PHONOLOGICAL AWARENESS COMBINED WITH EXPLICIT ALPHABETIC CODING INSTRUCTION IN KINDERGARTEN: CLASSROOM OBSERVATIONS AND EVALUATION

by

Karen Sumbler

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Human Development and Applied Psychology
Ontario Institute for Studies in Education of the University of Toronto

© Copyright by Karen Sumbler (1999)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-41321-7
EXPRESS ALPHABETIC CODING INSTRUCTION IN KINDERGARTEN:  
CLASSROOM OBSERVATIONS AND EVALUATION  
Doctor of Philosophy, 1999, Karen Sumbler  
Graduate Department of Human Development and Applied Psychology  
Ontario Institute for Studies in Education of the  
University of Toronto  

ABSTRACT  

Recent research has consistently demonstrated that through explicit teaching, beginning readers can develop the abilities (i.e., alphabetic coding and phonological awareness) considered to be key to early reading success. However, there is controversy as to the best means of promoting the development of these skills in the classroom. Through comparisons with control classrooms, the current study evaluated the effectiveness of a commercially available early-reading teaching program (Jolly Phonics - Lloyd, 1993) which combines these key elements. Time-sampling observations of experimental and control classrooms were undertaken to investigate which literacy components were related to reading and spelling success. Midway through the school year, children in 10 experimental and 10 control senior kindergarten classes (265 participants) were pretested on phonological awareness and early literacy measures to ensure group comparability. Classroom observations took place throughout the spring. At the end of the year children were tested on phonological awareness, and on a range of reading and spelling measures. Results showed that on 16 of the 19 measures including complex literacy skill knowledge (alphabetic coding) and application (reading and spelling both real words and nonwords), the experimental group significantly (and practically -- i.e., most effect sizes were large) outperformed the control children. In addition, the performance of children considered to be "at-risk" for future reading failure (determined by very low pretest letter-naming scores) was analyzed. Results showed that
this subsample of children also benefited from the Jolly Phonics program and outperformed their at-risk control counterparts. Furthermore, on most measures, the at-risk experimental group performed as well as the control children who were designated as having average literacy skills (and on one measure, scored significantly higher), thus, appearing to have changed their at-risk status. Correlational and multiple regression analyses found several links to literacy achievement, with the Phonics Literacy Component being the strongest predictor, followed by time spent involved in letter formation. A unique mnemonic feature of the Jolly Phonics program (meaningful actions associated with letter-sounds) was also related to outcome measures.
ACKNOWLEDGMENTS

There are many people who have contributed in different ways to the completion of this research, and to whom I wish to express my gratitude. First, and foremost, I wish to thank my thesis supervisor, Dr. Dale Willows. Through her unfailing guidance and support, she has not only sustained my interest and enthusiasm through the inevitable ups and downs associated with the completion of a dissertation, but has been both a mentor and a friend. Her range of knowledge in the field of literacy is remarkable, and I am fortunate to have had the opportunity to benefit from her expertise.

I also wish to thank Dr. Esther Geva and Dr. Tom Humphries for serving on my thesis committee and providing direction, feedback, and important fresh insights. A special thanks is extended to Dr. Phil Nagy, who, as a "Good Samaritan", helped immensely in statistical crises. Thanks also, to Dr. Richard Kruk for statistical advice and helpful pointers.

I wish to express my gratitude to the many individuals involved in assisting with the actual mechanics of this project, with regard to both data collection and the intricate scoring procedures.

I am grateful to have had the help and "company" from Cynthia, a friend and colleague who was concurrently going through the same thesis tribulations as I, and with whom I exchanged numerous ideas and strategies.

Special appreciation is extended to my family, who have provided continuous heartfelt support during this particular project, and throughout the entire process of my career change.

Finally, I wish to thank a dear friend and my staunchest advocate, Deborah Duggan. Her encouragement, frequent reminders to keep things in perspective, and especially her sense of humour contributed greatly to the realization of my goals.
# TABLE OF CONTENTS

Title page ................................................................. i
Abstract ......................................................................... ii
Acknowledgments ......................................................... iii
Table of Contents ....................................................... iv
List of Table and Figures ............................................... vi
List of Appendices ..................................................... ix

**Introduction** ........................................................................ 1
   Agreement Upon Issues .................................................. 2
   Issues with Growing Support ....................................... 2
   Controversial Issues .................................................... 2
   Literature Review ........................................................ 3
   Whole Language .......................................................... 3
   Phonics .................................................................... 4
   Whole Language vs. Phonics and Implicit Phonics vs. Explicit Phonics ........................................ 5
   Predictors of Reading Success ..................................... 12
   Spelling-Sound Correspondence: Phoneme Level Vs. Onset and Rime ........................................ 13
   Phonological Awareness Training: Alone vs. in Combination ..................................................... 18
   Summary .................................................................. 24
   Rationale for the Current Research ............................... 26
   Research Objectives .................................................... 29
   Research Questions ..................................................... 31

**Method** ........................................................................ 33
   Overview .................................................................... 33
   Participant Characteristics and Group Comparability .... 34
      Participants ............................................................. 34
      Group Comparability .............................................. 35
   Time-Line .................................................................. 36
   Test Measures and Procedures ................................... 39
      Phase 1 Test Measures .............................................. 40
      Phase 1 Procedures: Testing and Scoring .................. 42
      Phase 2 Test Measures .............................................. 44
      Phase 2 Procedures: Testing and Scoring .................. 46
      Additional Scoring Procedures ............................... 48
   Program Implementation and Classroom Observations .... 52
      Jolly Phonics Program ............................................. 52
      Jolly Phonics Teacher Training and Program Implementation .................................................. 56
      Control Programs ................................................... 58
      Classroom Observation Measure: Time-Sampling .... 59
      Literacy Component Categories ............................... 60
      Classroom Observations: Training, Reliability, and Time-Sampling Procedures .................. 63

**Results and Discussion** ............................................... 67
   Overview .................................................................... 67
   Variable Screening .................................................... 67

PART 1 .............................................................................. 68
   Phase 2 (Posttest): Group Comparisons on Outcome Measures ................................................. 68
      Outcome Results ...................................................... 70
      Comparison of Change Scores in Phonemic Awareness ......................................................... 80
      Correlations of Phase 2 Outcome Measures .......................................................... 81
### Tables and Figures

#### Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Overview of Study</td>
<td>33</td>
</tr>
<tr>
<td>Table 2</td>
<td>Participant Characteristics and Group Comparability</td>
<td>36</td>
</tr>
<tr>
<td>Table 3</td>
<td>Time-Line Design</td>
<td>38</td>
</tr>
<tr>
<td>Table 4</td>
<td>Key to Abbreviations</td>
<td>39</td>
</tr>
</tbody>
</table>

**Outcome Comparisons**

| Table 5 | Set 1 (Standard Scores Measures)                                       | 71   |
| Table 6 | Set 2 (Full Word Measures)                                             | 72   |
| Table 7 | Correlations of Phase 2 Outcome Measures of Phonemic Awareness, Basic Literacy Skills, and Alphabetic Coding | 83   |
| Table 8 | Intracorrelations Among Phase 2 Outcome Measures of Phonemic Awareness, Basic Literacy Skills, and Alphabetic Coding | 84   |

**Observations**

| Table 9 | Non-Literacy Time Period Comparisons: Jolly Phonics Classes vs. Control Classes | 86   |
| Table 10| Percent of School Day Spent in Literacy-Related Activities: Jolly Phonics Classes vs. Control Classes | 87   |
| Table 11| Correlations Between Literacy Components (and Other) and Set 1 (Standard Scores Measures) | 90   |
| Table 12| Correlations Between Literacy Components (and Other) and Set 2 (Full Word Raw Score Measures) | 92   |
| Table 13| Correlations Between Literacy Components and Set 3 Measures -- Subsets A and B (Phonemic Awareness, Literacy Skills and Alphabetic Coding) | 94   |
| Table 14| Correlations Between Literacy Components and Set 3 Measures Subset C (Phonemic Analysis) | 97   |

**Regression**

| Table 15 | Literacy Components Predicting Set 1 (Standard Scores Measures) | 103  |
| Table 16 | Literacy Components Predicting Set 2 (Full Word Measures)       | 105  |
| Table 17 | Literacy Components Predicting Set 3 - Subset A (Phonemic Awareness and Basic Skills Measures) | 107  |
| Table 18 | Literacy Components Predicting Set 3 - Subset B (Alphabetic Coding Measures) | 108  |
| Table 19 | Literacy Components Predicting Set 3 - Subset C (Phonemic Analysis-Real Word Measures) | 111  |
| Table 20 | Alternate Analysis for Literacy Components Predicting Phonemic Analysis of WRAT Reading and Burns and Roe Word Recognition | 112  |
| Table 21 | Literacy Components Predicting Set 3 - Subset C (Phonemic Analysis-Nonword Measures) | 114  |
Phonics Subcategories

Table 22. Percent of School Day Spent in Phonics Subcategory Activities:
Jolly Phonics Classes vs. Control Classes ........................................... 118

Table 23. Intracorrelations Among Phonics Subcategories ......................... 119

Table 24. Correlations Between Phonics Subcategories and Set 1 Measures (Standard Scores) .......................................................... 121

Table 25. Correlations Between Phonics Subcategories and Set 2 Measures (Full Word Raw Scores) ......................................................... 122

Table 26. Correlations Between Phonics Subcategories and Set 3 Measures
(Phonemic Awareness, Literacy Skills and Phonemic Analysis) ................ 126

At Risk Comparisons

Table 27. Phase 1 Comparisons (Pretest Measures):
Jolly Phonics AR vs. Control AR Children ........................................... 130

Table 28. Phase 1 Comparisons (Pretest Measures):
At-Risk vs. Non-At-Risk Children ..................................................... 132

Table 29. At-Risk Phase 2 (Posttest) Comparisons:
Set 1 (Standard Score Measures + Burns & Roe Word Recognition) ........ 134

Table 30. Correlations Between Phase 1 (Pretest) and Phase 2 (Posttest) Scores on Measures of Phonemic Awareness and Literacy Skills
For Jolly Phonics At-Risk Children ..................................................... 141

Table 31. Correlations Between Phase 1 (Pretest) and Phase 2 (Posttest) Scores on Measures of Phonemic Awareness and Literacy Skills
For Jolly Phonics At-Risk Children ..................................................... 142

Table 32. Phase 1 (Pretest) Comparison:
Jolly Phonics At-Risk Children vs. Control Average Children ............. 144

Table 33. Phase 2 Comparisons (Outcome Measures):
Jolly Phonics At-Risk Children vs. Control Average Children ............. 140
Set 1 and 2 (Standard Score and Full Word Raw Score)

Table 34. Phase 2 Comparisons (Outcome Measures):
Jolly Phonics At-Risk Children vs. Control Average Children
Set 3 (Phonemic Awareness, Basic Skills, Coding, Phonemic Analysis) ..... 147

Figures

Figure 1. Jolly Phonics: Sample Sound Picture ..................................... 54

Figure 2. Observation Scoring Example .............................................. 66

Outcome Comparisons

Figure 3. Phase 2 (Posttests) Alphabetic Coding Performance:
Jolly Phonics vs. Control ............................................................... 75

Figure 4. Spelling Examples .............................................................. 76
Figure 5  Phase 2 (Posttest) Phonemic Analysis Performance - Real Words:
Jolly Phonics vs. Control................................................................. 78

Figure 6. Phase 2 (Posttest) Phonemic Analysis Performance - Nonwords:
Jolly Phonics vs. Control................................................................. 78

At-Risk Comparisons

Figure 7. Phase 2 (Posttest) Full Word Raw Score Performance:
Jolly Phonics AR vs. Control AR ....................................................... 135

Figure 8  Phase 2 (Posttest) Alphabetic Coding Performance:
Jolly Phonics AR vs. Control AR ....................................................... 137

Figure 9  Phase 2 (Posttest) Phonemic Analysis Performance
Jolly Phonics AR vs. Control AR ....................................................... 139
Appendices

Appendix A
   Nonstandardized Measures, Examiner Instructions, and Scoring Procedures ........ 178

Appendix B
   Jolly Phonics Program Samples and Outline .................................................. 195

Appendix C
   Observation Manual and Sample Protocol Page .............................................. 202

Appendix D
   Observation Reliability ................................................................................. 230

Appendix E
   Alternate Analysis of Phase 2 Outcome Measures: Class Means ....................... 236

Appendix F
   Alternate Analysis of At-Risk Subsample: ESL children removed from sample .... 239
INTRODUCTION

The importance of equipping children early in their school careers with the skills necessary for successful reading and writing cannot be overstated; and, there are few who would argue with this view. There is, however, considerable disagreement about: which components are important and/or necessary for reading acquisition, the order in which they should be taught, and the method by which they should be conveyed. The current study addresses these issues through "on-line" monitoring and evaluation of a commercially available reading program which is comprised of many of the elements which have been found in the literature to be important for early reading acquisition.

Agreed Upon Issues

Although there have been many controversies over the years with regard to the process of reading and reading acquisition, agreement has developed on a few main issues. For instance, there is little doubt about the importance of word recognition in successful reading. If words are not recognized, their meanings cannot be discerned, and the entire purpose of the reading process -- comprehension -- is undermined. There is also agreement among most researchers that a key element influencing the ease with which children learn to read is the underlying process of phonological awareness\(^1\) -- the ability to detect and manipulate the sounds making up spoken words (e.g., Adams, Foorman, Lundberg, & Beeler, 1996; Ehri & Wilce, 1980, 1985; Fowler, 1991; Williams, 1979, 1980). The literature abounds with not only correlations of this process with reading ability, but much of the research demonstrates its predictive value as well (e.g., Ball, 1993; Bradley & Bryant, 1983; Juel, 1991; Rieben & Perfetti, 1991; Share, Jorm, Maclean, & Matthews, 1984; Stanovich, Cunningham, & Cramer, 1984).

\(^1\) Often the terms phonological awareness and phonemic awareness appear to be used interchangeably, but technically, phonemic awareness refers to the understanding that words are made up of phonemes -- the individual units of speech sounds; and phonological awareness is a more global term encompassing larger units of speech as well.
Issues with Growing Support

There is a growing literature which supports a reciprocal relationship of phonological awareness and reading. That is, although being "tuned" to the sounds in words facilitates reading acquisition, it seems that learning to read -- the process of decoding words -- also enhances phonological awareness (Ehri, 1998; Ehri & Wilce, 1987; Goswami & Bryant, 1990; Perfetti, Beck, Bell, & Hughes, 1987; Stanovich, 1990; Stuart & Coltheart, 1988; Wagner, Torgesen, & Rashotte, 1994).

Another area of research about which agreement is continuing to grow involves the importance of alphabetic coding as a facilitator of reading. It appears that this factor is also highly predictive of word recognition (Vellutino, Scanlon, Small & Tanzman, 1991; Vellutino, Scanlon, & Tanzman, 1991). Alphabetic coding refers to the ability to make the connection between sounds and the letters that represent them. It lays the foundation for comprehension of the alphabetic principal; and the application of this principal -- decoding -- is the single best predictor of reading comprehension (Stanovich, 1990; Vellutino, 1991).

Controversial Issues

When putting all the new developments in basic reading research into practice, the main controversial question is still "What is the best way to teach children how to read?" There are many areas related to this question over which old debates still ensue (i.e., the "Great Debate" -- see below); and there are newer areas of contention which have arisen from the research regarding phonological awareness and alphabetic coding. Principal concerns center around the type and the timing of phonological awareness and alphabetic coding instruction/training: -- should the elements of alphabetic coding be taught systematically and explicitly, or should their acquisition be implicit in nature (i.e., incidental), highlighted only when needed "as an aid to a child's ongoing process of getting meaning from a text..." (Stahl, Duffy-Hester, & Dougherty Stahl, 1998, p. 339); should direct letter-sound correspondences (i.e., grapheme/phoneme connections) be employed, or is the process of onset and rime (utilizing units of speech larger than phonemes) a less "difficult way into reading" (Goswami, 1994, p. 33); or, should instruction in alphabetic
coding be delayed altogether, until phonological training can first be thoroughly developed as a priming step; and, do any conclusions regarding the above issues apply to children considered to be at risk for reading failure?

The core of this literature review will present research related to the more recent areas of controversy. However, it will start by reviewing older issues at the heart of the "Great Debate" since, although a large body of research has consistently resulted in essentially the same findings pointing to essentially the same conclusions, this research has also developed some newer questions -- changed in nature, from "Which is best?", to "How much of each, and when?"

**Literature Review**

As mentioned in the introduction, there has been a "Great Debate" (Chall, 1967, 1983, 1996) over the years as to which is the "best method" of teaching children how to read. This controversy is well known to most educators and has, in the past, developed into a polarization of views with those promoting "whole language" (meaning emphasis) at one extreme, and supporters of "phonics" (code emphasis) at the other. Although the extremes of each perspective have now softened to some extent, and most researchers agree that both meaning and the alphabetic principle are important aspects of reading acquisition, there are still controversies over the extent to which each should be emphasized. As well, timing is an issue, as some research suggests different approaches at different stages of beginning reading (e.g., Stahl, McKenna, & Pagnucco, 1994; Stahl & Miller, 1989).

**Whole Language**

Over the years, there has been an evolution of meaning-emphasis teaching methods with whole language being the most recent incarnation. All of these approaches have included differences in specific details of programming, and even within whole language itself there is no single clear-cut definition (Adams, 1991; Begeron, 1990). However, generally speaking, these
whole language/language experience-type approaches share many key concepts, and so, here, the term "whole language" will be used to refer to all meaning emphasis approaches. The following general elements are central to the whole language philosophy: a) emphasis is put on the functionality and meaning of reading and writing; b) learning is collaborative with teachers providing guidance, suggestions, and invitations to explore learning activities, and children are "encouraged to suggest alternative strategies or propose new directions for themselves" (Newman & Church, 1990, p. 22); c) activities should be authentic and motivating; d) children's literature is used as the basis of reading instruction, rather than readers with controlled vocabulary; e) the "mechanics" of reading and writing, including letter-sound associations, should be taught incidentally, in context, and as the need arises; and f) since it is felt that children acquire literacy as "naturally" as they acquire oral language, direct systematic instructional practices are considered "unnatural" (Aukeman, 1984; Bergeron, 1990; Goodman, 1986; Goodman & Goodman, 1982, Newman & Church, 1990; Stahl, et al., 1998).

Whole language practices minimize the teaching of phonics (letter-sound instruction), avoiding any "artificial" decomposition of language into its components (e.g., phonemes); and whole language advocates suggest that direct systematic phonics will actually interfere with the process of learning to read and write (Goodman, 1986). The whole language philosophy is based on a belief that through the many authentic literacy experiences, children will develop word recognition and spelling skills without ever having to be consciously aware of the intrinsic make-up of words or their linguistic counterparts, and that most will learn the alphabetic principal "intuitively" through activities such as inventive spelling (Clarke, 1988).

**Phonics**

Phonics is the term customarily used for any kind of instruction in the alphabetic principle. It is the teaching of letters and their sounds, and can be implemented in a number of ways ranging from direct, systematic instruction of letters and sounds in a specified order and supplemented by corresponding reading materials (i.e., synthetic phonics), to more indirect approaches (i.e.,
analytic, embedded) where letter-sound relationships are taught through analysis of complete words, often within the context of text -- usually basal readers (Chall, 1983, 1996).

More specifically, synthetic phonics starts with directly teaching children individual letters (usually, but not always, by letter name) and the sounds that these letters represent (Stahl, et al., 1998). Then children are taught how to blend these letter-sounds together to make words (in isolation). After the particular letter-sound (or sound pattern) is introduced, several other words with the same sound are presented. Then children might read a simple story containing many words with the sound just taught. Sometimes these stories are stilted and contrived. Synthetic phonics programs differ in the sequencing of the letters taught (or in how many letters are presented before words are introduced), but their overall approach is the same (Aukerman, 1984; Willows, Borwick, & Hayvren, 1981).

Analytic phonics starts with a complete word that children already know (often embedded within text), and it is then broken down (analyzed) into smaller parts (Chall, 1983, 1996). This may include analyzing the individual phonemes (each letter[s]-sound correspondence), or sound patterns may be examined, as in word families -- for example, hat, cat, pat (Stahl, et al., 1998). Usually, a sight-word vocabulary is first established.

