, (1977) and Lean and Arbuckle (1984) also found that a majority of prereaders were sensitive to the phonetic similarity effect. Furthermore, the Kindergarten results also compare favourably with the longitudinal research. The present study found that the Kindergarten children's phonetic similarity effect continued over one year of testing. Both Mann & Liberman (1984) and Share et al. (1984) found evidence of a phonetic similarity effect in Kindergarten children which was stable over one year of development.

In summary, the combined results support the major hypothesis that there is evidence of a phonetic similarity effect which appears to be relatively stable throughout the period tested, particularly on tasks requiring recall of sequences of information.

Phonetic Coding in Relation to Reading Achievement

The relationship between phonetic coding and reading achievement after the effects of IQ had been accounted for was examined with three multiple regression analyses. IQ was the first step entered into the hierarchical multiple regression analysis. Ratios that measured the degree of the phonetic similarity effect of the three memory tasks competed equally for the next steps to predict both concurrent and future reading achievement for each cohort. Generally, the letter task was related to concurrent and future reading achievement in Cohorts 1 and 2 after the effects of IQ had been accounted for. Thus, children in these cohorts who were most sensitive to the phonetic similarity effect as measured by the letter task were most likely to score better on reading tests in both years. Therefore, the Haskins group's fundamental position has been supported within cetain cohorts, without floor or ceiling effects.

Cohort 1's performance on the letter and word tasks when they were in Kindergarten, uniquely accounted for 14 and 12 percent, respectively, of the variance of Kindergarten prereading skills, shown in Table 7. Table 8 shows that when Kindergarten phonetic coding performance was compared with Grade 1 reading skills, only the letter task accounted for 14 percent of the variance of reading. When the Cohort 1 children were in Grade 1, as seen in Table 9, their performance on the letter and word tasks accounted for 11 and 8 percent respectively, of the variance of Grade 1 reading. This finding suggests that at the first year of testing, children in Kindergarten are coding phonetically, and that individual differences in this ability are related to pre- and early reading achievement, but only for phonetic memory for letter and word stimuli. This finding parallels the development of reading skills. Children at this age are very involved with letter knowledge presented visually and words presented orally.

Cohort 2's phonetic coding ability measured when they were in Grade 1 uniquely accounted for 20 percent of the variance in Grade 1 reading achievement (as seen in Table 7) and 40 percent of the variance of Grade 2 reading achievement (as seen in Table 8). This finding shows that the letter task became more important in predicting reading achievement at later levels of reading acquisition.

What is interesting is that none of the phonetic coding tasks predicted reading scores when phonetic coding and reading achievement were measured concurrently in Grade 2 in Cohorts 2 and 3 and when Cohort 3 was in Grade 3. This finding suggests that by Grade 2, the children's ability to code phonetically was not very strongly related to reading achievement. These results contrast with the findings of most of the studies in the literature that showed that good and poor readers could be differentiated by their ability to phonetically code stimuli in working memory. However, it is difficult to make a direct comparison because these studies that used children in Grade 2 and higher grades tested for word recognition or word attack skills, whereas the present study tested for reading comprehension in Grades 1 to 3. Furthermore, the phonetic coding tasks were not exactly the same. Most importantly, few of the studies in the literature examined the relationship between phonetic coding ability and reading achievement by multiple regression analyses with the effects of IQ statistically controlled. The lack of association between phonetic coding ability and reading achievement in the higher grades supports Stanovich's (1986) specificity hypothesis which states that at later ages, it becomes more difficult to uncover specific correlates of reading difficulty because older disabled readers have more generalized academic and cognitive problems.

The development of the relationship between the children's phonetic coding ability and reading achievement found in the present study concur with the activities occurring during the normal development of reading acquisition (e.g., Chall, 1983; Smith, 1983). Children in Kindergarten learn about letters, gain insight into the nature of words, and realize that words consist of syllables and phonemes which can be analyzed and synthesized. These abilities were measured in the Kindergarten prereading achievement test used in the present study. The relationship between phonetic coding ability as measured by the letter task was significantly related to prereading achievement in the present study. The next stage of normal reading development which usually occurs during Grade 1 involves "breaking the code", that is, learning decoding skills. Children realize "what the letters are for" and "how to know that a bun is not a bug (Chall, 1983, p. 16)". In this period of reading development, children gain insight about the nature of the spelling system. In the present study, the relationship between phonetic coding ability and reading achievement was the largest and most significant at this period. This stage of reading acquisition may thus be the best time for determining children's ability to code phonetically. The next stage of normal reading development occurs during Grade 2 and 3. Children at this stage consolidate what has previously been learned. They learn to use their decoding knowledge, the redundancies in the language and the stories that they read. The present study did not find any relationship between phonetic coding ability and reading achievement at this stage. In a regular class sample, phonetic coding ability was not strongly related to reading achievement measured concurrently or one year later. [illegible] be that the phonetic coding ability is only related to reading achievement concurrently or one year later during the early stages of reading acquisition, when children are learning about decoding. This speculation is in agreement with Baddeley's (1978, 1979) original hypotheses that phonetic coding is important in the early stages of learning how to read.

General Conclusions

This study has provided support for the two hypotheses that were developed. Firstly, there was evidence of a phonetic similarity effect throughout the early reading acquisition period. It is a strong effect, and is evident in tasks that represent three levels of written expression (letters, words and sentences). The word task's lesser sensitivity to the phonetic similarity effect in the first year of testing was attributed to its inadequacy in tapping working memory. Secondly, phonetic coding ability as measured by the letter task, and, to a lesser extent, the word task, was found to be positively related to reading achievement independent of IQ in Kindergarten, Grade 1 and predictively in Grade 2. The present findings, in combination with those already reported in the literature, provide converging evidence that the phonetic similarity effect is related to reading achievement.

Future Directions for Research

The present study and previous research have found that phonetic coding is positively related to concurrent and one year later reading achievement in the early reading acquisition period. A next step in this research area would be to determine whether or not phonetic coding, as measured by the phonetic similarity effect, is predictive of later reading achievement, and especially to determine whether it could help to identify children who will experience later difficulties in reading. Stanovich (1986) discussed the importance of examining the determinants of later reading difficulties at an early stage. Unfortunately, all that is known about later developmental periods is that the relationship between phonetic coding ability and reading achievement decreases. Cross-sectional studies indicate that poor readers generally do "catch up" to good readers in phonetic coding abilities in later periods of development. However, these poor readers remain poor readers, even if they can code phonetically. The present study suggests that phonetic coding ability in Kindergarten and Grade 1 is a moderately successful predictor of reading ability one year later, but no study has examined more long range predictors nor whether early phonetic coding ability can reliably discriminate the subset of children who will experience major difficulties in reading acquisition. If the phonetic similarity effect could be shown to be a reliable predictor of later reading achievement, then it could be utilized as a simple and easily administered screening tool when children enter Grade 1.

A projected research programme would begin with testing prereaders for phonetic coding ability. These tests would be repeated at least yearly and continued for a five or six year period. Then, regression analyses for the entire sample of children would be performed with phonetic coding scores in the first year of testing as the main independent variable and each year's reading scores as the dependent variable. These results would help to determine the relationship between early phonetic coding skills and later reading achievement. The next step would be to determine when the children begin to code phonetically by their age and levels of reading achievement. The children would be placed into two groups of phonetic coders and noncoders according to their performance on memory tasks in the first testing session. If it was found that noncoders in Kindergarten eventually show a phonetic similarity effect but do not improve in reading achievement, then one could conclude that phonetic coding ability measured in prereaders is a strong and reliable predictor of reading achievement. Then, the next goal of the research programme would be to determine whether noncoders could ever become good readers. Noncoders could be divided into two groups where one group received intensive instruction in reading that encompassed language skills, (similar in structure to what many children with language-related reading problems are given) and one group a normal reading instruction programme. If the children who receive the intensive reading instruction programme perform better on reading tests than children who receive normal reading instruction, then it would be meaningful to detect poor coders at an early stage. These children could be identified as having a possible reading disability before they begin to learn how to read. Remediation and intensive instruction could be started much earlier than usual, and perhaps some children might actually avoid the failure experience of having reading difficulties.

References

Alegria, J., & Pignot, E. (1979). Genetic aspects of verbal mediation in memory. Child Development, 50, 235-238.

Backman, J. E., Mamen, M., & Ferguson, H. B. (1984). Reading level design: Conceptual and methodological issues in reading research. Psychological Bulletin, 96, 560-568.

Baddeley, A. D. (1976). The psychology of memory. New York: Basic.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. Psychological Review, 85, 139-152.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), Processing of visible language (Vol. 1) (pp. 355-370). New York: Plenum.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), The psychology of learning and motivation (Vol. 8). New York: Academic Press.