The following section examines the research literature on whole language versus phonics as initial approaches to literacy, and within this context the issue of implicit (indirect, analytic, embedded phonics) versus explicit (systematic, synthetic phonics) is also examined. As will become evident, the research generally points to not only superior results of phonics programs over whole language, but that the most successful phonic programs were those that utilized explicit systematic strategies to teach the alphabetic principle.

**Whole Language vs. Phonics and Implicit Phonics vs. Explicit Phonics**

There have been many reviews of the reading research literature comparing whole language programs with phonics. In evaluating these findings, it should be kept in mind that although a program was labeled as either "whole language" or "phonics", there was a great deal of variation in the actual methods and materials employed within each of these categorizations of instructional
approaches. In fact, much of the research compares classes using whole language methods (and language experience, etc.) with those using basal reading programs rather than ones that were specifically termed phonics (e.g., Stahl & Miller, 1989). Because most basal programs have some instruction in phonics this was deemed as an appropriate comparison, however, it must be noted that "while there are some basals which primarily use phonics instruction to teach reading, most begin instruction with sight words and place more emphasis on comprehension than on phonics." (Cunningham & Allington, 1994, p. 13). Furthermore, very few studies included any type of monitoring to ensure that the purported approach was the one that was actually being adhered to in the classrooms. As researchers and reviewers have pointed out, there has been little documentation of actual instructional behavior and classroom activities (Evans, 1979, 1985; Sacks & Mergendoller, 1997; Stahl et al., 1994). This lack of observational data and the broad range of inclusiveness (for both methods) may account for some of the ambiguity in the research literature. However, despite these problems, the same general findings keep appearing in research syntheses, and they are consistent with those elucidated 30 years ago by Chall (1967).

The review (and observations, interviews and program analyses) conducted by Chall (1967) was one of the first (and most influential) large-scale syntheses of information on reading of the time. Included in her analysis were many studies of varying quality comparing different programs on various outcome measures at different times. The limitations inherent in analyzing such a vast array of data are numerous and thus, Chall's findings and conclusions were not without criticism (e.g., Carbo, 1988; Rutherford, 1968). However, her findings have withstood the test of time and have been supported repeatedly, both by other reviewers, and by research from different disciplines (e.g., Adams 1990; Bond & Dykstra, 1967; Perfetti, 1985; Stahl et al., 1994; Stahl & Miller, 1989), as well as her own update on more recent literature (Chall 1983, 1989, 1996).

Among other important findings (to be discussed later), Chall found generally, that programs using systematic phonics resulted in better performance on measures of word recognition, spelling, vocabulary, and comprehension up to second and third grade. Furthermore, these programs were at least as beneficial for children considered to be at risk for future reading
failure (because of lower socioeconomic backgrounds and/or lower levels of readiness skills). She also found some early advantages of whole language programs with regard to reading fluency and motivation, thus pointing to the importance of practice in reading, and challenging text to provide fluency and to maintain interest. Some educators and researchers have interpreted Chall's findings as strong support for a systematic, code emphasis teaching approach. Others view Chall's results as support for a combination view of reading -- that of meaningful and connected reading along with a systematic phonics program (Adams, 1990; Snow, Burns, & Griffin, 1998).

The importance of a combination of systematic phonics and meaning emphasis was also supported by a large-scale research project (i.e., the Cooperative Research Program) undertaken by the U.S. Office of Education (USOE) in the mid-sixties. This project coordinated by Bond and Dykstra (1967) was designed to assess the different approaches to reading instruction as well as to determine other contributors (i.e., characteristics of children, teachers, and schools) to reading success. The coordinators consistently found that systematic code emphasis programs surpassed straight basal programs (which consisted of a range of more eclectic approaches which may or may not have had some teaching of phonics, but not direct and systematic) in word recognition performances. Moreover, when these programs were supplemented by connected meaningful reading practices, they exceeded the straight basal programs on almost all of the outcome measures. Student characteristics did not influence the outcome. That is, if a program worked, it worked for at-risk children just as well as it did for those considered not to be at risk, and so as Adams remarks in her review of this project, "There appeared...to be no basis for the widely held belief that systematic phonic instruction is useful for only brighter children" (1990, p. 42). Other reviewers and researchers have drawn similar conclusions (Gersten, 1984; Pflaum, Walber, Karegianes, & Rascher, 1980).

More recently, a meta-analysis of the literature was undertaken by Stahl and Miller (1989) comparing whole language/language experience approaches with "traditional basal" programs (undefined, but appear to be reflective of at least some phonics content). In their re-analysis of some of the older studies, as well as additional new ones, Stahl and Miller drew conclusions slightly different from those presented thus far. They found stronger effects for the whole
language/language experience approaches in kindergarten where these approaches were used as a readiness program (i.e., preparing children for another reading approach) compared with the readiness version of the basal programs. However, these results may not hold true when kindergarten comparisons of whole language are made with programs which involve more than readiness skills teaching (i.e., programs which implement systematic letter-sound instruction in kindergarten).

In the overall analysis, Stahl and Miller also found that when whole language/language experience approaches were used as the "beginning reading approach" (as opposed to readiness), the effects were generally no different than those produced by the basal programs. However, when the quality of the studies was taken into account (i.e., comparing only studies with higher rated methodologies), results tended to favor the basal programs. The authors concluded that whole language/language experience methods may be more appropriate for kindergarten where children are learning about print and print concepts, and that once these skills are mastered, a more systematic code emphasis approach will likely be needed. In an update of this review Stahl, McKenna, and Pagnucco (1994), still basically confirmed this interpretation but added that, overall, they feel a more eclectic approach to reading in general should be taken -- one that combines the best of both methods (i.e., emphasizing decoding, comprehension, and motivation) and tailored to meet the needs of the children being taught. However, Stahl and colleagues (1994) still favor whole language approaches in kindergarten, especially for at-risk children so they can be introduced to the experience of reading.

The view that whole language is more appropriate for kindergarten is supported by Sacks and Mergendoller (1997). In one of the few studies that included classroom observations, these researchers examined reading achievement (and affect) of kindergarten children with different initial reading abilities who were taught by teachers with either a whole language or a phonics orientation. Teachers' theoretical orientation was determined by their scores on the Theoretical Orientation to Reading Profile (DeFord, 1985). This questionnaire assesses teachers' orientation by their responses on items representing three dimensions: phonics, skills, and whole language; the phonics and skills domain are considered to be closely related. Through observations the
authors were also able to compare the classroom context and literacy tasks established by the teachers.

Only one reading measure (discussed below) was used to assess reading ability at pretest and posttest. End-of-year performance showed that lower ability children (as determined by pretest scores) in the classrooms of whole-language oriented teachers improved their scores significantly more than their counterparts in the classrooms of phonics-oriented teachers. However, the authors' definition of reading gains is problematic. The measure used for testing children's reading achievement (The Test of Early Reading Ability-2, Reid, Hresko, & Hammill, 1989) was "designed to mitigate some of the concerns of whole language advocates regarding standardized assessment" (Sacks & Mergendoller, 1997, p. 725). As these authors describe, this instrument measures three aspects of reading: constructing meaning from print (e.g., awareness of print in the environment and in connected discourse), knowledge of the alphabet (e.g., letter names, numerals, and "oral reading" -- undefined), and children's awareness of the concepts of print (e.g., proper book handling, print conventions). To call improved performance on this measure an indicator of reading gains seems a bit misleading. Improvement in reading "readiness" performance would be a more appropriate description. There were no measures of spelling.

Observation results pertinent to literacy showed that out of the 13 "classroom activities" Sacks and Mergendoller examined, the only difference in time spent was found on worksheet activities. Children in the "phonics" classrooms (i.e., classes with phonics-oriented teachers) spent marginally significantly more time completing worksheets than did those in the whole language classrooms. "Task selection" comparisons revealed that children in the phonics classrooms spent marginally more time engaged in writing activities. With regard to types of writing task, children in the phonics classrooms spent marginally more time copying letters, words and sentences, and children in whole language classrooms spent more time using invented spelling and dictating stories. Concerning reading tasks, phonics classroom children spent more time looking through books silently, marginally less time "attending" to environmental or nonbook print, and marginally more time completing reading worksheets (the term marginally here refers to significance levels in the range of .05 < p < .10). However, since all analyses were individual
ANOVA's and therefore subject to inflated Type I errors (the authors did not perform an initial omnibus MANOVA, or use a Bonferonni correction), any differences reported at these marginal significance levels appear to be highly questionable. Overall, the authors themselves report that although the two groups of teachers held different theoretical views toward literacy, in reality, their classrooms were very similar. The differences that were found were ones more of degree than of kind, and, since the authors did not elaborate further on specific content (i.e., there was no description of any type of alphabetic coding instruction or word analysis), it is difficult to draw the same conclusions as the authors. In other words, it does not appear that the "phonics-oriented" teachers conducted a very explicit phonics program in their kindergarten classes. Furthermore, none of the literacy behaviors (i.e., percentages of time spent on the different literacy tasks) predicted outcome scores, and, therefore, it is difficult to attribute scores to any particular constellation of program components. This study underscores the importance of classroom observation for not only aiding in performance data interpretation, but in determining whether actual program make-up corresponds to the label being applied.

In her comprehensive synthesis of the reading literature, Adams' (1990) conclusions were strikingly similar to those of Chall almost 25 years earlier. She determined that effective reading instruction includes direct, systematic phonics as well as practice reading and much experience with interesting, motivating, and challenging literature. As did the reviews before her, Adams found that these results applied at least as well to at-risk children. What is particularly important about Adams' review was that her conclusions were based on literature from several different areas of psychology including developmental, cognitive and psycholinguistic (Snow, et al., 1998). Converging evidence from different disciplines always makes for stronger inferences. By drawing upon these different literatures, Adams was able to examine some of the cognitive processes underlying the act of reading, and also provide support for the importance of phonological awareness in developing alphabetic coding skills and for reading in general.

Before moving on to the next area of contention, there is one final point about phonics versus whole language programming that should be noted. The focus on direct, systematic phonics instruction in early grades compared to the more language-experience focus of whole
language programming could lead to concerns about children's oral language development in the direct-code classrooms. It appears, however, that language development is not compromised in classrooms practicing a more routinized, basal/phonics-oriented approach. In a classroom observational study conducted by Evans and Carr (1985), the authors examined a number of abilities, including linguistic and reading achievement of 20 classes of children (involving grades 1 and 1-2), comparing individualized "language-oriented" programs with those emphasizing a basal/phonics-oriented approach. The scores on measures of oral language for children in the language-oriented classes were no higher than those for the children in the basal/phonics programs. In fact, the children in the basal/phonics program actually scored slightly (but significantly) higher on one measure of expressive language than did the children attending language-oriented classes. In addition, they achieved significantly higher scores in reading comprehension than their counterparts in the language-oriented classrooms. Classroom observations verified that there were differences in instructional approaches in these primary grade classes, with significant differences in time spent on various literacy activities; through these observations, positive correlations between time spent on some of the literacy activities (e.g., word analysis, guided silent reading, printing) and reading achievement were revealed.

Similar results with regard to phonics instruction not encroaching on oral language development were reported for children at the junior kindergarten level -- even when classes included children for whom English was not their first language (Morgan & Willows, 1996). There were no differences on measures of oral language development between experimental and control groups, -- that is, the children in the classes where the phonics program (i.e., the same program that is under investigation in the current study -- Jolly Phonics) was implemented as part of their normal 1/2 day curriculum and those in control classrooms who received their normal curriculum alone (i.e., traditional junior kindergarten which focused mainly on play and language development). What was of particular interest, however, was performance outcomes on the literacy and phonological measures for the children who were "at risk" (due to having limited English language skills associated with less English language exposure). Not only did the ESL (English as a second language) children in the Jolly Phonics classes perform at par with their
non-ESL (Jolly Phonics) classmates on almost every measure of reading, spelling, and phonological processing, but their performance on these measures was significantly better than the control non-ESL children (i.e., children whose first language was English).

*Predictors of Reading Success*

Besides examining the general efficacy of different reading programs, Chall (1967) also tried to find relationships between specific abilities/skills and reading performance. Her findings, and subsequent research by others, have established that for children entering kindergarten, letter-name knowledge and phonological awareness, are the two most significant predictors of successful reading (Adams, 1990; Bond & Dykstra, 1967; Chall, 1967, 1983, 1996; Perfetti, et al., 1987, Scarborough, 1998). There are, of course, a number of other correlates of reading acquisition, such as socio-economic status, and having a language/dialect other than standard English as a first language (Snow et al., 1998). Various groups of "reading readiness" measures and language skills in general also predict reading performance (e.g., Adams, 1990; Badian, 1994; Hammill & McNutt, 1980). Memory also seems to be involved, and recent research has shown that verbal memory for sentences is close in prediction strength to the two top single predictors of phonological awareness and letter-name knowledge (Scarborough, 1998). Verbal and general IQ are considerably further down the list of predictors, being approximately equal to receptive language measures, and nonverbal IQ scores are poor predictors (Snow et al., 1998). Letter identification in and of itself is probably not a direct facilitator of reading, but is instead an indicator of the literacy environment from which a child comes, and in that sense is strongly predictive of future reading success. On the other hand, research has conferred a more causal (and reciprocal) role in the predictive ability of phonological awareness.

Share et al. (1984) found that phoneme segmentation ability of kindergarten children was predictive (i.e., 39% explained variance) of their Grade 1 reading achievement. Research by Stanovich et al. (1984) also showed that kindergarten children's phonological ability (i.e., scores on a set of phonological tasks) was highly predictive of their reading performance in Grade 1. Multiple regression analyses indicated that the phonological variables explained 66.2% of the
variance in reading performance, and were better predictors than IQ. Other researchers have found similar prediction results (e.g., Juel, Griffith, Philip, & Gough, 1986).

Spelling-Sound Correspondences: Phoneme level vs. Onset/rime

The finding that phonological awareness is a good predictor of future reading performance highlighted the need to take this factor into consideration when conducting research and developing teaching methods for beginning reading. As more researchers examined the concept of phonological awareness, definitional issues arose. Phonological awareness can be defined by varying levels of task and linguistic complexity (Adams 1990; Stahl & Murray, 1994; Yopp, 1988). Simpler levels include the ability to segment words into syllables and into "onset" (the beginning consonant[s] of a written syllable) and "rimes" (the vowel and remaining consonants). Breaking words into further units representing each individual sound (i.e., phoneme) presents difficulties for many prereaders and represents a more complex level of phonological awareness. This refining of phonological awareness into levels of sophistication has led to questions of causality regarding the relation between phonological awareness and reading. Findings from research investigating these questions (see below) have implications for methods of reading instruction.

It has been suggested that the different levels of phonological awareness emerge at different points in development (Goswami, 1994), and although most researchers acknowledge that a basic level of phonological awareness facilitates successful beginning reading, there is disagreement about the more complex aspects. Some authors suggest that sophisticated phonological analysis (e.g., phonemic awareness and manipulation) can be developed through specific phonological awareness training (e.g., Lundberg, Frost, & Petersen, 1988; Truch, 1991). Others maintain that it is a consequence of reading (Goswami, 1994; Goswami & Bryant, 1990; Morais, Bertelson, Cary, & Alegria, 1986), and because of this, alphabetic coding used as a beginning reading approach would not be appropriate, since it targets phonemes. Goswami contends that "learning about individual spelling-sound correspondences is a rather difficult way into reading, as they [children] cannot hear these individual sounds in the words that they are trying to decode"
(Goswami, 1994, p. 33). Instead, she advocates an onset and rime approach. Goswami proposes that by introducing reading and spelling/sound correspondences at the onset and rime level (rather than the phonemic level) children (even nonreaders) will be able to read new words by analogy, and this skill in reading will then help increase phonological awareness to the more complex level of phonemic analysis. Support for the suggestion that early readers are only at a basic level of phonological awareness comes from studies (Goswami, 1986, 1988) showing that beginning readers were able to use only simple onset-rime analogy awareness when reading new words. They could successfully read a new word by analogy only if the rime was the same but onset was different (e.g., beak-peak). They did not employ the more sophisticated rime-segmentation skill needed to read words with the same onset and first part of the rime, but a different rime ending (e.g., beak-bean).

A logical analysis conducted by Stahl and Murray (1994) on data collected from kindergarten and grade-one subjects' phonological awareness task performance, indicated that the basic level of phonological awareness (onset-rime awareness) is necessary for both word reading and more complex levels of phonemic analysis, suggesting that onset-rime knowledge enables basic word recognition which may facilitate higher-order phonological awareness. These, and other studies (e.g., Morais et al., 1986) seem to suggest that development of more sophisticated levels of phonological awareness (i.e., phonemic analysis) is a result of reading. If this is the case, they lend support to Goswami’s proposal that phonemic-based phonics instruction "is not a good way into learning to read" (Goswami, 1994, p. 36). Although other researchers agree that reading facilitates phonemic awareness, they do not believe that this relationship is unidirectional, and instead propose a reciprocal relationship of phonemic awareness and reading (e.g., Perfetti et. al., 1987). They also argue that spelling-grapheme instruction directed at the phoneme level is a prerequisite for distinguishing onsets from rimes (e.g., Ehri & Robbins, 1992).

In response to Goswami’s (1986, 1988) claim that nonreaders can use analogies to read words, Ehri (1987) proposes that in order to "read by analogy" children must first be good "recoders" (decoders). She suggests that readers must have sufficient phonological recoding skills (i.e., knowledge of phoneme/grapheme correspondence and its application) to be able to "store the
spellings of words in their lexical memory," and to differentiate between these stored words and their analogues. In other words, readers must have some initial phoneme/grapheme knowledge in order to utilize the reading-by-analogy strategy. In support of this contention, evidence from research assessing analogical transfer of training words conducted by Ehri and Robbins (1992) showed that participants in the experimental condition (i.e., who had training on analogues of transfer words) who were nondecoders (i.e., had no phoneme/grapheme skills) rarely read any words by analogy in nonsense-word-training tests designed to measure reading by analogy. Instead, they mistook the transfer words (i.e., analogues of the training words) for "training words" — that is, they did not realize that although the transfer word had the same rime, it did not have the same onset, reflecting their lack of phonemic analysis of the onset. In fact, there was no difference in performance on transfer-word reading between experimental nondecoders (whose transfer words were analogues of the training words) and control nondecoders (whose transfer words were not analogues of the training words, but did share some orthographic similarities). These nonreaders were unsuccessful in using the onset/rime strategy to correctly "read" new (analogous) words, contrary to Goswami's position.

In contrast, results on the same measures for subjects classified as decoders revealed that those in the experimental group significantly out-performed control subjects when reading transfer words. Although these latter results support one of Goswami's claims that beginning readers read by analogy more easily than by decoding each phoneme, they also verify Ehri's contention that in order for the onset/rime strategy to be successful, the child using it must already have enough grapheme/phonemic knowledge to distinguish the different onsets. The strategy was of no use to the children without these skills, suggesting that onset-rime spelling/sound correspondences may not be the best level of initial reading instruction. Ehri concludes that it is the decoders phonological recoding analytic skills which allowed them to take advantage of the analogical relation between training words and their transfer analogues. Thus, while onset/rime strategies may enhance reading skills that are already in place (thereby perhaps improving existing phonemic awareness abilities), they do not appear to be the appropriate first level of reading instruction.
The Ehri and Robbins study (1992) supports other research indicating that sequential phonemic decoding skill\(^2\) (as developed through instruction in phoneme/grapheme correspondence) is needed in order to read by analogy (e.g., Marsh, Friedman, Desberg, & Saterdahl, 1981; Marsh, Friedman, Welch, & Desberg, 1981; Zinna, Liberman, & Shankweiler, 1986). It would appear that rather than posing difficulties for the beginning reader, the very nature inherent in alphabetic coding instruction (i.e., emphasizing the fact that letters make sounds and that these sounds make up words), would actually help children become aware of (or "hear") those sounds in words which Goswami argues elude prereaders. Letters actually make the phonemes "visible" (Perfetti, 1995).