Balow, I. H., Farr, R., Hogan, T. P., & Prescott, G. A. (1978). Metropolitan Achievement Tests. New York: Harcourt Brace Jovanovich.

Beggs, W. D. A., & Howarth, P. N. (1985). Inner speech as a learned skill. Journal of Experimental Child Psychology, 39, 396-411.

Belmont, I., & Belmont, L. (1978). Stability or change in reading achievement over time: Developmental and educational implications. Journal of Learning Disabilities, 11, 80-88.

Belmont, I., & Belmont, L. (1980). Is the slow learner in the classroom learning disabled? Journal of Learning Disabilities, 13, 32-35.

Bisanz, G. L., Das, J. P., & Mancini, G. (1984). Children's memory for phonemically confusable and nonconfusable letters: Changes with age and reading ability. Child Development, 55, 1845-1854.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read. Nature, 301, 419-421.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, 35, 345-367.

Brown, R. M. (1977). Visual and phonetic memory in preschool children. Perceptual and Motor Skills, 45, 1043-1050.

Brown, R. M., Sanocki, T., & Schrot, D. (1983). Phonetic coding in marginally competent readers. Journal of General Psychology, 109, 87-94.

Byrne, B. (1986, March). Learning to read the first few items: Evidence of a nonanalytic acquisition procedure in adults and children. Paper presented at the Conference on Early Reading, Centre for Cognitive Science, University of Texas at Austin, Austin, TX.

Byrne, B., & Shea, P. (1979). Semantic and phonetic memory codes in beginning readers. Memory & Cognition, 7, 333-338.

Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading: Kindergarten through twelfth grade. Journal of Educational Psychology, 64, 293-298.

Chall, J. S. (1979). The great debate: Ten years later, with a modest proposal for reading stages. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1) (pp. 29-55). Hillsdale, NJ: Lawrence Erlbaum.

Chall, J. S. (1983). Stages of reading development. New York: McGraw-Hill.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Conrad, R. (1971). The chronology of the development of covert speech in children. Developmental Psychology, 5, 398-405.

Conrad, R. (1979). The deaf schoolchild: Language and cognitive function. London: Harper and Row.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusions, and memory span. British Journal of Psychology, 55, 429-432.

Dunn, L. M. (1965). Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Markwardt, L. M. (1970). Peabody Individual Achievement Test. Circle Pines, MN: American Guidance Service.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word-learning visual or phonetic? Reading Research Quartely, 20, 163-179.

Felzen, E., & Anisfeld, M. (1970). Semantic and phonetic relations in the false recognition of words by third- and sixth-grade children. Developmental Psychology, 3, 163-168.

Fox, B., & Routh, D. K. (1983). Reading disability, phonemic analysis and dysphonetic spelling: A followup study. Journal of Clinical Child Psychology, 12, 28-32.

Fox, B. & Routh, D. K. (1984). Phonemic analysis and synthesis as word attack skills: Revisited. Journal of Educational Psychology, 76, 1059-1064.

Freund, J. S., & Johnson, J. W. (1972). Changes in memory attribute dominance as a function of age. Journal of Educational Psychology, 63, 386-389.

Goodman, K. S., & Goodman, Y. M. (1979). Learning to read is natural. In L.B. Resnick & P.A. Weaver (Eds.), Theory and practice of early reading (Vol. 1) (pp. 137-154). Hillsdale, NJ: Lawrence Erlbaum.

Gough, P. B., & Tumner, W. E. (1986). Decoding, reading and reading disability. Remedial and Special Education, 7, 6-10.

Hall, J. W., Ewing, A., Tinzmann, M. B., & Wilson, K. P. (1981). Phonetic coding in dyslexics and normal readers. Bulletin of Psychonomic Society, 17, 177-178.

Hall, J. W., Wilson, K. P., Humphreys, M. S., Tinzmann, M. B., & Bowyer, P. M. (1983). Phonemic similarity effects in good versus poor readers. Memory & Cognition, 11, 520-527.

Helfgott, J. (1976). Phonemic segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. Contemporary Educational Psychology, 1, 157-189.

Hochhaus, L. (1972). A table for the calculation of d' and B'. Psychological Bulletin, 77, 375-376.

Jastak, J., Bijou, S. W., & Jastak, S. R. (1965). Wide Range Achievement Test. Wilmington, DE: Guidance Associates.

Johnston, R. S. (1982). Phonological coding in dyslexic readers. British Journal of Psychology, 73, 455-460.

Karlsen, B., Madden, R., & Gardner, E. F. (1977). Stanford Diagnostic Reading Test. New York: Harcourt Brace Jovanovich.

Lean, D. S., & Arbuckle, T. Y. (1984). Phonological coding in prereaders. Journal of Educational Psychology, 76, 1282-1290.

Liberman, I. Y. (1984). A language-oriented view of reading and its disabilities. Thalamus, 4, 1-50.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY conferences (pp. 207-225). Hillsdale, NJ: Larwrence Erlbaum.

Mann, V. A., & Liberman, I. Y. (1984). Phonologial awareness and verbal short-term memory, Journal of Learning Disabilities, 17, 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mark, L. S., Shankweiler, D., & Liberman, I. Y. (1977). Phonetic recoding and reading difficulty in beginning readers. Memory and Cognition, 5, 623-629.

McCleod, I. (1965). The GAP Reading Comprehension Test. London: Heinemann.

McNinch, G., & Richmond, M. (1972). Auditory perceptual tasks as predictors of first grade reading success. Perceptual and Motor Skills, 35, 7-13.

Neale, M. D. (1966). Neale Analysis of Reading Ability. London: Macmillan.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers. Journal of Experimental Child Psychology, 37, 187-206.

Perfetti, C. A. (1985). Reading ability. New York: Oxford Press.

Rosner, J., & Simon, D. P. (1971). The auditory analysis test: An initial report. Journal of Learning Disabilities, 4, 40-48.

Sattler, J. M. (1974). Assessment of children's intelligence. Philadelphia: Saunders.

Schonell, F. E. (1963). Reading and spelling tests. London: Oliver & Boyd.

Shankweiler, D., & Liberman I. Y. (1976). Exploring the relations between reading and speech. In R. M. Knights & D. J. Bakker (Eds.), Neuropsychology of learning disorders: Theoretical approaches (pp. 321-387). Baltimore: University Park Press.

Shankweiler, D., Liberman, I. Y., & Mark, L. S. (1982). Phonetic coding in dyslexics and normal readers, by Hall, Ewing, Tinzmann and Wilson: A reply. Bulletin of the Psychonomic Society, 19, 78-79.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. _Journal of Experimental Psychology: Human Learning and Memory_, _5_, 531-545.

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. _Journal of Educational Psychology_, _76_, 1309-1324.

Siegel, L. S., & Linder, B. A. (1984). Short-term memory processes in children with learning disabilities. _Developmental Psychology_, _20_, 200-207.

Siegel, L. S., & Ryan, E. B. (1984). Reading disability as a language disorder. Unpublished manuscript.

Slosson, R. L. (1963). _Slosson Intelligence Test for Children and Adults_. New York: Slosson Educational Publications.

Smith, F. (1973). _Psycholinguistics and reading_. New York: Holt, Rinehart & Winston.

Stanovich, K. E. (1986, March). _Speculations on the causes and consequences of individual differences in early reading acquisition_. Paper presented at the Conference on Early Reading, Centre for Cognitive Science, University of Texas at Austin, Austin, TX.

Tabachnick, B. G., & Fidell, L. S. (1983). _Using multivariate statistics_. New York: Harper & Row.

Torgesen, J. K. (1978-1979). Performance of reading disabled children on serial memory tasks: A selective review of recent research. _Reading Research Quarterly_, _14_, 56-87.

Torgesen, J. K. (1985). Memory processes in reading disabled children. Journal of Learning Disabilities, 18, 350-357.

Treiman, R., & Baron, J. (1981, April). Phonemic analysis training with prereaders. Paper presented at the meeting of the Society for Research in Child Development, Boston, MA.

Torneus, M. (1984). Phonological awareness and reading: A chicken and egg problem? Journal of Educational Psychology, 76, 1346-1358.

Tumner, W. E. (1986, March). Cognitive and linguistic factors in learning to read. Paper presented at the Conference on Early Reading, Centre for Cognitive Science, University of Texas at Austin, Austin, TX.

Vellutino, F. (1979). Dyslexia: Theory and research. Cambridge: M.I.T. Press.

Waters, G. S., Komoda, M. K., & Arbuckle, T. Y. (1985). The effects of concurrent tasks on reading: Implications for phonological recoding. Journal of Memory and Language, 24, 27-45.

Wechsler, D. (1967). Wechsler Preschool and Primary Scale of Intelligence. New York: Psychological Corporation.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children - Revised. New York: Psychological Corporation.