Although Ehri's interpretation of her results is convincing, it is not clear if "laboratory" findings can be generalized to classroom practice. Moreover, it cannot be determined from this research whether the conclusions apply to children considered to be at risk for reading difficulty. It would seem logical that they would, since reading by analogy presupposes enough familiarity with letters (recall lack thereof is a marker of at-risk children) to be able to recognize the sameness of rimes. However, actual classroom research is needed to begin to address these questions.

A recent study conducted by Foorman, Francis, Fletcher, Schatschneider, and Mehta, (1998) lends ecological support for the superiority of phoneme/grapheme instruction over reading by analogy for at-risk children, as well as over whole language methods. These researchers compared the effects of three different reading programs on children who were economically disadvantaged and who represented the bottom quartile in their school district's emergent literacy survey of first- and second-grade classrooms. The phoneme/grapheme program (termed "direct code" instruction) focused on systematic "direct letter-sound correspondences practiced in decodable text" (p. 39). The reading-by-analogy program (termed "embedded code") focused on less direct instruction in systematic spelling-sound patterns (i.e., onset/rimes) embedded in connected text, and the whole language program (termed "implicit code") focused on "indirect,

---

\(^2\) Although more recent research suggests that some readers may not need a complete grasp of all phoneme/grapheme correspondences (Scarborough, Ehri, Olson, & Fowler, 1998), they require enough phonemic understanding to be able to decode at least the differing onsets, vowel nuclei, and codas of words.
incidental instruction in the alphabetic code embedded in connected text" (p. 39). All three programs occurred in a literature-rich environment with meaningful reading activities, and both direct and embedded code programs also included phonemic awareness activities, thus controlling for this component in the first two approaches. Teachers undertook special training in their respective programs, and twice a month, classroom monitoring was conducted to ensure program fidelity. Testing on measures of phonological processing and word reading skills took place throughout the year and at year-end (i.e., May).

Overall results on the reading achievement measures showed that at-risk children in the direct code program (explicit grapheme/phoneme) benefited the most. They improved more quickly on word reading, and their end-of-year achievement on a combination of letter-word identification and nonword reading was significantly and practically higher than that of the other two groups. Furthermore, only 16% of the children in the direct code group demonstrated no growth in word reading, compared to 44% of the children in the embedded code approach (i.e., onset/rime), and 46% of the children in the implicit code (i.e., whole language) group. As well, children in the direct code group were 2.4 times and 3.1 times less likely to score below the 25th percentile in word decoding than the children in the implicit code and embedded code approaches respectively. However, the researchers found no differences in spelling performance, and overall, spelling scores were "not impressive" (p. 52). The highest end-of-year spelling standard score (for the direct code group) was only 85.7 (SD = 12.2), where 100 is the average.

Finally, pretest phonological processing scores for children in the direct code group were much more weakly related to posttest reading scores than those of the other two groups. This weaker relation suggests that besides improving reading performance, direct code instruction also improved phonological skills. The authors explained that "we would expect initial phonological processing to be less related to outcome in DC [the direct code group] " (p. 48) if explicit instruction in the alphabetic code does enhance phonological ability; because if direct code instruction does indeed facilitate phonological processing, it would do so for all children thereby minimizing "the importance of the level of this skill that children bring to the classroom in the fall." So it seems, for a number of reasons, that reading instruction focusing on letter-sound
correspondences at the level of phonemic units of words, rather than onset/rimes, is a more successful approach to teaching reading. This study also supports the superiority of explicit code instruction over whole language methods.

**Phonological Awareness Training: Alone vs. In Combination**

Some researchers are less interested in what kind of spelling/sound correspondences should be taught. Rather, they are more concerned with the development of phonological awareness, and believe that specific training of this skill should be a main focus of early reading/prereading programming (e.g., Byrne & Fielding-Barnsley, 1989, 1990; Hurford et al., 1994). Others propose that explicit development of this skill should occur before any grapheme instruction is introduced (e.g., Lindamood & Lindamood, 1975; Truch, 1991).

A large-scale longitudinal study of Danish preschool children conducted by Lundberg and colleagues supports the view that phonological training alone can have positive effects on subsequent reading skills (Lundberg, et al., 1988). A complex sequenced training program of metalinguistic games and exercises involving rhyming, segmenting, and blending was developed by the authors. Training of the preschool teachers included an initial inservice course, and their involvement in a previous year-long pilot study undertaken to develop the phonological training program. Within their regular preschool program (which focused on social development and had no reading instruction) children in the experimental group had 20-minute daily training sessions over a period of 8 months. The control children followed the regular preschool program minus the phonological training. On posttest measures, the authors found that the children in the training program performed better than the controls on phonological tasks, particularly those at the phoneme level, and that these skills facilitated significantly superior reading and spelling performance in Grades 1 and 2.

Investigating whether or not phonological training can be beneficial for children considered to be at-risk, Brady and colleagues (Brady, Fowler, Stone, & Winbury, 1994) conducted a study examining phonological awareness training on inner-city kindergarten children (who scored below the norms on pretest measures of reading and phonological awareness). Two teachers were trained
in and implemented a complex program (devised by the experimenters, and much of which was based on the ADD method described below) in their classrooms in three 20-minute sessions per week over an 18 week period. Teachers were aided by research assistants one to two hours each week. The authors found significantly greater gains in phonological awareness (but not reading or spelling achievement) for these children over control kindergarten participants at the end of kindergarten. A one-year follow up showed that children from the experimental kindergarten classrooms had been significantly more likely to be promoted to first grade (rather than to pre-one), and, a trend toward better reading skills (compared with controls) was now evident. In Grade 1, the experimental group demonstrated significantly better performances in word identification and analysis than the control children. No testing of spelling performance was reported.

Another study supporting the success of phonological training for the at-risk population was conducted by Hurford et al., 1994. Half of the at-risk students (the experimental group) in this study were individually trained in phonological discrimination, segmenting, and blending skills in 40 sessions (of approximately 15-20 minutes) over 20 weeks. At end-of-year testing, aside from superior performance on the phonological processing tests, the experimental (trained) at-risk group significantly outperformed the control at-risk children on the reading measures (i.e., word identification and nonword reading), and were also able to halve the pretest gap between themselves and the not-at-risk readers on the word identification measure. Spelling performance was not assessed. The authors concluded that phonological training not only improves phonological awareness in at-risk students, but that this improvement translates into improved reading ability.

Research examining phonological awareness training for individuals of all ages considered to be learning disabled also resulted in clear reading improvement (Truch, 1994). Truch used the Auditory Discrimination in Depth program (ADD Program -- Lindamood & Lindamood, 1975), which begins instruction at a "sublexical phonemic processing level," for training clients ranging in age from 5 to 55 years. The ADD method first trains students in a phonological awareness of the consonants and vowels, emphasizing the articulatory actions that produce them. Through manipulating colored tiles which represent the phonemes, students learn to make increasingly more
complex phonological judgments (i.e., from simple phoneme isolation, to deletions and shifting of phonemes within a syllable) before any real word decoding or encoding instruction occurs. It is thought that once the necessary phonological skills are developed, these can be easily transferred to orthography. Truch found highly significant gains for his learning disabled clients on measures of phonological awareness, and on measures of sound/symbol correspondence, word identification, spelling, and decoding in context as well. The fact that it is possible for individuals to successfully develop complex phonemic awareness before even being introduced to basic reading, suggests a causal role for this ability in reading acquisition. Although this does not rule out that phonemic awareness may also be a consequence of reading, it does negate the supposition that it is only a consequence.

Although the ADD method appears to be successful on a one-to-one basis, it may not be a realistic alternative for general classroom methods for teaching reading. It was initially developed for, and is primarily used with the learning disabled population, and the more in-depth description of the program procedure reveals a great deal of one-to-one guided questioning and responding between student and teacher. Furthermore, the ADD Program requires many hours (80) of intensive teacher training before it can be taught, and little is known about using this program as a means of teaching reading in regular classrooms. Although Truch mentions two studies which have tried such implementation (Truch, 1991), neither of the studies is published, and there is not enough information to make an informed judgment.

The studies reviewed to this point provide ample evidence for the importance of instituting some type of phonological training early in a child's school career; however, they do not necessarily support doing this at the expense of specific instruction in reading (e.g., phonics). In fact further research suggests that delaying grapheme/phoneme instruction in order to develop phonological awareness may simply not be necessary for successful reading acquisition -- even for children at risk.

In a phonological training (i.e., specifically phoneme training) study conducted by Byrne and Fielding-Barnsley (1990), the researchers found that for preschool children, "phoneme identity" and segmentation could be easily trained, and that this skill transferred to phonemes on
which children had not been trained. However, it was not until they were trained in letter-sound correspondence that the children could apply this information to a reading task (thus indicating acquisition of the alphabetic principle). Another study by the same authors (1991) focusing on just phoneme identity training (rather than including segmentation as well) for preschoolers resulted in similar findings. Follow-up studies (1993; 1995) found that the children with the phoneme training demonstrated superior performance on nonword decoding 1, 2, and 3 years later. In addition, these children scored significantly higher than controls on reading comprehension at the third-year posttesting. However, there were no differences between groups in spelling performance at the 1- and 2-year follow-ups (spelling was not assessed at the third year).

Although the research focused on early phoneme training, it was evident that these preschoolers could also be taught letter-sounds and apply this learning. The authors noted that both phonemic awareness and letter-sound knowledge needed to be established for the acquisition of the alphabetic principle -- a conclusion supported by their earlier research and other researchers as well (e.g., Ball & Blachman, 1991; Vellutino, 1991).

Bradley and Bryant's research (1983) looked at the consequences of specifically combining phonological training with letter-sound learning. From a large-scale longitudinal study, they examined a subsample of prereaders (5-7 year-olds) who were considered to be at risk -- those with low phonological scores. The authors assigned children to one of four groups: (1) sound categorization -- where children were trained to categorize pictures of animals by shared sounds, (2) sound categorization training plus instruction in the letters which represented these sounds, (3) semantic categorization control group -- where children were trained to categorize pictures semantically, and (4) another control group with no special training. The training groups were instructed in 40 individual 10-minute sessions over a 2-year period. The study demonstrated that although participants with sound categorization training alone (i.e., phoneme and rime identity) performed somewhat better than both controls on reading measures by the end of second grade, results for subjects who received sound categorization training coupled with explicit letter/sound correspondence instruction were far more dramatic (and statistically significant) on both reading and spelling measures. These findings suggest that putting off phonics instruction until
comprehensive phonological awareness training has been initiated is not necessary, and that a combination approach could be a more effective strategy. As successful as this training program was, caution should be exercised in generalizing these findings to the average classroom context. The children who participated in this study received individual training sessions, separate from their classroom programs.

In the Bradley and Bryant's study (1983) reported above, phonological training alone, or phonological training with letter-sound learning was compared. Ball and Blachman (1991) conducted a study with regular kindergarten children to examine how a group with just letter-sound training would compare with a combined phonological training (i.e., phoneme segmentation) and letter-sound group. They questioned whether or not instruction in just letter-sound correspondences would be enough to "heighten phonemic awareness and increase early reading and spelling skills" (p. 54). The interventions (i.e., training) were administered to small groups for 20 minutes, 4 times a week over 7 weeks. The rest of the time, the children attended their regular kindergarten programs (as did a control group). The results showed performance of the children in the combined training group (phonological/letter-sound) was significantly higher than the other two groups on measures of early reading and spelling skills, as well as phonological skills. Although the children with just letter/sound training, performed significantly better than controls (and comparable to the "combined" group) on a measure of letter/sound knowledge, they did not seem to apply this skill to reading or spelling measures, on which they performed at par with the controls. Furthermore, these latter two groups did not differ on the phonological measure. The results provide converging evidence for the importance of a combination approach to teaching beginning reading. Findings also support the notion that instruction in letter names and their sounds is not sufficient to improve reading skills -- that their connection to reading is not explicit enough; but the addition of phonological training (in this case segmentation training) that closely resembles the task of early reading may be enough to help children apply their alphabetic coding knowledge to word analysis. This combination approach may provide the "phonological linkage" (Hatcher & Ellis, 1994) described next.
Aside from Bradley and Bryant (1983) who looked at children with low phonological scores, more research is coming to light (e.g., Foorman et. al., 1998) showing that development of phonemic awareness combined with alphabetic coding instruction is particularly helpful for all types of children considered to be at risk (Hatcher & Ellis, 1994). Hatcher and Ellis designed a study to see whether an intervention that involved phonological training combined with specific reading instruction would be more effective with poor readers than an intervention consisting of either reading instruction alone, or phonological training alone. Four groups of 7-year-old poor readers (roughly corresponding to the lowest 25% of reading skill) were matched on sex, age, IQ and reading ability. These groups were then assigned to one of the three intervention groups, or to a control condition. Each intervention consisted of individual teaching for 40 thirty-minute sessions over a period of 20 weeks. The Phonological Training Alone program consisted of various phonological tasks roughly corresponding to the various levels of difficulty described in the literature (e.g., Stahl & Murray, 1994; Stanovich et al., 1984; Yopp, 1988). The Reading with Phonology program consisted of reading instruction plus phonological activities. The reading portion was loosely modeled on the work of Marie Clay (1985) but with changes which included: more "direct teaching of letter/sound associations, relating spellings to sounds using plastic letters (as advocated by Bradley & Bryant, 1983), and writing words while paying attention to letter-sound relationships" (pp. 47-48) -- activities which the authors termed "phonological linkage" activities. The Reading Alone condition was the same as the latter, except that no explicit reference to phonology or letter-sound relationships (and phonological linkage activities in general) was included. The authors did note that in their regular classes, all children were exposed to phonics reading instruction, but no phonological awareness exercises.

Posttesting toward the end of the school year showed that in spelling, and on every measure of reading, the children in the Reading with Phonology group improved significantly more than did the controls; whereas the children in the other two groups did not outperform the controls (except in one instance -- word recognition -- where the Reading Alone group also showed some significant improvement). On measures of phonological skills, the Phonological Training Alone group made significantly more progress than did the control children. The other
two groups did not improve in this domain more than the controls. At a 9-month follow-up testing, the effects of superior reading for the Reading with Phonology group were still apparent, but on the spelling measure, effects decreased such that there was now no significant difference from the control group. The authors conclude that in order for training in phonology to aid in reading and spelling, it must be connected (linked) explicitly to reading instruction that contains many elements of phonics-based programming (e.g., letter-sound correspondences). They note, however, that their study involved individual tutoring, and this may place constraints on how widely or easily their particular program can be implemented. Modification of some of the exercises for larger groups was suggested.

The above studies once again emphasize the importance and benefits of not only developing at-risk kindergartner's phonological awareness, but also the need to specifically link this to alphabetic coding instruction. Why does this combination seem to be so important for reading acquisition? One very plausible reason is posited by Ehri (1998). She points out that often phonemes do not have a one-to-one mapping onto letters (e.g., "sh" is one sound represented by two letters) and without concurrent instruction in the orthographic representation with phonemic awareness training, the introduction of text (and learning to spell) may be more difficult or confusing for the beginning reader. She suggests that it is "graphophonemic" awareness, rather than just phonemic awareness which should be the focus of beginning reading programs.

Summary

The following points may be a reasonable summary of what much of the research to date appears to suggest: (1) Phonics programs result in better literacy performance than do whole language methods. (2) Explicit systematic teaching of phonics is more successful than implicit, indirect (embedded) phonics instruction. (3) The addition to systematic phonics programs of the meaningful, motivating reading aspects of whole language may enhance reading performance even more. (4) Initial spelling-sound correspondences are best taught to beginning readers at the
phoneme/grapheme level. (5) Phonological training tends to facilitate reading acquisition, but its effects are more powerful when coupled with instruction in alphabetic coding. (6) All of these points appear to hold true for, and may be even more applicable to, children considered to be at risk for reading failure.

It is not so clear, however, what specific roles these factors play in kindergarten. Phonemic awareness development is clearly important and can be started at the kindergarten level. However, opinions differ as to whether alphabetic coding should be started at kindergarten entry as well. Since children come to kindergarten with such a broad range of home (preschool) literacy experience, and some may not have even the basic rudiments of letter or book knowledge (Adams, 1990), many researchers feel that kindergarten time is best spent developing these readiness skills in order to prepare children for more formal training (Sacks & Mergendoller, 1997; Stahl & Miller, 1989, Stahl et al., 1994). Although the Sacks and Mergendoller study showed that programs which provide much literacy experience and instill concepts of print go far in preparing children in these necessary reading readiness skills, conclusions cannot be drawn about whether or not this is better than including alphabetic coding instruction as well (since explicit letter-sound instruction did not appear to be part of the phonics-oriented teachers' programs in this study).

Meta analyses of the reading research literature (Stahl & Miller, 1989, Stahl et al., 1994) do seem to indicate that whole language may be more effective in kindergarten than basal reading programs, particularly for at-risk children (due to a literacy impoverished home environment) at this age. However, as mentioned earlier, since most studies did not include classroom observations, it is difficult to determine what was actually included in each of the instructional approaches being compared. Furthermore, it seems that mere exposure to print and implicit literacy experiences for a few hours a day during the school year, would not go far enough to make up for the hundreds or thousands of lost literacy hours (Adams, 1990) for many at-risk children entering kindergarten. For children with very little literacy experience, explicit teaching of letters and sounds would likely give them a chance to "catch up", in some respects, to their peers who have had years of reading-related experience, and thus, have had many more hours of opportunity to pick up at least an implicit understanding of the alphabetic code.
Rationale for the Current Research

As indicated in the review section, an enormous amount of reading research has been amassed which addresses varying aspects of early reading acquisition. And although trends are clearly evident, hard and fast conclusions about most issues cannot yet be drawn, particularly for the kindergarten level. Research needs replication and converging evidence from many sources (Stanovich, 1992) for conclusions to move from the realm of speculation and trends, toward becoming fact.

The research to date can be broadly grouped into three types: (1) scientific/laboratory experiments, (2) research examining experimenter-designed training programs for either classroom implementation or one-to-one tutoring (usually targeting a particular process and or particular children), (3) classroom research comparing existing programs by theoretical approach (or comparing good and poor readers within the same approach).

Laboratory research has been extremely profitable theoretically, and has had many practical implications. However since it is conducted under unnaturalistic conditions, generalizability to the typical classroom environment is questionable. As a result, some of these implication have been field tested, in the form of experimenter-designed training programs, within a more naturalistic school atmosphere. However, even some of these studies have been artificial in nature with special one-to-one student "training sessions" set aside from the regular classroom activity, and/or have targeted only at-risk populations (e.g., Brady et al. 1994). There have been some very comprehensive, in depth studies examining instructional/training programs which were integrated into regular classroom programming for all students (e.g., Foorman et al., 1998); and results from this research have helped determine the efficacy of various instructional components. However, the experimental programs under investigation have often been very complex and required a great deal of initial teacher coaching (e.g., 30 hours in the Foorman et. al., 1998 study) with follow-up retraining, and regular on-site consultation. Although important results have been gleaned from these classroom studies, the complexity and amount of teacher training for implementation of the actual methods that were designed and examined seems to prohibit their widespread use as regular
programming. Studies comparing existing classroom teaching approaches, although obviously more applicable to general classroom implementation, have rarely included observations (with a few notable exceptions -- e.g., Evans & Carr, 1985). This lack of classroom monitoring leads to dubious conclusions about the relative effectiveness of the programs being compared since there is no way of determining whether a particular theoretical approach is actually being implemented. As the observations of the Sacks and Mergendoller study showed (1997), teachers with self-proclaimed whole language or phonics orientations actually conducted similar classroom programs (although a few subtle differences were reported).

There have been a few influential studies in classroom observational research; however, these studies mainly examined academic achievement in general across grades (rather than specific focus on early reading instruction), and were concerned more with the process of teaching rather than the content of the programs (e.g., Brophy & Evertson, 1978; Stallings, 1980; Stallings & Krasavage, 1986). Although these studies have provided valuable information regarding: lesson delivery, teaching style and teacher/child interaction, classroom management and context variables, and importance of time-on-task, they have shed little light on actual program content, particularly with regard to reading instruction. They have, however, pointed to the effectiveness of direct, explicit instruction in any subject domain.