Woodcock, R. W. (1973). Woodcock Reading Mastery Test. Circle Pines, MN: American Guidance Service.

Woodcock, R. W., & Johnson, M. B. (1977). Woodcock-Johnson Psycho-educational Battery. New York: Teaching Resources.

Zifcak, M. (1981). Phonological awareness and reading acquisition. *Contemporary Educational Psychology*, *6*, 117-126.

APPENDICES

APPENDIX 1

Stimuli for Letter Task

High Visual Similarity

Rhyming		Nonrhyming	
Block 1	Block 2	Block 1	Block 2
b d p g	d g p b	m u h r	u r h m
d g b p	g b d p	u r m h	r m u h
p b g d	p d b g	h m r u	h u m r
g p d b	b p g d	r h u m	m h r u

Low Visual Similarity

Rhyming		Nonrhyming	
Block 1	Block 2	Block 1	Block 2
v t d e	t e d v	k l s w	l w s k
t e v d	e v t d	l w k s	w k l s
d v e t	d t v e	s k w l	s l k w
e d t v	v d e t	w s l k	k s w l

APPENDIX 2

Stimuli for Word Task

	Target	Foils			
		Rhyming	Control	Semantic	Control
1.	here	ear	hang	there	next
2.	city	pretty	zebra	town	life
3.	high	my	her	low	first
4.	home	comb	glove	house	ship
5.	white	right	fire	black	round
6.	bed	said	cat	sleep	soup
7.	new	two	work	old	boot
8.	bird	word	wet	robin	nose
9.	sweet	heat	big	sour	plump
10.	wood	could	bank	tree	rest
11.	take	ache	help	give	hill
12.	thief	beef	ripe	robber	apple

Fillers

mud *	window *	milk *	room *	stove *	truck *	corner
laugh	shelf *	swim	feed *	nurse	run *	pen
soft *	tag *	chair *	loud **	dime *	jump	
king	learn	coat **	wish *	space *	bought	

* presented 2 times
** presented 3 times

APPENDIX 3

Stimuli for Sentence Task

(in order of presentation)

Nine-word sentences

1. Don't roar any more at the store's four doors. **
2. Tuesdays at three, Lucy is free to see T.V. **
3. Kim saw the big doll in the old barrel.
4. Plain Jane remained in Spain when the rain came. **
5. Mondays at four, Johnny is able to play baseball.
6. Tom and Bill piled books on the chair.
7. Kate ate a steak and a plate of cake. **
8. Lou threw the blue shoe in the new canoe. **
9. Sam drank a coke and a glass of punch.
10. Jack and Mac stacked the sacks on the track. **
11. Poor Jim played inside when the snow was falling.
12. Peg's brown dog bit the bone on the floor.
13. Pat's bad cat bat at the rat on the mat. **
14. Don't play all the time at the teacher's chair.

** Rhyming Sentences

Appendix 3 (Cont'd)

Thirteen-word sentences

1. Don't roar any more at the store's door or Miss Moore will get sore. **
2. Tuesdays at three, Lucy is free to see T.V. with Dee and Lee. **
3. Kim saw that the big doll in the old barrel belonged to her.
4. Plain Jane remained in Spain when the rain washed out the main lane. **
5. Mondays at four, Johnny is able to play baseball with Mike and Bert.
6. Tom and Bill piled books on the chair in front of the door.
7. Kate ate a steak and a plate of date cake that Jake baked. **
8. Lou knew that the blue shoe in the new canoe belonged to you. **
9. Sam drank a coke and a glass of fruit punch that Joan made.
10. Jack and Mac stacked sacks on the track in the back of the shack. **
11. Poor Jim played inside when the snow covered up the back yard.
12. Peg's brown dog bit at the bone that fell on the clean floor.
13. Pat's bad cat bat at the rat that sat on the flat mat. **
14. Don't play all the time at the teacher's chair or Miss Smith will get mad.