In response to the experimental research described above, early intervention programs focusing on many of the important reading components (e.g., Reading Recovery, Success for All) have been developed, and are becoming widely implemented (Clay, 1985; Pinnell, DeFord, & Lyons, 1988; Slavin, Madden, Karweit, Dolan, & Wasik, 1991). These too have been subjected to research and have reported considerable success (e.g., Clay, 1985; Pinnell, Lyons, De Ford, Bryk, & Seltzer, 1994), although some of the evaluation studies are not without problems (e.g., Nicholson, 1989; Robinson, 1989; Shanahan & Barr, 1995). However, these programs are intervention approaches which often deal with the "casualties" of less-than-sufficient regular classroom programming. They focus on intensive small group or individual remediation; are "pull-out" in structure (i.e., removing children from their regular classroom), and can be extremely expensive to implement. They are reactive, not proactive. It appears that what is still needed is
research on a reading approach which will prevent, or at least reduce, the need for intensive and expensive intervention programs -- a commercially available reading program that not only speaks to the researched-based literacy components that have been found to be important for reading acquisition, but also one that is designed for regular classrooms comprised of a wide range of differently-skilled children.

There are, of course, numerous commercial basal reading programs currently in use, but evaluations of kindergarten classroom implementation have simply not been conducted as reported by members of the Committee of the Prevention of Reading Difficulties in Young Children in their recent comprehensive compendium synthesizing research on early reading development (Snow et. al., 1998). These packages usually include all sorts of activities and materials (e.g., games, big books, posters, audio tapes, workbooks, etc.) as well as instruction manuals with lesson plans; and they are greatly influenced by market research with regard to which components to include and emphasize. Stein et. al. (1993: in Snow, et. al., 1998) undertook an analysis of the major reading programs available, which showed that most packages provide activities incorporating many of the important early reading components already discussed. However, analysis of the four best-selling programs (Simmons, 1994) found that activities involving one of the most important components -- phonological awareness-- were limited in quantity and quality, with little emphasis placed on phonemic units of speech. Furthermore, the authors point out that these program analyses do not elucidate what teachers actually do with these programs. Since these market-driven programs try to include optional activities and alternatives to instruction (e.g., student self-selection) in order to cover all theoretical bases, "except in the hands of the most competent teachers, each of these strategies carries its own variety of risks to classroom order and instructional coverage" (Snow, et. al., 1998, p. 192). These programs are usually publisher developed, not teacher developed, and since financial success is a guiding incentive, development often follows a route of mass production rather than one of research and liaison with the eventual consumer of the product -- the classroom teacher. As Ball and Cohen (1996) point out, programs developed with teacher input and feedback could result in a much better match between publisher's ideals and what really happens in the day-to-day classroom context.
So, on the practical side of early reading instruction, there have been reviews on commercial educational offerings (e.g., Simmons, 1994) examining the content of the currently available basal programs. However, these have not included investigations into how reading packages translate into classroom implementation. On the theoretical/empirical side of early reading research, there have been several high quality investigations of many pieces of the reading instruction puzzle, but it appears none has examined an instructional approach which has put all these pieces together, -- one that is formatted in such a way that it is relatively easy to implement and integrate into existing programs, provides instructional guidance along with theoretical reasoning, is geared to a typical kindergarten classroom (as opposed to individual tutoring), is developed with teacher input, and is also commercially available. As reported by the members of the Committee of the Prevention of Reading Difficulties in Young Children (Snow, et. al., 1998) with regard to research involving commercially available basal reading programs at the kindergarten level, they have yet to find any "objective, empirically sound evaluation of major kindergarten offerings", and they go on to say "such evaluation should be a priority for public policy" (p.194).

The current study offers a comprehensive evaluation of a teaching program for beginning reading in kindergarten (Jolly Phonics, Lloyd, 1993). Jolly Phonics is not a basal reading program, but it is a commercially available instructional approach that is meant to be integrated into a classroom's existing program. It includes all (or most) of the literacy puzzle pieces. That is, its goals and objectives reflect current research findings on early reading acquisition -- the emphasis of phonological awareness combined with explicit, systematic letter-sound correspondence instruction. This study not only assessed the efficacy of the Jolly Phonics method (by comparing with controls, achievement on reading/spelling measures of children receiving the Jolly Phonics instruction), but through on-line monitoring of program implementation and content, the current study was also designed to explore the relation between program components and successful reading/spelling performance.
Research Objectives

Through comparisons with control classrooms, the current study was designed to investigate practical questions regarding alphabetic coding, phonological awareness, and reading instruction within the context of the commercially available Jolly Phonics teaching program. The first objective was to see whether or not kindergarten children in the Jolly Phonics program (which included explicit instruction in these components) performed better than did controls (who were exposed to more traditional kindergarten programming) on end-of-year nonstandardized measures of reading, spelling, reading readiness, and word analysis, and on standardized measures of reading, spelling, word analysis, and phonemic awareness. These tests measure not only how well children have been able to learn letter/sound correspondences within their respective programs, but also whether or not they are able to apply this knowledge to word analysis and synthesis in reading and spelling.

The second objective of this research was to compare the experimental and control classrooms with regard to time spent on various literacy activities (which included different key reading components). Because program implementation can vary greatly due to any particular teacher's style, interest, and philosophy, systematic time-sampling observations made it possible to determine what was actually going on in the classrooms under investigation. In this way, not only could the experimental program be monitored in terms of program component emphases, but the various activities making up the control programs could be examined as well.

The third objective was to see how important these components are to the acquisition of early reading (and spelling) skills. By examining correlations of time spent on these various activities with performance on outcome measures, and by using multiple regression methods, it was possible to determine which literacy components best predicted reading and spelling performance.

The final objective of this study addressed concerns related to proportion of the population of children considered to be at risk for reading failure. Researchers have suggested that children with low levels of reading readiness (e.g., alphabet knowledge) may profit more from a literature-
rich language-experience approach since these children may not be "developmentally prepared" to receive explicit alphabetic coding instruction (e.g., Sacks & Mergendoller, 1997; Stahl et. al., 1994). However, this study examined whether or not at-risk children benefited from an instructional approach which included explicit phonics (and word analysis) combined with an emphasis on phonological awareness in addition to the conventional kindergarten fare of rich language/literature experiences.

**Research Questions**

Reflecting the four main objectives, the research questions were as follows:

1. Do participants who received the Jolly Phonics teaching approach perform significantly better than controls on measures of prereading skill (letter recognition and letter writing), phonological awareness (phoneme deletion), and phoneme/grapheme knowledge (letter-sound recognition, letter-sound recall)? If so, does this superiority in performance translate into significantly higher scores on measures reflecting application of these skills -- word recognition, spelling, word analysis (nonword reading and spelling)?

2. Do the Jolly Phonics classrooms spend more time involved in activities which focus on the key components found in the research to be strongly associated with reading acquisition (alphabetic coding, phonics, phonemic awareness)? If so, is this accomplished at the expense of time spent on literacy components that are thought to be important elements in kindergarten programming (e.g., language experience, concepts of print, writing experiences)?

3. For all classes, does the time spent involved in the various literacy activities (phonics, phonemic awareness development, reading, writing, etc.) predict scores on the outcome measures? Specifically:
   a) How much variance in the outcome scores does the phonics component explain.
   b) After phonics, what is the next most significant predictor(s)?
c) How much additional variance does classroom time spent on development of auditory phonological awareness explain?

4. Do at-risk children who received the Jolly Phonics teaching approach perform significantly better than their at-risk counterparts in the control classes on measures of prereading skill (letter recognition and letter writing), phonological awareness (phoneme deletion), and phoneme/grapheme knowledge (letter-sound recognition, letter-sound recall)? If so, does this superiority in performance translate into significantly higher scores on measures reflecting application of these skills -- word recognition, spelling, word analysis (nonword reading and spelling)?

---

3 In the observations for this study, two different aspects of the phonological awareness component were placed in two different categories. Activities which were totally auditory in nature (i.e., without the influence or enhancement of printed letters or text) were recorded in the "Auditory Phonological Awareness" category; and activities which included print when phonemes or other phonological units were being emphasized were recorded in the "Phonics" category. Therefore, since the Phonics category contained this aspect of phonological awareness, as well as a number of other activities, it was considered to be the most likely predictor, and thus, was chosen to be entered first in hierarchical regression analyses.
METHOD
Overview

The study was undertaken in three stages. There were two testing phases separated by a treatment period. Phase 1 was a pretest phase involving the assessment of experimental (Jolly Phonics) and control children's phonemic awareness and literacy skills approximately midway through the school year to determine participant characteristics and ensure group comparability. The treatment period consisted of implementation of the Jolly Phonics program and monitoring (through a time-sampling observation procedure) of classroom activities in both the experimental and control classrooms. Phase 2 testing was the posttest phase, the results of which are the main focus of the analyses described in the Results and Discussion section. For Phase 2, participants were posttested on the same (Phase 1) pretest measures, and on a battery of additional end-of-year measures. Table 1 presents a brief overview of the study in the form of a time line.

Table 1.
Overview of Study

<table>
<thead>
<tr>
<th>Phase 1 Testing</th>
<th>Treatment</th>
<th>Phase 2 Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Measures (for equivalency of groups)</td>
<td>Program Implementation and Time-Sampling Observations (Experimental and Control Classrooms)</td>
<td>Posttest Measures (for assessing differences)</td>
</tr>
<tr>
<td>Mid Year</td>
<td>Spring</td>
<td>End of Year</td>
</tr>
</tbody>
</table>

Participant characteristics are first presented, along with Phase 1 results examining group comparability. Details of the study's time-line, testing measures, scoring, and procedures are then

---

One reason for waiting until midyear before initial testing was the fact that teachers wanted a few months to socialize the children to the school environment and instill classroom routines. It was also believed that this development of familiarity with new (school) experiences would help these young children to feel more at ease with the pretesting procedures.
described (Phase 1 and 2). Next, the Jolly Phonics program, teacher training, and program implementation is outlined, as well a brief description of control classroom programming. The observation protocol is then presented, followed by observer training, reliability, and the time-sampling observation procedure.

**Participant Characteristics and Group Comparability**

**Participants**

Children were drawn from 10 experimental and 10 control kindergarten classes from eight schools (with comparable ranges of socio-economic status -- mainly lower-middle class) in suburban areas outside of metropolitan Toronto. This area encompassed a very large region; therefore, the schools making up the experimental group were far removed from those in the control, and experimental and control teachers had no contact with each other. This obviated some of the pitfalls which can arise in educational treatment studies, such as: experimental and control teachers sharing materials, information, and or techniques (Borg, Gall, & Gall, 1993); or control teachers hearing/seeing the effectiveness of the experimental program and therefore either tending to "give up" on the one they are implementing, or being spurred on to try even harder than they might under a more "non-competitive" atmosphere (i.e., "John Henry effects", Borg & Gall, 1989, p. 191).

Teachers in the experimental classes had volunteered to try the Jolly Phonics program when hearing about it either through a teacher development workshop or from other teachers who had attended. Experimental classes integrated the Jolly Phonics program into their regular curricula (i.e., a traditional kindergarten focus of play, language, reading readiness, and social development). Control schools described their curricula as balanced literacy programs. Subsequent observations from this study revealed that control programs were similar in nature to those of the experimental classes, minus the integration of Jolly Phonics. That is, they emphasized play, socialization, language development, and reading readiness skills. In both the control and
experimental classes, the degree to which each of these domains was emphasized, varied considerably. Some of the teachers taught two classes (i.e., morning and afternoon kindergartens), resulting in eight different teachers in the Jolly Phonics group and seven, in the control. In the 10 experimental and 10 control classrooms there were 151 Jolly Phonics and 130 control participants, totaling 281 children. After initial data screening for atypical participants (e.g., gifted or developmentally challenged designation, or extremely high pretest word recognition scores suggesting children were from a different population), 265 children remained (145 experimental and 120 control).

The top of Table 2 presents participant characteristics. There was no difference ($t[1, 263] = .57, p = .574$) between the mean age for the Jolly Phonics children (5.91, SD = .27) and control children (5.93 years, SD = .31) at the end of the year testing period. Nor was there a significantly different female to male ratio (Jolly Phonics = 74 females and 71 males; control = 51 females and 69 males), $\chi^2(1, N = 265) = 1.92, p = .115$. The number of children whose first language was not English (ESL) was significantly higher in the Jolly Phonics than in the control classes (ESL:notESL ratio: Jolly Phonics = 49:96; control = 22:98), $\chi^2(1, N = 265) = 8.0, p < .01$, suggesting that outcome testing of the Jolly Phonics program effectiveness might be conservative.

**Group Comparability**

As described in the Phase 1 Measures and Procedures section, children were pretested on phonemic awareness and early literacy skills prior to program implementation. There was a large range in levels of literacy skills in both the experimental and control groups. Table 2 shows that the mean score for the control children on each of the pretest measures was slightly higher than that for the experimental participants (see Measures and Procedures section for description of measures). This was to be expected since due to time constraints and logistical reasons in securing permission for the control classes, they were tested slightly later in the school year (as described later). These differences were not significant, as demonstrated by a MANOVA computed for this group of measures, $F(5, 259) = 1.330, p = .252$), and therefore the only measure whose univariate ANOVA had a $p$ value <.05 (i.e., Letter-Sound Recall, $F[1, 263] = 4.34, p < .05$)
cannot be considered significant. The Jolly Phonics mean for this measure was 4.48 (SD = 8.69), slightly lower than the mean of 6.1 (SD = 7.0) for the control group.

Table 2.

Participant Characteristics and Group Comparability at Phase 1 Pretest

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 145)</th>
<th>Control (n = 120)</th>
<th>$\chi^2$</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>74 girls</td>
<td>51 girls</td>
<td>1.92</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>71 boys</td>
<td>69 boys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESL</td>
<td>49 (34%)</td>
<td>22 (18%)</td>
<td>8.00</td>
<td>.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics</th>
<th>Control</th>
<th>t</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>5.91 0.27</td>
<td>5.93 0.31</td>
<td>0.57</td>
<td>ns</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics</th>
<th>Control</th>
<th>F</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAAS</td>
<td>2.21 2.21</td>
<td>2.72 2.21</td>
<td>3.52</td>
<td>ns</td>
</tr>
<tr>
<td>LtNm</td>
<td>16.35 8.69</td>
<td>17.82 7.39</td>
<td>2.15</td>
<td>ns</td>
</tr>
<tr>
<td>L/SRecog</td>
<td>5.57 6.30</td>
<td>7.04 6.91</td>
<td>3.31</td>
<td>ns</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>4.48 5.63</td>
<td>6.10 7.01</td>
<td>4.34</td>
<td>.05*</td>
</tr>
<tr>
<td>B&amp;R</td>
<td>0.78 2.04</td>
<td>0.90 2.12</td>
<td>0.22</td>
<td>ns</td>
</tr>
</tbody>
</table>

MANOVA $F(5, 259) = 1.33, p = .252$

*Note: Since the MANOVA was not significant, this difference cannot be considered significant.

Time-Line

As mentioned in the overview, there were two testing phases separated by a period where the treatment was implemented in the experimental classrooms, and, for both groups (i.e., Jolly Phonics and control), the classroom observations took place. Table 3 presents the study's overall design, along with the measures administered in each testing phase. Phase 1 pretesting was conducted to determine if the children in the experimental (Jolly Phonics) and control groups were of a similar reading, spelling, and reading-readiness skill level, as well as being comparable in their level of phonological awareness. This testing took place approximately mid-way through the school year. Teachers in the experimental group were to commence implementation of the Jolly Phonics reading program in January, and so Phase 1 testing for the Jolly Phonics kindergarten children was completed just prior to Christmas vacation in December. Due to time constraints, and delays obtaining permission from the Boards of Education for participation of the control schools, Phase 1 testing for the control group began late January and was completed prior to the end of February. Part of the intervening period was non-school time (i.e., winter vacation).

Since the Jolly Phonics group was pretested earlier than the control group, their Phase 2 posttesting (on those measures that were administered in Phase 1) started earlier as well, and was conducted in the month of April. All of the other Phase 2 testing (i.e., the additional measures which had not been included at Phase 1 pretest) was conducted at the very end of the school year (during the month of June). For the control group, all posttesting (Phase 1 measures and additional Phase 2 measures) was conducted in June. Classroom observations (three separate observation days per class) took place during March, April, and May (see Table 3).

---

5 Note that since the earlier "end-of-year" testing time (i.e., April) on the Phase 1 portion of the outcome measures for the experimental group would favor the control group (who were tested in June), this slightly different testing time-table for the two groups made the experimental and control comparisons (on the posttest set of Phase 1 measures) a conservative test.
### Table 3. Time-Line Design

#### Phase 1 Testing

<table>
<thead>
<tr>
<th>Nov/Dec</th>
<th>Jolly Phonics</th>
<th>TAAS</th>
<th>Jan/Feb</th>
<th>Jolly Phonics</th>
<th>Apr</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

#### Classroom Observations

<table>
<thead>
<tr>
<th>Nov/Dec</th>
<th>OBS1</th>
<th>OBS2</th>
<th>OBS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov/Feb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar/May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr/Jun</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Phase 2 Testing

<table>
<thead>
<tr>
<th>Nov/Dec</th>
<th>Jolly Phonics</th>
<th>TAAS</th>
<th>Jan/Feb</th>
<th>Jolly Phonics</th>
<th>Apr</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

#### Control

<table>
<thead>
<tr>
<th>Nov/Dec</th>
<th>OBS1</th>
<th>OBS2</th>
<th>OBS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov/Feb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar/May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr/Jun</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Key:** Next Page
Table 4.

Key to Abbreviations

<table>
<thead>
<tr>
<th>PHASE 1 MEASURES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TAAS</td>
<td>Rosner Test of Auditory Analysis Skills</td>
</tr>
<tr>
<td>LtNm</td>
<td>Letter Name Task</td>
</tr>
<tr>
<td>WrtAlph</td>
<td>Write Alphabet</td>
</tr>
<tr>
<td>L/SRecog</td>
<td>Letter-Sound Recognition Task</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>Letter-Sound Recall Task</td>
</tr>
<tr>
<td>B&amp;R</td>
<td>Burns and Roe Word Recognition</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADDITIONAL PHASE 2 MEASURES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WrRdSS</td>
<td>WRAT Reading Standard Scores</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>WRAT Spelling Standard Scores</td>
</tr>
<tr>
<td>WaSS</td>
<td>Woodcock Word Attack Standard Scores</td>
</tr>
<tr>
<td>NwRd</td>
<td>Nonword Reading Task</td>
</tr>
<tr>
<td>NwSpPh</td>
<td>Nonword Spelling Task -- Phonemic Analysis</td>
</tr>
</tbody>
</table>

Test Measures and Procedures

The Phase 1 pretest measures were administered to see if the Jolly Phonics and control groups had comparable levels of literacy/preliteracy skills and phonological awareness. The second (posttest) phase of testing included these measures along with other outcome measures to assess various reading and spelling skills (see Phase 2 measures) for a total of 11 tasks administered. For purposes of analyses, some of the additional outcome measures were scored two or three different ways (see Phase 2 Testing Procedures and Scoring), to provide a more complete examination of children's literacy knowledge and abilities (resulting in 19 sets of scores for Phase 2 analyses). A copy of all the nonstandardized test measures, examiner instructions, and alternate scoring procedures can be found in Appendix A.
Phase 1 Test Measures

Phonological Awareness

Rosner's Test of Auditory Analysis Skills (TAAS; Rosner & Simon, 1971)

The TAAS assesses mainly a "complex level" of phonological awareness (Yopp, 1988). After the first few items requiring the segmenting of words into syllables, task requirements increase in difficulty, as participants must perform various phoneme manipulations (e.g., splitting blends or deleting initial, medial, or final sounds). Raw scores range from 0 to 13.

Rhyme Task (dropped from study)

In order to tap a "basic level" of phonological awareness (Yopp, 1988), an experimental Rhyme Task had been designed and was administered at Phase 1 and 2 testing times. However, this measure was dropped from the analyses since several of these kindergarten children did not understand the requirements and simply answered in a response set (i.e., yes/no/yes/no...etc.).

Basic Literacy/Readiness Skills

Recite Alphabet and Write Name (dropped from study)

To put children at ease, they were asked to recite the alphabet. Since, most children reached ceiling, scores on this measure were not included in analyses. They were also asked to write their names. This too was not included in analyses.

Writing Alphabet Task (WrtAlph)

WrtAlph was an experimental task designed to assess how many letters children could write (print) in 60 seconds. The task was administered at both Phase 1 and Phase 2 testing times. However, results of the Phase 1 administration have not been used in any analyses because some classes had missed pretesting on this measure (due to logistical reasons of timing).

Letter Name Task (LtNm)

LtNm was an experimental task designed to assess how many letters the children could recognize. The 26 letters of the alphabet were printed in a simple lower case font (i.e., Geneva, 48 point) on an 8" x 11" card, 6 or 7 letters to a line (black print on white background). The order of the letters was in a non-alphabetical sequence (the same sequence, except for the letter "1", as
employed by the Jolly Phonics program for letter introduction -- see Appendix B). Highest possible score was 26.

*Complex Literacy Skills (Alphabetic Coding)*

**Letter-Sound Recognition (L/SRecog)**

L/SRecog was an experimental task designed to assess children’s recognition of letter sounds. It requires the child to point to the appropriate letter in response to a letter-sound presented to them (in a non-alphabetical sequence) by the examiner. All 26 letters were printed in alphabetical order (in the same font as the other measures) on a 4.5" x 12" card. Highest possible score was 26.

**Letter-Sound Recall (L/SRecall)**

L/SRecall was an experimental task designed to assess children’s ability to produce (recall) the sound made by a visually presented letter(s). Items included the 26 letter sounds as well as 17 two-letter phonemes (i.e., digraphs and diphthongs) printed (in the same font and case as the other tasks) on two 8" x 11" cards. This resulted in a total of 43 correspondences for the 42 phonemes (two different letters have the same sound -- "c" and "k"). The order of the phonemes was in a non-alphabetical sequence (the same sequence as the LtNm Task). Highest possible score was 43, which included phonemes having two different pronunciations (e.g., "oo" as in moon or book, -- see Testing and Scoring section for details); however, for final scoring, one of the phonemes was dropped ("ng") due to inconsistent scoring by the examiners, and so the highest possible score for the task in this study was 42.

*Real Word Recognition*

**Burns and Roe Word Recognition List (B&R; Burns & Roe, 1993):**

The B&R word lists (from the Burns and Roe Informal Reading Inventory, 1993) are graded from a preprimer to 12th grade level. Each level consists of 20 words. For the Phase 1 testing, only the preprimer list was administered, but for Phase 2 testing, as many lists as required were used until each child met a discontinue criterion (7 errors out of the 20 words at a particular
level). Two alternative lists (List 1 and List 2) are available for each level. For this study, the List 1 graded word lists were used. Words were presented on 4.25" x 12" cards (one card per level) listed in a column in the same font as in the inventory (Times, 18 point).

**Phase 1 Procedures: Testing and Scoring**

The Phase 1 measures were administered to each child individually by research assistants in a uniform manner. To keep testing sessions short for these very young children, Phase 1 testing for each child was completed in two separate sessions on two separate days. One session (Session A) included the Rosner Test of Auditory Analysis Skills (TAAS) and the Rhyming task (which is not described since it was dropped from the analysis). The second session (Session B) included the remainder of the Phase 1 tests.

**Session A**

The Rhyming task was administered first, followed by the TAAS. The TAAS was administered according to test requirements. This involved two trial demonstration words, followed by test items. The child was asked to repeat a word spoken by the examiner, and then repeat it again but without saying a particular phoneme (e.g., saying "fish" and then saying it again without the /f/). The test was discontinued after the child made two consecutive errors. The child's score was recorded as the item number prior to the two errors (consistent with TAAS instructions).

**Session B**

In session B, the Recite Alphabet task (dropped from study) was first administered. Writing Name (also dropped) and Writing the Alphabet task were presented next. Children were given a sheet of lined paper and asked to write their name on the top. Then children were asked to write the letters of the alphabet on the page, or as many of the letters as they knew how to write. After 60 second (unobtrusively timed with a stop-watch), children were asked to stop. If children clearly had finished before the time was up, the task was discontinued sooner. Scores in this...
study's analysis were the total number of "unique" letters that the child was able to write within the 60 seconds. That is, if a child wrote nothing but "a's" on the paper, the score would be 1. Upper and lower case representations of the same letter were only given a single point.

The Letter Name task was then administered. Participants were shown the card with all the letters of the alphabet represented out of order (for description of materials see Test Measures). Children were asked to point to, and name each of the letters as the examiner revealed one row at a time (by sliding a cover sheet down the page). Approximately four seconds response time was allotted before moving to the next letter. After five consecutive errors, the examiner revealed the whole card and asked if there were any other letters on the card that the child might know. The child's score equaled the number of correctly named letters (out of 26).

The Burns and Roe Word Recognition task was administered next. For the Phase 1 pretest, children were shown a card with the twenty-word preprimer list. In keeping with the Burns and Roe format, the words were listed vertically and were shown one-at-a-time by moving a cover-card down the list. Five seconds were allowed for a response before moving on to the next word. Children were asked if they could read the word shown and after seven errors (not necessarily consecutive) the remainder of the card was revealed and children were asked if there were any other words on the list that they knew how to read. The child's score equaled the number of correctly read words minus the first word "a." This was the only word that the majority of the children "read correctly" at pretest, and it was clear that they were usually reading it as a letter name, and not a word. To remove this confound, the score for "a" was dropped for all the pretest and posttest scores; thus in pretest, the highest possible score a child could receive was 19 (instead of 20).

The Letter-Sound Recognition Task was then administered. A card with the lower-case alphabet was placed in front of each participant. Children were told that the examiner was going to make a sound and that they were to point to the letter that makes that sound. Since reading requires visual recognition of the letters that correspond to particular sounds, it was decided that only a pointing response (rather than verbal) would be accepted. This was to ensure that the children recognized the visual representations of these sounds. To make the time required for this test
manageable, and to alleviate any discomfort the children may have felt resulting from many failed responses, a discontinue rule was invoked so that testing was stopped after six consecutive errors. The child's score equaled the number of correct letter-sound responses (out of a possible 26).

Following the Letter-Sound Recognition Task, the Letter-Sound Recall Task was administered. This included the 26 letter sounds as well as 17 two-letter phonemes (i.e., digraphs and diphthongs). Printed phonemes on cards were revealed to the children one row at a time. The examiner pointed to each of the phonemes and asked “What sound does this letter make? What does it say?”, allowing five seconds for a response. If children did not respond within five seconds, the examiner went on to the next phoneme. When showing a digraph (e.g., "sh"), the examiner said, "Sometimes when two letters are stuck together like this, they make one sound. Do you know what sound these two letters make when they are stuck together like this?" After 6 consecutive errors, the rest of the phonemes were revealed all at once and children would be asked if they knew any other sounds that these letters might make. The child's score equaled the number of correct letters-sound responses. As mentioned in the description of measures, one of the phonemes was dropped ("ng") due to inconsistent scoring by the examiners, and so the highest possible score for the task in this study was 42. A more detailed explanation of instructions and scoring is found in Appendix A.

Phase 2 Test Measures

The same measures as were used in the Phase 1 testing were administered again at posttesting. In addition, the following measures were used:

**Real Word Reading and Spelling**

**WRAT Reading and WRAT Spelling subtests** (WrRdSS and WrSpSS; Wilkinson, 1993)

The WrRdSS and WrSpSS (subtests of the Wide Range Achievement Test: 3rd Revision) are standardized measures assessing word recognition and spelling. For kindergarten-age children, both the reading and spelling subtests start with individual letters (i.e., naming/printing letters), before full words are introduced. The WRAT3 provides absolute scores, standard scores,
and grade scores. It was normed on a stratified sample of individuals aged 5 to 75. Reliability range from .92 to .95 using test coefficient alpha. It has two alternate forms. The Blue Form was used for this study.

**Nonword Reading and Spelling**

**Woodcock Word Attack** subtest (WASS; Woodcock, 1987)

The WASS (subtest of the Woodcock Reading Mastery Test - Revised -- WRMT-R) is a standardized measure assessing decoding ability (i.e., applied phonics and structural analysis). This test requires the child to decode nonwords (pseudowords). In order to receive a score, the pronunciation must be completely blended and totally correct. Initial items are short words, mainly vowel-consonant (VC) or CVC. The difficulty level quickly progresses to multisyllabic pseudowords incorporating digraphs and complex blends. The kindergarten children, in this study, often met the discontinue criterion of 6 consecutive errors almost immediately. The WRMT-R was normed on a stratified sample of individuals from kindergarten to Grade 12, college, and adults. Split-half reliability estimates are reported for alternate grades starting with Grade 1. Median reliability coefficient is .87. The WRMT-R has two alternate forms. Form G was used for this study.

**Nonword Recognition Task** (dropped)

The Nonword Recognition Task was designed to see if children could recognize (i.e., choose between two similar nonwords) a nonword when spoken by the examiner. However, this measure was dropped from the analyses since several of these kindergarten children did not understand the requirements and simply answered in a response set (i.e., yes/no/yes/no...etc.).

**Nonword Reading Task** (NwRd)

NwRd was an experimental task with less complex nonwords (24 nonwords) designed to obtain a fuller picture of children's word analysis skills. To ensure that children had a greater opportunity to demonstrate what they knew, the discontinue rule for this measure was 10 nonconsecutive mistakes, thus allowing participants a chance to come across some orthographic pattern(s) they might know. Since nonwords on this measure incorporated many of the digraphs
introduced in the Jolly Phonics program, this task could also assess whether the children in the Jolly Phonics classes were able to apply what they were taught.

**Nonword Spelling Task – Phonemic Representation (NwSpPh)**

NwSpPh was an experimental task designed to assess children's understanding of the alphabetic principle through their phonemic representations of nonwords. Since items on the WRAT Spelling subtest are comprised of real words (and letters), it cannot be determined whether children understand the letter-sound mappings when they produce a correct spelling, or if they are simply drawing on the whole word memory in their sight-word lexicon. To generate an appropriate spelling of a nonword, children must rely solely on their alphabetic coding knowledge. This task consisted of 22 nonwords (primarily one syllable) comprised of the 42 phonemes. Examples of these words are "sim, gad, thorp". Because the correct spelling of nonwords requires the "hearing" of their component phonemes, this task was considered a further measure of children's phonological awareness.

**Phase 2 Procedures: Testing and Scoring**

There were 3 separate individual testing sessions as well as 1 session where the spelling measures were group administered. Due to time constraints, or requirements of "clean-up" testing (i.e., administering tests to children who had missed testing for various reasons) sometimes test session were combined, or the order of administration differed. For the most part, however, the following procedures were followed.

*Session A*

Session A included the same measures and procedures as (described previously) in Phase 1 pretesting.

*Session B*

Session B was also conducted in the same manner as described in Phase 1 except for one change. For the Burns and Roe Word Recognition task, additional words were included. If
children completed the preprimer list of 20 words without accumulating seven errors, they would go on to additional twenty-word lists (primer to grade 8) until they reached a ceiling of seven (non-consecutive) errors in a particular list. The child's score equaled the number of words read correctly up until the discontinue criterion was met (minus the first word "a" as described in Phase 1 testing).

**Session C**

Session C included the additional end-of-year tests that had to be individually administered. First, the WRAT Reading subtest was administered, followed by the Woodcock Word Attack subtest. These were both administered in the standardized manner. On the score sheet, the examiner wrote down approximations of the child's response (if not correct). Both tests were scored in the standardized manner.

The experimentally designed Nonword Tasks were then administered. The Nonword Recognition Task (dropped from the study) was first, followed by the Nonword Reading Task. This test was explained in more detail to the child than was the Woodcock Word Attack test (which necessitated following standardized instructions). The child was reminded that letters make sounds, and that sounds can be put together to make words (see Appendix A for detailed instructions). The examiner demonstrated the first visually presented trial word, and then helped the child sound out the second trial word. The test words were then presented, with the examiner uncovering 1 word at-a-time. Up to ten seconds were allowed prior to uncovering the next word. The examiner wrote down approximations of the child's response (if not correct). The test was discontinued when the child made 10 non-consecutive errors. There were 24 items. The child's score equaled the number of words read correctly up to the discontinue point.

**Session D**

Session D was comprised of two spelling tests which were group administered. Because kindergarten children required much supervision, the tests were administered in small groups of 4
or 5. The Spelling subtest from the WRAT3 was administered first, and scored in the standardized manner.

The second group-administered task was the experimental Nonword Spelling Task -- Phonemic Analysis, consisting of 22 nonwords. It was administered in much the same way as the WRAT3 test with the exception that there was no discontinue rule, with all 22 nonwords being administered (see Appendix A, Nonword Spelling Task for details). Children were encouraged to put down any sounds that they might recognize.

Scoring for the Nonword Spelling task -- Phonemic Representations was more complex. This task was scored according to the degree of phonemic representation found in children's responses (a similar alternate phonemic method of scoring for several of the Phase 1 measures is described later). All reasonable phonemic representations in this Nonword Spelling Task were considered correct. For example, the following would receive a full score for the spelling of the nonword "jope": joap, joep, goap, gope, goep. Leniency was exercised, because at the kindergarten level, children most likely would have not yet learned any of the more advanced spelling rules (e.g., hard and soft "g"). Scoring for this task involved a total of four possible scores for each word. If the word was spelled completely correctly, it would receive 3 points. If there were at least two (but not the complete word) correct phonemes (in the proper order) it would be worth 2 points. The initial phoneme correct would receive 1 point, and a blank, or no correct phonemes would be scored 0. Total possible score was 66.

Additional Scoring Procedures

In addition to the above scoring for each measure, different methods of scoring were used for some of the measures which resulted in 19 outcome scores for the 11 measures.

Full Word Scoring

WRAT Reading - Words

As mentioned earlier, this measure begins with children naming letters, and when scored in the standardized manner for children at kindergarten level, most of the "reading" score results from
letter-name knowledge. So the WRAT Reading subtest was scored a second way, which involved just the raw score total for the number of words the child read completely correctly.

**WRAT Spelling - Words**

For the same reason as the WRAT Reading, the WRAT Spelling was scored this second way as well, using the same procedure as above -- total words correct raw scores.

**Word Attack - Raw Scores**

The standardized scoring of the Woodcock Word Attack subtest for children at the kindergarten level does not reveal much information about how many phonemes a particular child may be able to decode. Moreover, a raw score of 0 (i.e., no words correct) could result in a standard score ranging from 58 to 77, depending upon the child's age. So it was also desirable to score performance on this test different ways. For Word Attack - Raw Scores, the number of words children got totally correct were used in the analyses.

**Phonemic Analysis Scoring**

The secondary method of scoring just described for the WRAT Reading and Spelling and Woodcock Word Attack tests gives some additional information as to the children's reading skills. However, it is a rather gross and conservative measure, and much information is still not utilized. So a third, finer-grained analysis -- which was similar in style and reasoning as the phonemic scoring used for the Nonword Spelling Task described earlier -- was undertaken for the following measures. A type of "phonemic analysis" method of scoring was developed and used to re-score the following four tasks:

- **WRAT Reading -- Phonemic Analysis** (out of 10 words)
- **Burns and Roe Word recognition -- Phonemic Analysis** (out of 19 words)
- **Word Attack -- Phonemic Analysis** (out of 6 words)
- **Nonword Reading Task -- Phonemic Analysis** (out of 10 words)

The phonemic analysis scoring for each of the above measures included only those words that children would have had the opportunity to attempt even if they reached the discontinue criterion immediately. For the WRAT Reading, this meant the first 10 words, for the Burns and Roe Word
Recognition test, it was all 19 of the preprimer list, for Word Attack, the first 6 word were scored, and for the Nonword Reading Task, the first 10 words were included.

Briefly, the scoring contained the following criteria (see Appendix A for greater detail) which allowed for recognition of approximately correct (AC) phonemic decoding attempts as well as those which were completely correct (CC) within the context of the word (or digraph of a nonword). That is, if while a child was sounding out a word, the child gave a letter sound that would have been considered correct for a particular letter when seen in isolation, but was not correct for that letter within the context of the word (or within the context of a digraph within a nonword), it would be considered AC (approximately correct). This may be clearer with an example. For the word "she", a sounding out process which resulted in the child's saying /s/, /h/, /e/, would be scored as /s/ (AC) /h/ (AC) /e/ (CC). The "s" and the "h" in the word "she" are really parts of the digraph making the "sh" sound, and so the child's attempts for these two letters would be considered only approximately correct when attempted within the context of the digraph "sh" (and within the word "she"). The child's attempt for "e" was completely correct. The child would receive a point value even if the final attempt was not completely blended. With this terminology in mind, the following scoring system was developed:

a) 1 pt. - only initial phoneme completely correct.

b) 2 pts. - only 2 or more (but not all) phonemes/letters CC (completely correct), AC, or a mix of AC and CC. They have to be in proper sequence.

c) 3 pt. - all phonemes/letter sounded out AC or a mix of AC and CC, blended or not blended.
    Example: the word "spell" would receive 3 points for: /s/ /p/ /e/ /I/ or "speel".

d) 4 pts. - all phonemes sounded out CC (for the word in question) but not blended

e) 5 pts. - word read correctly.

For the WRAT Spelling -- Phonemic Representation, the scoring method was slightly different and more in depth. The system used was adapted by Haas (1998) from the spelling schema developed by Tangel and Blachman (1992, 1995) to represent the various stages found in the literature of spelling development in children (e.g., Beers & Henderson, 1977; Gentry, 1982).
The details of this scoring system can be found in Appendix A. The essence is briefly presented here. This spelling system was applied to the following words:

and
in
him
make
cook
enter
light
reach

These eight words were included in the phonemic analysis scoring for the following reason. In Haas' research, an earlier version of the WRAT had been used, and the spelling system had been specifically developed to accommodate 20 of the words in this version. In the current research, the above eight of the first ten words of the WRAT3 were within the 20-word set that Haas had studied. Words were graded on an 11 point scale (0-10) which was specifically developed for each word. This 11-point scale represented various stages of spelling development (SOURCE: from Haas, 1998, reprinted with permission).

**Precommunicative Stage**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>random letter strings.</td>
</tr>
</tbody>
</table>

**Semiphonetic Stage**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>one phonetically related letter.</td>
</tr>
<tr>
<td>2</td>
<td>one correct letter/phoneme or two phonetically related letters/phonemes.</td>
</tr>
<tr>
<td>3</td>
<td>multiple phonetically related letters/phonemes or two or more correct letters in a minimum combination of one correct and one related letter/phoneme.</td>
</tr>
</tbody>
</table>

**Phonetic Stage**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>each phoneme is represented by a correct and/or related letter/phoneme.</td>
</tr>
<tr>
<td>5</td>
<td>transposition error in otherwise phonetically correct representation.</td>
</tr>
<tr>
<td>6</td>
<td>additional letter(s) in otherwise phonetically correct representation.</td>
</tr>
</tbody>
</table>
**Transitional Stage**

(indicates phonemic correctness and increasing degree of orthographic accuracy)

- 7 phonetically correct representation.
- 8 above plus one orthographic feature correct.
- 9 above with second orthographic feature correct.

**Correct Stage**

- 10 conventional spelling.

It must be noted that "phonetically correct" can entail a very wide range of spellings. These alternate methods of scoring helped to determined whether children could at least partially apply what they had learned.

---

**Program Implementation and Classroom Observations**

**Jolly Phonics Program**

The Jolly Phonics program, developed by a British educator, Sue Lloyd (1993), was designed for beginning readers (kindergarten and Grade 1). It is not as comprehensive as a basal reading program, but is a commercially available teaching method which emphasizes phonological awareness combined with letter-sound correspondence instruction. This instructional approach is meant to be integrated into a classroom's existing program. Jolly Phonics provides theoretical explanations for the various teaching suggestions so that the teachers can understand why certain elements are so important, and thus teachers can adapt them accordingly to the unique characteristics of their individual classes (and the ability levels of the children they teach).

This program introduces the phoneme/grapheme correspondences to children in a sequenced, yet motivating way. Accompanying the "lessons" are physical actions, stories, sound and word games, home reading, and a number of other activities designed to be motivating to young children. In addition, this method places an emphasis on phonological awareness (termed auditory training by Lloyd). This is accomplished through various types of blending,
segmentation, and deletion exercises which span the different levels of phonological complexity discussed on the literature (Stahl & Murray, 1994; Yopp, 1988). This synthesis of decoding instruction, phonological training, and multi-modal experiential learning attempts to put theory into practice in an effective, motivating way.