** Rhyming Sentences

APPENDIX 4

ANOVA Summary Table for Developmental Analysis of Letter Task

Source	MS	DF	F
Between Subjects			
Cohort	612.84	2	7.03 **
Error	87.15	85	
Within Subjects			
Year of Testing (YrT)	2,045.47	1	46.88 ***
YrT x Cohort	134.23	2	3.08
Error	43.64	85	
Visual Similarity (VS)	50.40	1	2.08
VS x Cohort	6.57	2	.27
Error	24.67	85	
Phonetic Similarity (PhS)	2,758.89	1	111.08 ***
PhS x Cohort	82.34	2	3.32 *
Error	24.84	85	
YrT x VS	9.93	1	.56
YrT x VS x Cohort	44.46	2	2.50
Error	17.82	85	
YrT x PhS	290.85	1	17.30 ***
YrT x PhS x Cohort	43.38	2	2.58
Error	16.81	85	
VS x PhS	107.64	1	8.27 **
VS x PhS x Cohort	23.95	2	1.84
Error	13.02	85	
YrT x VS x PhS	1.25	1	.08
YrT x VS x PhS x Cohort	13.02	2	.87
Error	14.97	85	

*** $p < .001$
** $p < .01$
* $p < .05$

APPENDIX 5

ANOVA Summary Table for Developmental Analysis of Word Task

Source	MS	DF	F
Between Subjects			
Cohort	4.64	2	1.45
Error	3.19	85	
Within Subjects			
Year of Testing (YrT)	7.75	1	4.03 *
YrT x Cohort	1.23	2	.64
Error	1.92	85	
Foil Type (FT)	10.33	2	36.30 ***
FT x Cohort	.33	4	1.15
Error	.28	170	
YrT x FT	2.48	2	10.65 ***
YrT x FT x Cohort	.16	4	.69
Error	.23	170	

*** $p < .001$
* $p < .05$

APPENDIX 6

ANOVA Summary Table for Developmental Analysis of 9-Word Length Sentence Task for Cohort 1

Source	MS	DF	F
Year of Testing (YrT)	237.22	1	3.53
Error	67.28	27	
Phonetic Similarity (PhS)	6,136.08	1	291.34 ***
Error	21.06	27	
PhS x YrT	$.8^{9} \times 10^{2}$	1	.00
Error	21.62	27	

*** $p < .001$

APPENDIX 7

ANOVA Summary Table for Developmental Analysis of

13-Word Length Sentence Task for Cohort 3

Source	MS	DF	F
Year of Testing (YrT)	1,816.22	1	9.18 **
Error	197.88	28	
Phonetic Similarity (PhS)	1,753.46	1	26.03 ***
Error	67.37	28	
PhS x YrT	118.01	1	1.81
Error	65.21	28	

*** $p < .001$
** $p < .01$

APPENDIX 8

Tables of Intercorrelations of Variables for Multiple Regression Analyses

Table A

Intercorrelations for Cohort 1

		(1)	(2)	(3)	(4)	(5)
(1)	Letter 1					
(2)	Word 1	-.11				
(3)	Sentence 1	.06	.25			
(4)	IQ	-.17	.29	-.07		
(5)	Reading 1	.35	.34	.19	.13	
(6)	Letter 2	--	--	--	.20	--
(7)	Word 2	--	--	--	.08	--
(8)	Sentence 2	--	--	--	-.21	--
(9)	Reading 2	.29	.34	.13	.44	--

		(6)	(7)	(8)
(7)	Word 2	-.10		
(8)	Sentence 2	.37	.01	
(9)	Reading 2	.41	-.29	.07

Table B

Intercorrelations for Cohort 2

	(1)	(2)	(3)	(4)	(5)
(1) Letter 1					
(2) Word 1	-.09				
(3) Sentence 1	.17	.05			
(4) IQ	-.07	.06	.13		
(5) Reading 1	.43	.06	.28	.26	
(6) Letter 2	--	--	--	.04	--
(7) Word 2	--	--	--	.02	--
(8) Sentence 2	--	--	--	-.21	--
(9) Reading 2	.61	-.06	.10	.24	--

	(6)	(7)	(8)
(7) Word 2	.09		
(8) Sentence 2	.28	.04	
(9) Reading 2	.15	-.07	.13

Table C

Intercorrelations for Cohort 3

		(1)	(2)	(3)	(4)	(5)
(1)	Letter 1					
(2)	Word 1	-.18				
(3)	Sentence 1	-.14	.35			
(4)	IQ	.14	.05	.05		
(5)	Reading 1	.11	.03	.12	.12	
(6)	Letter 2	--	--	--	.07	--
(7)	Word 2	--	--	--	.47	--
(8)	Sentence 2	--	--	--	.16	--
(9)	Reading 2	-.25	.13	.09	.37	--

		(6)	(7)	(8)
(7)	Word 2	.26		
(8)	Sentence 2	.20	-.17	
(9)	Reading 2	.07	.02	-.12