All materials necessary for program implementation (e.g., instructions to teachers, several reproducible worksheets, games, parent information forms) are found within the main manual -- the *Phonics Handbook* (Lloyd, 1993). However, there are additional materials available -- little hardcover books featuring the letter sounds by sets, a colorful phonics frieze, workbooks, and videos.

*Letter-Sound (Grapheme/Phoneme) Introduction*

Grapheme/phoneme is a more appropriate term than letter-sound correspondence because some of the phonemes introduced are represented by more than a single letter (e.g., "sh"). However, for ease of discussion, the term "letter-sound" correspondence will be used. The Jolly Phonics program introduces letter correspondence for the 42 phonemes representing all the sounds that the English language is generally considered to include (Adams, 1990; Lloyd, 1993). These are introduced in a particular sequence (see Appendix B for details) designed to enable children to start forming words with the letters learned almost immediately. For example, the first sequence includes "s, a, t, i, p, n". Numerous simple words can be made with these letters. The sequence is also designed to separate potentially confusable letters (e.g., "b" and "d") as research suggests these are more easily learned and remembered if taught at separate times (Williams & Ackerman, 1971).

Each letter-sound correspondence is introduced through a story and an accompanying physical action. This action is logically and meaningful connected to the sound it represents by means of the story. When children perform an action, they over-articulate the sound it represents, thus further emphasizing the phonemic aspect of the letter-sound correspondence. The story-line is

---

6 Linguist disagree, however, as to the exact number of phonemes, with a range of 30 to 45 basic phonemes, depending upon the classification system used (Wagner et al., 1994).

53
suggested by the program, but details and elaboration are left up to the imagination of the teacher so that it can be adapted to be more meaningful to the children. In this study, teachers often used classroom children's names in the story, and would incorporate local events to link the entire process to the children's personal experience. The Jolly Phonics program provides materials to coincide with each of the letter-sound "lessons" (e.g., matching pictures representing the story, printed words, and actions; accompanying letter-formation practice sheets; games; puzzles). The pictures often present items that contain the sound of interest, not only in the initial position of words, but the final and medial as well. The teacher points these out, emphasizing the sound for each position. For example, Figure 1 shows a picture accompanying the story for the /p/ sound (suggested story-line is printed below). Some of the words the teacher would highlight are: pink (the birthday cake is pink), pig, stripe (on the candles), popcorn, drips (from the straw in the cup), etc.

![Figure 1](image-url)

**Figure 1.** Jolly Phonics' sample picture depicting letter sound and related story.

Note: Figure and story-line copyright 1993 by Jolly Learning Ltd. Adapted and reprinted with permission.
Suggested story-line for the /pl sound:

It is a child's birthday -- favorite animal is a pig -- mother has made a cake in the shape of a pig and has put on the cake five candles that relight when blown out -- child puffs the candles out, making a /pl/ sound, but every time the candles appear to have gone out, they light up again.

Teacher shows the "p" (flash card).
Children imagine their finger is a candle and try to puff the trick candle out saying "/pl/ /pl/ /pl/ /pl/ /pl/.

Phonological Awareness Development

Aside from the obvious emphasis on the phonological make-up of words that occurs within the introduction of letter-sound correspondences, the program suggests other activities aimed at developing phonological/phonemic awareness. One blending activity is "identifying the word", where the teacher may say "I'm going to say some sounds and I want you to tell me what word you hear"... and she slowly sounds out (segments) a word for the children to blend (e.g., /pl/ /ol/ /tl/). Segmenting activities range from simple to more complex. Some are as simple as clapping out the number of syllables in a word. More complex activities would involve segmenting words into onset and rimes (br-at, sp-at, r-at, str-ong, wr-ong, l-ong, etc.), or individual phonemes (sh-ee-p, f-i-sh, f-i-s-t, t-r-ai-n). This includes having children holding up a finger for each phoneme as they slowly say a word. Another activity is "finish the word" task, where the teacher says the whole word, and then breaks it down, asking the children to finish the word. For example, for the word "split":

teacher says "s"... class says "plit"
"sp"... class says "lit"
"spl"... class says "it"
"spli"... class says "t"

The program suggests that throughout the day, the teacher can "slip in" a phonological awareness activity. For example, when dismissing children to get their coats for outdoor play, teachers can say "children who have a /ml/ sound in their name (beginning, middle, or end) can go to get their coats."
The Jolly Phonics program consists of a number of reinforcing activities and materials, such as: sound sheets, personal sound book to take home and practice with parent help, homework writing sheets, word boxes, games, matching letter/sound/picture activities, sentence sticking (cutting and pasting sentences to match a picture), and much more. Most activities are made from reproducible sheets found in the main manual for this program. There are also reproducible parent information forms that teachers can send home in hopes of involving the parents in their children's learning process.

**Jolly Phonics Teacher Training and Program Implementation**

Prior to the teachers' integration of the Jolly Phonics program into their curricula, most were introduced to the method through a 2-hour teacher information session where the developer, Sue Lloyd, presented the main components and discussed the program development process. This session was video-taped for those teachers who could not attend. Other than this, the teachers simply used the Phonics Handbook manual to guide their programming. To help encourage the teachers to make use of the various program activities suggested, the experimenter prepared packages for each teacher containing the reproducible sound sheets, flash cards, and games along with abbreviated guidelines of the programs key components and time of introduction (see Appendix B).

The main elements of the Jolly Phonics program, at the kindergarten level, included tuning children into the sounds of words through the phonological development activities, and the explicit introduction of letter-sound correspondences. There were several reinforcing activities to accompany these main themes. Although there are some specifics in timing that the program suggests, the teachers were free to follow their own timeline and include only the activities that they felt worked best within their respective classrooms. For example, the program suggests that kindergarten children can learn 6 letter sounds a week, and that these sounds can be introduced two on Monday, two on Wednesday, and two on Friday. Tuesdays and Thursdays are spent in reinforcing games and activities. The teachers in this study felt this schedule was far too ambitious, and so introduced the letter sounds at their own pace, ranging from 1 a week, to 3 or 4
a week. It is interesting to note that some of the teachers stepped up their pace of introduction of letter sounds when they saw how quickly children were able to pick them up, and that teachers who have remained with the program in the year(s) following initial implementation, find little difficulty with the recommended instructional pace. However, for this first year, program integration and implementation varied widely.

A typical lesson would look something like this:

The teacher may first review previous days' sounds by holding up flash cards with the printed letter, and having the children perform the actions and sounds. She would then introduce the new letter, sound, and action through a personalized story (see description of Jolly Phonics Program) and have the class perform the accompanying action while saying the letter sound. Next, she would explicitly describe, visually and verbally, how the letter is formed, and have the children draw the letter in the air, articulating the sound while "drawing" the letter. She may ask what other words have the sound just learned, and query as to whose name might contain the sound somewhere in it (i.e., beginning, middle, or end). Children would then go to tables and work on a letter-formation practice sheet (Sound Sheets) with teacher guidance, and then color the pictures which correspond to the story and sound. Later in the day, they would cut and paste a copy of the letter into their sound books, which they take home and might practice with their parents, sisters, or brothers. The new letter sound would be carried through to other activities, where children might paint the new letter, trace it in the sand box, or make it out of clay during craft time.

Teachers are encouraged to come up with their own ideas for reinforcing the sounds throughout the day. During story time, for example, teachers may point out the sound whenever she comes across it. Some teachers had the children bring in a toy or item for "show and tell" which incorporated a particular sound. Other teachers might choose teams for games or activities, the membership of which was based on the sounds in the children's names (e.g., "everybody with a /p/ in their name will be in group number 1").

Once enough letters are learned to form some simple words (within the first week if the suggested sequence and timing is followed -- e.g., pin, pat, tap), these words are explored, and children are encouraged to figure out new words through various word-analysis activities. 

57
soon as the children are able to identify sounds in simple words and are able to form the letters, the program encourages the beginning of writing/spelling, and provides homework writing sheets.

Reading and writing activities continue into Grade 1.

As mentioned earlier, the teachers implemented some or all of the suggestions to varying degrees, depending upon the classroom make-up (e.g., number of ESL children) and the teachers' own preferences. However, all of the classroom implemented explicit instruction of at least some of the letter-sound correspondences, and, to some degree focused on sounds in children's names and other words, rather than letter names.7

Control Programs

As mentioned in the Participants section, control schools described their curricula as balanced literacy programs. Initial informal classroom observations (and the subsequent time-sampling procedure) revealed that many of activities in the control classes were similar in nature to those found in the experimental classrooms except for the specific activities associated with the Jolly Phonics program. There was a wide range among all classes in the time spent in the various activities.

Classes focused on positive social interaction during the school day. There was a considerable amount of time devoted to non-literacy activities including outdoor play (which was mainly supervised "recess-type" activities -- running, playground equipment climbing, tricycle riding, etc.), and indoor "free play" (which could include some literacy-related behavior)8 where children were able to spend independent time at different activity centers. Some of these centers focused mainly on play (e.g., "house", toys), some were geared more to fine-motor and creativity development (coloring, painting, building blocks, crafts), and some had puzzles, games, plastic letters and alphabet puzzles. Other activity centers were "reading centers" where children perused

7 This less than complete implementation of the Jolly Phonics approach makes the outcome (Phase 2) testing a conservative test of the effectiveness of the treatment program.

8 All indoor free play time was observed in the time-sampling procedure, and appropriate literacy categories were marked when evidence of some type of literacy-related activity occurred.
through books, and some centers were designed to encourage the exploration of math concepts such as balancing weights or measuring water, sand, and so forth. All classes had "circle time" where the date, and weather were discussed, and usually children brought something for "show and tell" at this time. Some classes also used circle time for literacy activities such as learning letter names and sounds (there was a wide range among classes in the types of literacy skills stressed), or discussed certain events. Teachers might introduce some points of geography, history, or science as themes that would be continued throughout the day (or week) with respect to whatever activity the children were engaged in (e.g., making Valentine's day cards). Children in all classes spent some of their time listening to the teacher read stories, and most teachers encouraged story discussion and prediction. There was usually some time during the day where children sang songs and might "chant" a poem written on a chart as the teacher guided their "reading" with her finger. And finally, all classes had "snack time", -- again, the length of time spent preparing, eating, and cleaning up after snack varied.

Classroom Observation Measure: Time-Sampling

In order to investigate and quantify the types of class activities children were engaged in during the treatment period of the study, very detailed time-sampling classroom observations were undertaken in the 10 experimental and 10 control classes. The time-sampling observation instrument and method was loosely modeled on one developed by Usher and Evans (1971; Evans & Carr, 1985) for research evaluating primary education programs. The current research observation protocol (Appendix C) was adapted from Usher and Evans' "pupil activity scan" and was designed to detect the various components of only language arts programs, rather than complete curricula as in the Usher and Evans study. Decisions of how to organize/group the various literacy activities into categories which made theoretical and practical sense were based both on a model of reading development and programming -- *The Balanced and Flexible Literacy Diet* -- developed by Willows, (1994), and on several initial informal classroom observations to get a "feel" for the programs to be observed.
Observations of activities were recorded under any of the main categories termed Literacy Components (described next) that were considered appropriate to the activity observed. All nonliteracy-related activities were recorded under a category termed "Other". These activities were also monitored (via time-sampling) to ensure that any unplanned (or incidental) instances of literacy were recorded. Specific discriminations among different types of play, arts and crafts, and so on, were not specifically analyzed (although observers provided written descriptions). The length of time spent in non-classroom activity (e.g., outdoor play) was also documented for comparison purposes. Appendix C includes the specifically designed observation protocol and training manual outlining the entire observation procedure and categories in depth. A brief description will be presented here of the "adjusted"9 Literacy Component categories used for the analyses in this study. Most of the main Literacy Component categories contained subcategories (all of which are described in the appendix). Here, only the subcategories for the Phonics Literacy Component will be described since these are further analyzed in the Results section.

Literacy Component Categories

Print-Related and Oral Phonics (Phonics)10

The Phonics category included all phonics-related activities that occurred in connection with printed letters/words, or in connection to verbally presented letter-names/spellings. Any teaching of letter-sound correspondence would be recorded here, as well as all word analysis activities when they involved print. This would included blending, segmenting, detection of sounds in words, explicit rhyming when working with the printed form of the words. This does not include

---

9 For ease of discussion and visual representation in the current text, some of the Literacy Components (and subcomponents) are labeled (and organized for analyses) slightly differently than they were in the manual and observation protocol (Appendix C).

10 Originally, sight words and learning of letter names had been included in the Phonics category for easier recording of observations (and description of these activities is found under the Phonics category in the manual in Appendix C). However, for purposes of analyses, these two activities were removed from this category and treated as two separate Literacy Components since they reflect quite separate skills.
reading connected text; it involves only explicit forms of letter-sound correspondence teaching or word analysis. The following subcategories were part of the Phonics main category:

**Oral Spelling**, included any verbal spelling of words, when visual representations of print were not available. This would included "chanting" with the teacher (e.g., "c-a-t spells cat").

**Practice Spelling** was marked whenever the same type of activity (as above) was occurring, but the visual printed referents were available.

Note that both the "Spelling" subcategories were marked only if the spelling activity was very basic in nature -- virtually just reciting/repeating the letters of words. If any word analysis was occurring, this would be marked in the "Word Analysis" subcategory below.

**Letter-Sound Correspondence** was included mainly for explicit teaching or reviewing of letter-sound correspondences. If letter-sounds were being used to decode a word (and were therefore a by-product of the main activity of word analysis) the observation was recorded in the Word Analysis subcategory, not here.

**Jolly Phonics Actions** (unique to the Jolly Phonics classes) was marked any time the actions (see program description above) were observed. This usually occurred along with the letter names and/or their sounds, but not always. It also sometimes occurred along with word analysis, and occasionally with auditory phonological awareness activities (where print or actual letter names were not being used, but children were performing the actions to highlight a phoneme.)

**Word Analysis** was included for all activities involving learning about word structure and spellings (with print). This was where the bulk of the observations would be scored if segmenting or blending of printed words was occurring.

**Auditory Phonological Awareness (APA)**

The APA category referred to phonological awareness activities which were not print related or letter-name related. The sounds, not the letters, had to be the focus of emphasis for activities to be recorded in this category; children's attention was not being specifically directed to printed letters or printed words (or letter names) when participating in some type of word or sound
analysis activity. The APA category included (auditory) activities involving blending, segmenting, detection of sounds in words, explicit or implicit rhyming activities, poems, chanting, or rhyming songs when no print was available.

**Sight Word (StWd)**

The StWd category included all sight word activities. Sight words are words that the children learn as a whole (a single unit), without sounding out. In kindergarten, children's sight word vocabularies are very limited. Observations marked in this category involved activities like reading isolated words on the word wall, flash cards, and so forth. This was not reading sentences, and did not include any word analysis.

**Grammar**

The Grammar category was marked whenever activities included grammar explanations, punctuation explanations, or description of sentence structure.

**Real Reading (RlRd)**

The RlRd category included any reading by children that appeared to be real reading, either a book (or sentences) at their own level. If this occurred, it usually happened when children were reading to, or along with the teacher, aide or volunteer.

**Concepts of Print (ConPrint)**

The ConPrint category included activities where children were being read to and/or they were learning about reading, print and print conventions (e.g., books work from front to back, reading goes from left to right). It also included situations where children were "pretend reading" -- mainly chanting or repeating, or "reading" memorized stories or pattern books.

**Real Writing (RlWrite)**

RlWrite included any real attempts at writing connected text, usually guided by the teacher. Instances of this were few and mainly consisted of the teacher closely helping children to work out a simple 3- or 4-word sentence to go along with a picture they had just drawn. This was not single word analysis.
Letter Formation (LForm)

The LForm category involved any specific description of how letters are formed, as well as children practicing this, either specifically, within word-copying exercises, or printing their name.

Letter Name Learning (LtNmLn)

The LtNmLn category involved only the specific learning of letter names, not their sounds, and not words. This could be marked in conjunction with Letter-Sound Correspondence if the letter's name was also taught at the time of the letter-sound introduction.

Oral Vocabulary/Language Development (Vocab)

Vocab included all oral language development and language-related activities ranging from very routine discussions of the date and weather to elaborate discussions of stories, ideas and events. Because observations were recorded under Vocab whenever any teacher-supervised discussion or talking was occurring, it became a type of "catch-all" category (i.e., "housekeeping" activities -- giving instructions, discipline, etc.) as well.

Other

The Other category included all nonliteracy-related activities, such as playing, crafts, painting, drawing, coloring, pasting, etc.

Classroom Observations: Training, Reliability, and Time-Sampling Procedure

Training

An information session was first held with the four research assistants who, along with the experimenter, were to be the observers. At this meeting the observation protocols, the procedure, and the Literacy Components were thoroughly explained along with the types of activities that fit into the various Literacy Component categories. The research assistants/observers were then given the manual to study (see Appendix C). Throughout January and early February, several pilot (i.e., practice) classroom observations with the examiner took place for each observer. Throughout the observation session, the trainee would be checked to see that the appropriate categories were being marked. The session was discussed afterwards and any questions were answered. Once a satisfactory level of performance had been achieved for each trainee (where the trainee consistently
recorded observations in the appropriate categories), the "real" observations began. These were conducted by the 4 trainees and the experimenter mainly in March, April, and May. The experimenter was available every evening to answer any questions and to give advice about any "odd" situations that may have arisen. In the majority of the questions, the observers' responses were confirmed as correct, indicating their comprehension of the observation procedures and Literacy Component categorizations. In addition, the observers wrote descriptions of the activities taking place (along with materials used) on the bottom of each page of the protocol.

There were three observations for each class in the 10 experimental and 10 control classrooms (totaling 60 observations), and each observation lasted for the entire class (i.e., half day). It had been planned that the experimenter would co-observe two in-class observations with each of the 4 assistants in order to establish inter-rater reliability for the observation measure; however, due to illness only 5 inter-rater observations were conducted. Reliability was calculated and established on these 5 observations.

Reliability

Reliability of the observation measure was established through inter-rater observations. Five of the 60 observations (8%) were conducted (the experimenter was always one of the inter-raters). The reliability for each of the literacy component totals was calculated using the percentage method, a traditional method used for calculating the reliability of duration data (Wyatt, Callahan, & Michael, 1985) in which two observers record the total time that a "behavior" (or in this case an activity) occurs within the observation session. Then, by dividing the smaller total by the larger total time, the reliability coefficient is calculated. Since, in the current study, the number of raw observations recorded in a Literacy Component category is a representation of the time spent in that type of activity, it was felt these could be considered duration data, and that the percentage method was an appropriate method of reliability measurement (see Appendix D for more detail).

Average category percentage coefficients, ranged from .60 (Letter Name Learn category) to .95 ("Other" category). Categories with few observations have lower reliabilities, simply due to the scarcity of the behaviors. This is a difficulty which often occurs with low-occurrence reliability
ratings (Saudergrass & Lentz, 1986). Thus, Chi-Square tests of significance were also performed on both the averages of the combined raw observations in each category (i.e., category averages of the five experimenter protocols compared with those of the five inter-rater protocols), and on individual sets of inter-rater categories with the lowest reliabilities. The results showed that there were no significant differences between Set 1 (5 experimenter protocols) and Set 2 (5 inter-rater protocols) in any of the Literacy Component categories tested, nor between the individual sets of inter-rater categories. Where virtually no instances of activities in a category were seen in the set of inter-rater protocols, reliabilities were not calculated (i.e., Grammar, Real Reading, and Real Writing).

*Time-Sampling Procedure*

As mentioned in the Observation Measure section, a time-sampling observation procedure using the pupil activity scan (Usher & Evans, 1977) was used. This time-sampling process involves taking a sample of the children’s behavior over a number of 10-second intervals (once-a-minute). A group of 4 children was observed for each 10-second "scan". The activities in which these children were engaged were recorded, and then, for the next 10-second scan (starting the next minute) a second group of 4 was selected, observed, information recorded, and so on.

Which activities children were involved in (i.e., Literacy Components or Other), how engaged children were (e.g., fully active, somewhat passive, not at all), and whether the activity was teacher-supervised or non-supervised (i.e., children involved in something on their own without direct teacher supervision or guidance) was recorded for each scan. If the activity in the scan had any language arts components, the observation would be recorded next to the appropriate Literacy Component(s) on the printed protocol. For example, Figure 2 presents a small portion of an observation protocol page showing how the following scenario would be recorded.
**Example of Jolly Phonics activity:**

The 4 children in the scan are involved in the following activity. The teacher is showing the letter "S" on a flashcard, and the children are responding with the "S" action (waving their arms in the form of a snaky "S") and they are saying "/S/ /S/ /S/ /S/." The scan is recorded as follows:

<table>
<thead>
<tr>
<th>Time:</th>
<th>10:41</th>
<th>10:42...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>SUPERVISION</strong></td>
<td>S</td>
<td>EN</td>
</tr>
<tr>
<td>T led</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. <strong>ENGAGEMENT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fully engaged:active</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. <strong>PHONICS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OralSpell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PracSpell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/Sound Corresp</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>JPActions</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Word Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. <strong>MATERIALS/ACTIVITY</strong></td>
<td>JP flash card</td>
<td></td>
</tr>
</tbody>
</table>

**Key:**
- S = supervised, NS = nonsupervised.
- EN = engagement (used only if children in the scan were differently engaged)
- Tled = Teacher led activity.
- 4 = represents the 4 children in that particular scan.

**Figure 2.** Example of observation scoring. This shows the recording of 4 children involved in a letter-sound teaching activity during a 10-second scan.

**Note:** This is a truncated version of the full observation sheet (with only pertinent categories showing).

Each ten-second scan gives a short "glimpse" as to what is occurring in the classroom. By collecting a series of glimpses of all the pupils in quick succession, the larger picture can be seen which will reflect the component emphases of the teaching program as revealed by the children's behavior. According to Usher and Evans (1977), "the greater the number of children ... engaged in an activity, and the longer ... [the] activity persisted, the greater the chance that it would be observed and recorded during a ten-second scan. Consequently, the activity would be predominant in the summarized composite picture." (p. 12).
RESULTS and DISCUSSION

Overview

The results section details findings from Phase 2 posttesting and the time sampling observations. It is divided into 4 main parts related to the 4 questions guiding this study. The first part, examines performance of the Jolly Phonics and control classes on the outcome measures. The second section compares the groups, via the observations, with regard to the time spent in various literacy activities. The third part examines the relation between literacy component times and outcome performance. The last section, investigates the performance of the subsample of children at risk for reading difficulty.

Variable Screening

Prior to analysis all Phase 1 (pretest) and Phase 2 (posttest) variables were examined (using various Minitab and SPSS programs) for accuracy of data entry, missing values, distributions and other assumptions of univariate and multivariate analyses. For analyses involving MANOVA's and ANOVA's, the variables were examined by group (i.e., for the experimental and control). For regression analyses, the scores for the experimental and control groups were combined. For grouped data analyses, two sets of analyses were conducted when variables were found to possibly violate the normality assumption. Analyses were run on both transformed (via square root and logarithmic transformations) and nontransformed variables. Differences in these analyses were negligible, with no change in the multivariate level of significance, and no change with regard to which univariate ANOVA's contributed to the MANOVA results. Furthermore, when Box's M was run on the nontransformed variables, and was rejected, an examination of the determinants of S1 and S2 (the generalized variance for the two groups) revealed that the larger generalized variance was with the larger-sized group (i.e., the experimental group), indicating that the multivariate statistics will be conservative (Stevens, 1996 p. 260). This means that if differences are found between experimental and control groups on the measures, these results may be a slight underestimate of the true differences between groups. Therefore, analyses using nontransformed
variables will be discussed, since interpretation of these results is more meaningful. Variable screening involving subsets of the sample (e.g., the comparison of at-risk children in the Jolly Phonics and control groups) will be discussed as encountered (i.e., Part 4).

**PART 1**

**Phase 2 (Posttest): Group Comparisons on Outcome Measures**

The performance of the Jolly Phonics and control groups on the outcome measures was compared. For the analyses, Phase 2 measures were grouped according to how they were scored (i.e., standard scores, full word raw scores, or more detailed analyses) and what they measured. The three standardized tests (using the standard scores) made up the first set of measurements -- Set 1 (i.e., Woodcock Word Attack, WRAT Reading, and WRAT Spelling).

Set 2 consisted of "all or nothing" raw scores on tests measuring word recognition and spelling; that is, the word had to be totally correct in order to receive a point. Therefore, Set 2 measures are referred to as "Full Word Raw Score" measures. This set included Burns and Roe Word Recognition, Woodcock Word Attack raw scores (particularly to examine the mean number of full words the children were able to decode), the experimentally-designed Nonword Reading Task, and the WRAT Reading and WRAT Spelling raw scores just for the number of words successfully read (not including the letter portion of these tests -- at kindergarten level for the WRAT Reading and WRAT Spelling, the larger, and often the only, portion of a child's full standard score is achieved through naming/writing letters rather than words). In the reporting of the results, these Full Word Raw Score (Set 2) measures are listed in order of "real" word reading (i.e., WRAT Reading -- Words, and the Burns and Roe Word Recognition), real word spelling (i.e., WRAT Spelling -- Words), followed by the pseudoword reading tasks (i.e., Woodcock Word Attack -- Raw Scores and the Nonword Reading Task). So the second set consisted of scores for the 5 measures of reading or spelling success with words (and nonwords/pseudowords) as wholes.
Set 3 analyses were performed on the tests that measure various components of reading and spelling processes, as well as the application of these components/skills. This set of 11 measures was divided into three subsets: (A) a phonemic awareness test (the Rosner Test of Auditory Analysis Skills) and tasks of basic literacy skills (Letter Name and Writing Alphabet Tasks), (B) measures of more complex literacy skills -- the alphabetic coding measures (Letter-Sound Recognition Task and Letter-Sound Recall Task), and (C) the 6 measures which were scored phonemically (see description of measures in Methods section) to assess the application of alphabetic coding skills to both real word (WRAT Reading -- Phonemic Analysis, Burns and Roe Word Recognition -- Phonemic Analysis, WRAT Spelling -- Phonemic Analysis) and to nonword decoding and/or encoding (Woodcock Word Attack -- Phonemic Analysis, Nonword Reading Task -- Phonemic Analysis, Nonword Spelling -- Phonemic Analysis).

For all findings reported in the Results section, analyses were considered significant if alpha levels were less than .05. Occasionally, analyses only approaching significance -- .05 < p < .10 (i.e., quasi-significance) are reported when findings are of particular interest, or when N is small (e.g., reduced from an N of 265 for the analyses comparing individual scores, to N = 20 when comparing class means).

In order to avoid an inflated overall Type I error rate (Stevens, 1996, p. 152), 3 omnibus MANOVA analyses11 were first performed, one for each set of tests. The individual scores of the children in each group were used for all Phase 2 outcome comparisons. Although there is some concern regarding independence of observations (because of class groupings), most of the alpha levels were well beyond .01 and therefore reached a far more stringent criterion for significance, which is appropriate in cases of possible non-independence (Stevens, 1996, p. 241). Indeed all MANOVA's and most univariate ANOVA's reached significance levels beyond p < .001. Moreover, MANOVA's were also performed on class means (which circumvents the problem entirely, but also loses much information). Results of the class-means MANOVA's were all significant, and univariate ANOVA's demonstrated that the same variables (as reported in this

11 The technical term for the 2-group MANOVA is Hotelling's T², however, for ease of discussion, the term MANOVA will be used for these analyses.
section's analyses) significantly contributed to the results (with the exception of the Burns and Roe Word Recognition -- Phonemic Analysis measure). See Appendix E for a table of class-mean MANOVA results.

Outcome Results

On almost every measure comprising the three sets analyzed, the children in the Jolly Phonics group significantly outperformed their control group counterparts, indicating that these children learned the skills taught and were able to apply them. The following analyses describe the results of these three omnibus MANOVA's by set. Means, standard deviations, $F$ ratios and $p$ values, as well as effect sizes are presented in accompanying tables for Sets 1 and 2. Results are depicted graphically for most of the Set 3 measures.

Set 1 (Standard Score Measures)

Table 5 presents the results of the MANOVA and univariate ANOVA's comparing the Jolly Phonics and control groups on the first set of variables (standardized measures with standard scores). The MANOVA was significant $F(3, 261) = 9.133, p < .001$, and all three measures contributed significantly to the results: Woodcock Word Attack Standard Scores, $F(1, 263) = 26.73, p < .001$; WRAT Reading Standard Scores, $F(1, 263) = 16.09, p < .001$; and WRAT Spelling Standard Scores, $F(1, 263) = 12.23, p < .001$. The Jolly Phonics group scored significantly higher on all of the standardized measures than did the control group. Effect sizes were medium to medium/high (Cohen, 1977), and all would be considered to have practical significance (i.e., effect sizes >.33, Borg et al., 1993).
Table 5.

Phase 2 Comparisons (Outcome Measures)
Jolly Phonics vs. Control
Set 1 (Standard Score Measures)

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 145)</th>
<th>Control (n = 120)</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>WaSS</td>
<td>99.83 21.98</td>
<td>86.43 19.78</td>
<td>26.73</td>
<td>.001</td>
<td>.68</td>
</tr>
<tr>
<td>WrRdSS</td>
<td>106.73 12.91</td>
<td>100.57 11.96</td>
<td>16.09</td>
<td>.001</td>
<td>.52</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>103.89 13.88</td>
<td>98.03 13.18</td>
<td>12.23</td>
<td>.001</td>
<td>.44</td>
</tr>
</tbody>
</table>

MANOVA, $F(3, 261) = 9.133, p < .001$.

SET 2 (Full Word Raw Score Measures)

A similar pattern of results showing performance differences on the second set of measures is presented in Table 6. Once again the MANOVA was significant, $F(5, 259) = 5.59, p < .001$. Examination of the univariate tests reveal that all of the measures contributed significantly to the MANOVA result. Even though these measures are rather conservative, in that a child must get the word completely correct in order to receive a point, the tests do significantly separate the experimental and control groups' performances. The children in the Jolly Phonics group scored significantly higher than those in the control on each real word and nonword measure.

For the word portion of the WRAT Reading and WRAT Spelling tests, the experimental group averaged 2.63 (SD = 3.10) and 1.20 (SD = 1.64) words respectively, compared with the control children's average of 1.29 (SD = 2.06) and 0.48 (SD = 0.95). Both of these differences were significant at $p < .001$ (WRAT Reading, $F[1, 263] = 16.24, p < .001$; and WRAT Spelling, $F[1, 263] = 17.95, p < .001$). On the Burns and Roe Word Recognition test the mean difference was 3.25 words, $F(1, 263) = 6.17, p < .05$, with the experimental group averaging 6.28 words (SD = 13.23), and the control group averaging 3.03 words (SD = 6.02). The Jolly Phonics group also surpassed controls on the Woodcock Word Attack (nonword) test and the informal Nonword Reading Task. Their mean for the former (Word Attack) was 3.58 words (SD = 5.45) which was
significantly superior to the control group's average of 1.01 words (SD = 3.02), \( F(1, 263) = 21.24, p < .001 \). On the latter, the Jolly Phonics group averaged 2.9 words (SD = 4.42) compared with 0.78 words (SD = 2.26) for the control children, \( F(1, 263) = 22.69, p < .001 \). Large effect sizes were noted on the nonword measures, with real word measures demonstrating medium to medium/large effect sizes.

Table 6.

Phase 2 Comparisons (Outcome Measures)
Jolly Phonics vs. Control
Set 2 (Full Word Raw Score Measures)

<table>
<thead>
<tr>
<th>Jolly Phonics ( (n = 145) )</th>
<th>Control ( (n = 120) )</th>
<th>( F )</th>
<th>( p &lt; )</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>WrRdWrds ( M )</td>
<td>2.63</td>
<td>3.10</td>
<td>1.29</td>
<td>2.06</td>
</tr>
<tr>
<td>WrRdWrds ( SD )</td>
<td>6.28</td>
<td>13.23</td>
<td>3.03</td>
<td>6.02</td>
</tr>
<tr>
<td>WrSpWrds ( M )</td>
<td>1.20</td>
<td>1.64</td>
<td>0.48</td>
<td>0.95</td>
</tr>
<tr>
<td>WrSpWrds ( SD )</td>
<td>3.58</td>
<td>5.45</td>
<td>1.01</td>
<td>3.02</td>
</tr>
<tr>
<td>WaRS ( M )</td>
<td>2.90</td>
<td>4.42</td>
<td>0.78</td>
<td>2.26</td>
</tr>
</tbody>
</table>
| WaRS \( SD \) | MANOVA, \( F(5, 259) = 5.59, p < .001 \).

Key: \( \text{WrRdWrds} = \) WRAT Reading Full Words, \( \text{B&R} = \) Burns and Roe Word Recognition, \( \text{WrSpWrds} = \) WRAT Spelling Full Words, \( \text{WaRS} = \) Woodcock Word Attack-- Raw Score, \( \text{NwRd} = \) Nonword Reading Task

Set 3 (Phonemic Awareness, Basic Skills, Coding, and Phoneme Analysis)

The final Phase 2 posttest comparisons were performed on measures of the underlying skills/components and subskills related to the processes of reading and spelling, as well as on measures of the application of these skills. The first two subsets of Set 3 measures (i.e., Subsets A and B) contained all the same measures that had been administered at Phase 1 pretesting (except the Burns and Roe Word Recognition test which was included in Set 2). Since the earlier "end-of-year" testing time on these measures for the experimental group (i.e., April) would favor the control group (who were tested in June on these, and all of the other outcome measures), this slightly different testing time-table for the two groups made the experimental and control
comparisons (on the posttest set of Phase 1 measures) a conservative test (see Time-Line in Methods section for details of testing times). The MANOVA performed on the 11 measures of Set 3 was highly significant, $F(11, 253) = 15.20, p < .001$. Only 3 (Subset A) of the 11 measures did not significantly contribute to the MANOVA results.

Set 3 -- Subset A (Phonemic Awareness and Basic Skills)

The groups did not differ on the posttest measures of basic literacy skills (i.e., Writing Alphabet Task and the Letter Name Task), and the difference in group means on the phonemic awareness measure (i.e., Rosner Test of Auditory Analysis Skills -- TAAS) only approached significance, univariate $F(1, 263) = 3.24, .05 < p < .10$. The Jolly Phonics group raw score mean on this measure was 4.08 (SD = 3.25) compared with the control mean of 3.43 (SD = 2.54). The effect size was small (.26), however, examination of the test items at these score levels reflect a qualitative difference between the two groups. Items one to three involve segmenting words at the level of syllables. The fourth item requires segmenting the word at a phonemic level (i.e., initial phoneme deletion). On the average, control children were able to segment words into syllables. The Jolly Phonics group mean of 4 indicates that, aside from being able to segment words into syllables, these children were also able to show, on this auditory phonological test, a beginning understanding of the phonemic make-up of words. Phonemic analysis of reading and spelling measures (presented later) suggest that the Jolly Phonics children actually had a more in-depth awareness of phonemes than indicated by their performance on the TAAS (as did some of the control children), suggesting that the TAAS may not be sensitive enough to detect early (and perhaps more subtle) levels of phonemic awareness development.

On the Writing Alphabet Task, both groups performed comparably $F(1, 263) = 1.42, \text{ns.}$, with the Jolly Phonics group averaging 13.17 letters (out of 26) in 60 seconds (SD = 7.15), and the control group averaging 12.15 (SD = 6.58). On the Letter Name task, the Jolly Phonics and control means were 21.52 out of 26 (SD = 6.62) and 20.31 (SD = 6.15) respectively, $F(1, 263) = 2.35, \text{ns.}$ This lack of difference on the latter task is not surprising since both groups are near ceiling, and the Jolly Phonics Program does not focus on letter names. In fact, the program tends
to avoid naming letters and instead, highlights the visual letter(s) connected to its sound. The consequence of this strategy is reflected in the large differences in the alphabetic coding measures (i.e., letter-sound correspondence) reported below, as well as in the application of this knowledge to the measures of phonemic analysis.

**Set 3 -- Subset B (Alphabetic Coding)**

Not surprisingly, children in the Jolly Phonics group outperformed those in the control group on both aspects of alphabetic coding. These findings essentially show that the children in the Jolly Phonics group learned what they were taught. Figure 3 depicts the performance of the two groups on these measures of letter-sound correspondence knowledge (i.e., Letter-Sound Recognition Task and Letter-Sound Recall Task). When asked which letter made a given sound (Letter-Sound Recognition Task), the Jolly Phonics children significantly outperformed their control counterparts $F(1, 263) = 38.20, p<.001$, averaging 17.42 letter-sound correspondences (SD = 7.0) compared with 11.7 (SD = 8.07) for the controls. Their recall of letter-sound correspondences was also impressive with a mean of 19.69 (SD = 9.02), $F(1, 263) = 74.74$, $p < .001$. The control mean was 10.12 (SD = 8.91). Effect sizes were large, .71, and 1.07 respectively, indicating considerable practical significance.
Alphabetic Coding

Figure 3. Performance of Jolly Phonics and control children on the Phase 2 Outcome measures of alphabetic coding.

Set 3 -- Subset C (Phonemic Analyses)

The practical significance of the Jolly Phonics group's superior alphabetic coding skills is demonstrated on the phonemic scoring of reading and spelling measures, which reflect the children's application of their knowledge. These measures were scored so that children were given credit for their attempts at analyzing words phonemically either through sounding out, or through generating good phonemic approximations of spellings (termed "phonemic analysis" for this paper). In order to have a better mental picture of the kind of skills these measures assess, Figure 4 presents examples of some of the higher quality attempts at WRAT Spelling words written by children in the Jolly Phonics group (lines A), and a range of fairly typical attempts at the same words by control-group children (lines B). Although the first spellings (on lines A) of the words "make", "cook", and "reach" are not totally correct, the appropriate phonemic representations of most of the sounds, show a high degree of understanding of the alphabetic principle. When scored phonemically, these spellings receive credit (see Methods section and Appendix A for details of
this scoring procedure), but when they are scored in the standardized (or the Full Word Raw Score) manner (i.e., either right or wrong), the attempts displayed on line A would receive the same score value (i.e., 0) as the ones on line B, and thus, much important information about the child's true skill level would be lost.

1 A  mark
1 B  A W O R T O F

2 A  co K
2 B  K

3 A  re e C H
3 B  L L E

Key:
1 = make
2 = cook
3 = reach
A = attempts by children in the Jolly Phonics group from the upper range.
B = a range of typical attempts by children in the control group.

Figure 4. Samples of WRAT Spelling of Jolly Phonics and control children. When scored by the standardized criterion, both A and B responses would receive a score of 0.
Informal observations of the types of responses on the spelling measures indicated that the Jolly Phonics children typically demonstrated a better quality of letter formation, and that they used lower-case letters more consistently, although these were not scored in the analyses. Figure 4 illustrated the qualitative differences between groups in their responses. Although there was a broad range in caliber of spelling attempts in both groups, the upper range of quality was found more often in the Jolly Phonics group. Similarly impressive statistical differences were found on all of the measures that assessed children's practical use of what they learned in their respective classroom programs.

**Real Words**

Figure 5 represents results of the phonemic scoring of the real word measures. For WRAT Reading -- Phonemic Analysis, the Jolly Phonics group average score was 15.76 (SD = 12.46), which was significantly higher than the control mean of 9.26 (SD = 10.76), $F(1, 263) = 20.16$, $p < .001$, effect size = .60. A consistent pattern of results was seen on the Burns and Roe -- Phonemic Analysis, and the WRAT Spelling -- Phonemic Analysis. For the Burns and Roe measure, the experimental group mean was 25.83 (SD = 28.89) and the control mean was 16.75 (SD = 23.55), $F(1, 263) = 7.66$, $p < .01$, and effect size was .39. The Jolly Phonics group's average score on the WRAT Spelling -- Phonemic Analysis, was 28.08 (SD = 21.07) compared with the control average of 16.19 (SD = 15.76), $F(1, 263) = 26.12$, $p < .001$. The effect size for the spelling difference was large (.75).
Figure 5. Performance of Jolly Phonics and control children on the phonemic scoring of the Phase 2 Outcome real word measures.

Nonwords

When the children had to apply their word analysis skills to the nonword tasks (where one's sight word lexicon would not be able to facilitate execution of the task), the performance gap was very striking. It was evident that the children in the Jolly Phonics Program made very good use of their letter-sound knowledge on these "purer" measures of alphabetic coding application. Figure 6 shows that when these measures were scored phonemically, the Jolly Phonics group average scores were at least twice those of the control group. On the Woodcock Word Attack -- Phonemic Analysis the mean score for the Jolly Phonics group was 13.72 (SD = 9.64), compared with only 5.17 (SD = 7.34) for the control children, $F(1, 263) = 63.87, p < .001$. Phonemic scoring on the nonstandardized Nonword Reading Task yielded similar results, with the Jolly Phonics group's mean of 21.49 (SD = 13.89) being significantly superior to the control group's 8.81, (SD = 10.70), $F(1, 263) = 67.04, p < .001$. Scores on the Nonword Spelling
Task – Phonemic Analysis followed the same pattern, $F(1, 263) = 50.49, p < .001$. The experimental group averaged 28.08 (SD = 18.17) compared with the control group's mean of 13.65 (SD = 14.10). Effect sizes for all three nonword phonemic analysis measures were large, 1.17, 1.19, and 1.02 respectively.

### Phonemic Analysis - Nonwords

![Phonemic Analysis - Nonwords](image)

**Key:**  
- WaPh = Woodcock Word Attack -- Phonemic Analysis  
- NwRdPh = Nonword Reading Task -- Phonemic Analysis  
- NwSpPh = Nonword Spelling Task -- Phonemic Analysis

**Figure 6.** Performance of Jolly Phonics and control children on the phonemic scoring of the Phase 2 Outcome nonword measures.

To summarize, the Jolly Phonics group's performance was significantly higher than that of the control group on 16 of the 19 Phase 2 outcome measures. These 16 measures assessed complex literacy skills (i.e., alphabetic coding) and different aspects of reading and spelling. Of particular practical importance was the experimental group's performance on the phonemic analysis measures, which lends insight as to how well the children were able to apply their letter-sound knowledge. These findings indicated that the children in the Jolly Phonics group clearly knew what to do with what they had learned.
The children in the experimental classrooms did not demonstrate superior performance on the two measures representing basic (readiness) skills, and on the measure of phonemic awareness (the Rosner Test of Auditory Analysis Skills -TAAS). As explained earlier, the lack of significant differences on the basic skill measures (i.e., Writing Alphabet Task, and Naming Letters Task) was not surprising given that these areas are not emphasized in the Jolly Phonics program. However, the lack of a significant effect on the TAAS for the Jolly Phonics group was unexpected. Notwithstanding the earlier suggestion that the TAAS may be insensitive to early stages of phonemic development, it was still somewhat puzzling that a stronger effect on this measure was not apparent, especially since the literature suggests a close relationship between reading and phonemic awareness. Because of this, and because the univariate ANOVA for the TAAS comparison approached significance (i.e., .05 < p <.10), and, because the Jolly Phonics group Phase 1 (pretest) average on this measure was slightly, though not significantly, lower than that of the controls (see Method section), it was decided to take a closer look at performance on the TAAS. So the change scores of the two groups on this measure were examined. Change scores represent the difference between pre- and posttest performance.

Comparison of Change Scores in Phonemic Awareness

An ANOVA performed on the change scores indicated that the performance on the TAAS for the children in the Jolly Phonics group improved significantly more than did that of the children in the control group, $F(1, 263) = 15.08, p < .001$. The mean change in score for the Jolly Phonics group was 1.88 (SD = 2.62), compared with the control group's mean of 0.71 (SD = 2.19), effect size = .53. There may be some aspects of the Jolly Phonics programming, not found in the control classes, that tend to facilitate better the development of phonemic awareness.

Although the Jolly Phonics group outperformed the control group on all measures of reading and spelling, they did not, as already mentioned, score significantly higher on all the skills associated with beginning reading/spelling. Therefore, it was of interest to see how closely performance on the skill and phonemic awareness measures was linked to literacy achievement in this study.
Correlations of Phase 2 Outcome Measures:

*Phonemic Awareness, Basic Literacy Skills, and Alphabetic Coding with Literacy Performance*

Using the full sample including all experimental and control children, Pearson correlations were undertaken to see how closely phonemic awareness ability (TAAS), the ability to write and name letters (basic literacy skills), and knowledge of letter sounds (alphabetic coding) were associated with reading and spelling whole words (real and nonsense words) and to phonemic analysis (and representations) of words (see Table 7). Not surprisingly, given the large sample size (N = 265), all of the correlations were statistically significant, however, there was a large range in the actual size and practical significance of the correlations.

Although the p values reported in the correlation tables in this study indicate significance of individual correlations, it must be noted that with correlational analyses involving a large number of variables, there will be some correlations that reach significance purely by chance (since alpha is inflated with multiple tests). However, it is the relative size, and the consistency of the associations of specific components with reading and spelling measures which is of most interest to this research (particularly in the next section examining correlations of Literacy Components using class means where N = 20).

For the set of correlational analyses examined here, with N=265, the majority of correlations are significant even when accounting for the large number of variables, since critical r is approximately .365 at p = .05 for 25 variables and df = 263 (Shavelson, 1988); and all (but one) of the correlations for the components of most interest (i.e., alphabetic coding, see below) would still be considered significant at an alpha level of at least .01.

The correlational analysis showed that the two measures of alphabetic coding skills (i.e., Letter-Sound Recognition and Letter-Sound Recall) were particularly closely related to successful reading and spelling, with almost all correlations being at least r = .50. The highest alphabetic coding correlations were with the phonemic analysis scoring of the reading and spelling measures (i.e., Subset C of Set 3 measures), running as high as r = .76. The consistently moderate/high to high correlations of alphabetic coding with all reading/spelling achievement measures point to the
importance of these complex literacy skills, and to the need to include explicit instruction of these components in early reading programs.

The Letter Name Task showed moderate/high correlations with the WRAT Reading and Spelling standard scores ($r = .66$ for both), presumably because of the emphasis on letter names at the kindergarten level for these tests. The Letter Name Task’s moderate associations with some of the phonemic analysis measures were probably the by-product of the letter-sound learning activities, as suggested by Letter Name’s moderate/high correlations with the alphabetic coding measures (see Table 8 for intracorrelations of phonemic awareness and literacy skill measures).

It is interesting to note that phonemic awareness ability (as measured by the TAAS) was not particularly strongly correlated with any of the outcome measures (and in fact these correlations were similar in strength to those of the Letter Name Task), suggesting once again, that the TAAS may not discriminate phonemic awareness as well as expected.
Table 7.

Correlations of Phase 2 Posttest Measures of Phonemic Awareness, Basic Literacy Skills, and Alphabetic Coding with Literacy Performance (N = 265)

<table>
<thead>
<tr>
<th>Set 3 - Subsets A &amp; B</th>
<th>Phonemic Awareness</th>
<th>Basic Literacy Skills</th>
<th>Alphabetic Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TAAS</td>
<td>LtNm</td>
<td>WrtAlph</td>
</tr>
<tr>
<td>Set 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WaSS</td>
<td>0.300***</td>
<td>0.270***</td>
<td>0.161**</td>
</tr>
<tr>
<td>WrRdSS</td>
<td>0.345***</td>
<td>0.655***</td>
<td>0.351***</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>0.309***</td>
<td>0.663***</td>
<td>0.418***</td>
</tr>
<tr>
<td>Set 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WrRdWds</td>
<td>0.497***</td>
<td>0.439***</td>
<td>0.310***</td>
</tr>
<tr>
<td>B&amp;R</td>
<td>0.459***</td>
<td>0.297***</td>
<td>0.220***</td>
</tr>
<tr>
<td>WrSpWds</td>
<td>0.543***</td>
<td>0.397***</td>
<td>0.330***</td>
</tr>
<tr>
<td>WaRS</td>
<td>0.477***</td>
<td>0.334***</td>
<td>0.230***</td>
</tr>
<tr>
<td>NwRd</td>
<td>0.506***</td>
<td>0.339***</td>
<td>0.277***</td>
</tr>
<tr>
<td>Set 3 - Subset C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WrRdPh</td>
<td>0.469***</td>
<td>0.540***</td>
<td>0.382***</td>
</tr>
<tr>
<td>B&amp;RPh</td>
<td>0.533***</td>
<td>0.466***</td>
<td>0.389***</td>
</tr>
<tr>
<td>WrSpPh</td>
<td>0.586***</td>
<td>0.532***</td>
<td>0.402***</td>
</tr>
<tr>
<td>WaPh</td>
<td>0.444***</td>
<td>0.477***</td>
<td>0.381***</td>
</tr>
<tr>
<td>NwRdPh</td>
<td>0.489***</td>
<td>0.487***</td>
<td>0.356***</td>
</tr>
<tr>
<td>NwSpPh</td>
<td>0.545***</td>
<td>0.532***</td>
<td>0.423***</td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < .01; ***p < .001.

Key:
- Set 2: WrRdWds = WRAT Reading Full Words, B&R = Burns and Roe Word Recognition, WrSpWds = WRAT Spelling Full Words, WaRS = Woodcock Word Attack Raw Score, NwRd = Nonword Reading Task
- Set 3: Subset A: TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, & WrtAlph = Write Alphabet Task
- Subset B: L/SLRecog = Letter-Sound Recognition Task, L/SLRecall = Letter-Sound Recall Task
- Subset C: WrRdPh = WRAT Reading -- Phonemic Analysis, B&RPh = Burns and Roe Word Recognition -- Phonemic Analysis, WrSpPh = WRAT Spelling Phonemic Analysis, WaPh = Woodcock Word Attack -- Phonemic Analysis, NwRdPh = Nonword Reading Task -- Phonemic Analysis, NwSpPh = Nonword Spelling Task -- Phonemic Analysis.
Intracorrelations Among Phonemic Awareness, Basic Literacy Skills, and Alphabetic Coding

Since the alphabetic coding measures distinguished themselves with respect to their associations with outcome measures, their correlations with the other literacy skill measures (i.e., basic skills) and phonemic awareness task (TAAS) were examined. Table 8 presents these intracorrelations. As expected, the two letter-sound correspondence tasks were highly correlated with each other ($r = .89$). They showed considerably lower, but moderate and moderate/high correlations with Letter-Name Task ($r = .60$, and $r = .65$), and minimal correlations with the other measures, suggesting a unique relation of the alphabetic coding measures to reading and spelling performance.

Table 8.

<table>
<thead>
<tr>
<th></th>
<th>TAAS</th>
<th>LtNm</th>
<th>WrtAlph</th>
<th>L/SRecog</th>
</tr>
</thead>
<tbody>
<tr>
<td>LtNm</td>
<td>0.280***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WrtAlph</td>
<td>0.212**</td>
<td>0.472***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/SRecog</td>
<td>0.440***</td>
<td>0.652***</td>
<td>0.389***</td>
<td></td>
</tr>
<tr>
<td>L/SRecall</td>
<td>0.462***</td>
<td>0.599***</td>
<td>0.350***</td>
<td>0.885***</td>
</tr>
</tbody>
</table>

Note:  *$p < .05$; **$p < .01$; ***$p < .001$.

Key:  **TAAS** = Rosner Test of Auditory Analysis Skills, **LtNm** = Letter Name Task, **WrtAlph** = Write Alphabet Task, **L/SRecog** = Letter-Sound Recognition Task, **L/SRecall** = Letter-Sound Recall Task

These sets of correlational analyses convincingly demonstrate the strong relation between alphabetic coding ability and performance on the reading and spelling measures. The better children remembered which sound went with what letter, the better they were able to read, spell, and analyze words.
PART 2

Group Comparisons of Time Spent

Part 1 of the data analysis showed a large discrepancy between the Jolly Phonics and the control groups in reading and spelling performance. These results clearly showed that kindergarten children can acquire early reading and spelling skills when they are explicitly taught. Although these findings demonstrate the efficacy of the Jolly Phonics teaching program as a whole, they do not indicate whether any particular program components are key, and in what ways they may work. In order to investigate the differences in the experimental and control program components, analyses of the classroom observations were undertaken examining time spent in the main Literacy Component categories.12

These analyses were performed on the 20 classes as the unit of analysis, not individual children. First, the totals for the individual "raw" observations in each of the main categories (i.e., the different Literacy Components, and "Other" category) were converted to the "percentage of the school day" (PSD) spent in that activity. All other time intervals, termed "non-class activity" periods (e.g., additional time periods spent in transitions from activity to activity, recess/outdoor play) were also converted to PSD values (see Appendix D for details of conversions and reliability of observation measure).

Since three separate observations (on separate days) were conducted in each of the classes, these conversions yielded 60 PSD values for each of the main categories (as well as for the "non-class activity" time periods). In other words, each class had a set of 3 PSD values for each of the categories and time periods of interest. The mean of this set of 3 was then calculated for each class resulting in 20 class means (10 experimental and 10 control) for each of the Literacy Component categories, for the Other category, and for the non-class activity time periods. This average is considered to be more representative of a typical day than any single observation period.

12 As mentioned in the Method section, most of the main Literacy Component categories contained subcategories, and occasionally, the same observation could be scored in more than one subcategory (within the main Literacy Component category). However, when summing main category raw observations for analyses, any dual (or multiple) recordings of the same observation within a main Literacy Component category were only counted once.
Non-Literacy Time Comparisons

As mentioned above, all comparisons were conducted on class means (N = 20). First, a series of ANOVA's was performed to see if the experimental and control classes differed in the length of the school day, or in time spent on outdoor play, transitions, or interruptions. Table 9 presents the means, standard deviations, F ratios, significance levels, and effect sizes for these time periods. None was significant. The observation activity category labeled Other was also included in this set of analyses, and when examining this part of the school day activity, it can be seen that children in the control classes spend significantly more time involved in non-literacy (i.e., Other) activities, $F(1, 18) = 9.46, p < .01$, than do the children in the Jolly Phonics classrooms.

Table 9.
Non-Literacy Time Period Comparisons: Jolly Phonics Classes vs. Control Classes

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 10)</th>
<th>Control (n = 10)</th>
<th>F</th>
<th>p&lt;</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td><strong>School Day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(minutes)</td>
<td>154.00</td>
<td>5.16</td>
<td>153.50</td>
<td>10.01</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>Outdoor Play</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(% of school day)</td>
<td>10.56%</td>
<td>7.83</td>
<td>11.98%</td>
<td>8.11</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Transitions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(% of school day)</td>
<td>13.94%</td>
<td>4.19</td>
<td>13.90%</td>
<td>3.30</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Interruptions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(% of school day)</td>
<td>1.90%</td>
<td>1.90</td>
<td>1.73%</td>
<td>1.26</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Other activity category</strong></td>
<td>28.14%</td>
<td>5.09</td>
<td>35.28%</td>
<td>5.29</td>
<td>9.46*</td>
</tr>
<tr>
<td>(% of school day)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*ANOVA, $F(1, 18) = 9.46, p < .01$

Literacy Time Comparisons

The analyses of particular interest were those examining how much of the school day children spent involved in activities which were literacy related. An omnibus MANOVA
comparing control and experimental classes was significant $F(10, 9) = 4.65, p < .05$. Subsequent univariate analyses revealed that a number of literacy categories contributed to this result. Table 10 shows that the largest significant difference ($F[1, 17] = 28.9, p < .001$) was found in the Phonics category where the participants in the experimental classes spent an average of 9.92% (SD = 4.20) of the day (i.e., approximately 15 minutes) involved in phonics-related activities compared to an average of only 1.96% (SD = 2.06) (i.e., approximately 3 minutes) for the control children.

Significant differences were also found in the Sight Word (StWd), $F(1, 18) = 28.9, p < .001$, and Grammar categories, $F(1, 18) = 28.9, p < .001$ with control classes spending less time involved in each type of activity. In addition, the children in the Jolly Phonics classes spent significantly more time participating in auditory phonemic awareness activities, $F(1, 17) = 5.30, p < .05$, averaging 7.18% (SD = 4.19) of the school day (i.e., approximately 11 minutes), whereas the average for the control classes was only 3.70% (SD = 2.28) (i.e., approximately 6 minutes).

Table 10.

Percent of School Day Spent in Literacy-Related Activities: Jolly Phonics Classes vs. Control Classes

<table>
<thead>
<tr>
<th>Jolly Phonics</th>
<th>Control</th>
<th>$F$</th>
<th>$p&lt; $</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 10)</td>
<td>(n = 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phonics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>9.92%</td>
<td>4.20</td>
<td>1.96%</td>
<td>2.06</td>
<td>28.90</td>
</tr>
<tr>
<td><strong>APA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.18%</td>
<td>4.19</td>
<td>3.70%</td>
<td>2.28</td>
<td>5.30</td>
</tr>
<tr>
<td><strong>StWd</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.67%</td>
<td>1.29</td>
<td>0.42%</td>
<td>0.53</td>
<td>7.96</td>
</tr>
<tr>
<td><strong>Grammar</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.76%</td>
<td>0.61</td>
<td>0.07%</td>
<td>0.21</td>
<td>11.37</td>
</tr>
<tr>
<td><strong>RfRd</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.31%</td>
<td>0.39</td>
<td>0.07%</td>
<td>0.12</td>
<td>3.42</td>
</tr>
<tr>
<td><strong>ConPrint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.14%</td>
<td>1.91</td>
<td>7.20%</td>
<td>2.80</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>RfWrite</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.38%</td>
<td>0.55</td>
<td>0.25%</td>
<td>0.52</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>LForm</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.84%</td>
<td>1.15</td>
<td>1.27%</td>
<td>1.01</td>
<td>1.39</td>
</tr>
<tr>
<td><strong>LtmLn</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.85%</td>
<td>0.88</td>
<td>1.32%</td>
<td>1.94</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Vocab</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.78%</td>
<td>5.26</td>
<td>20.31%</td>
<td>8.50</td>
<td>0.64</td>
</tr>
</tbody>
</table>

MANOVA, $F(10, 9) = 4.65, p < .015$

Key: APA = Auditory Phonic Awareness, StWd = Sight Word, RfRd = Real Reading, ConPrint = Concepts of Print, RfWrite = Real Writing, LForm = Letter Formations, LtmLn = Letter Name Learning, Vocab = Vocabulary/Language Development

87
The results in of these analyses clearly demonstrate that, on a daily basis, the children in the Jolly Phonics classrooms spent considerably more time involved in certain literacy-based activities than did their control classroom counterparts; and children in the control classrooms spent more of their day in nonliteracy activities (e.g., play, crafts). These differences in time spent help to identify potential key elements of the Jolly Phonics program which may have contributed to the superior end-of-year performance for the recipients of this teaching method. The next step was to see which of these differences related to children's performance on the outcome measures.

PART 3
Relation of Time Spent to Outcome Measures

Correlations

As with the group comparisons of time spent in literacy activities, all of the correlational analyses were performed on class means, not individual participants. For this set of analyses the experimental and control classes were combined (N = 20). On examination of the intercorrelations among the activity categories (i.e., Literacy Components and the Other category), it was noted that Real Reading and Grammar were highly correlated ($r = .75, p < .001$). To avoid multicollinearity in regression analyses (Stevens, 1996), and because all observations in these two categories were mutually exclusive (and there were relatively few observations overall for both), these components were collapsed (by summing both sets of observations) into one category labeled RdGram. The next step was to examine correlations of the activity-literacy categories with the class means of the various outcome measures.

As noted previously, $p$ values reported for the correlational analyses indicate significance of individual correlations. Since correlations with multiple variables inflates the level of alpha, significance of the correlations should be interpreted with caution. However, as also mentioned earlier, it is the relative size, and the consistency of the associations of specific Literacy
Components with the range of reading and spelling measures which is of most interest to this research. Multiple measures were used for the purpose of exploring this consistency of associations. If tests of significance had been the prime focus of the correlational analyses, specific key Literacy Components and outcome measures (e.g., Phonics and Auditory Phonological Awareness with alphabetic coding and nonword decoding/encoding measures) would have been judiciously chosen to limit the number of variables, thus reducing the chance of spurious correlations. Had this occurred, many of the more notable reduced-variable correlations would still have reached significance (i.e., critical \( r \) for six variables = .660, or .686 for 7 variables) as can be seen in the following tables. On the other hand, some important associations would have also been missed (see below).

**Literacy Components with Set 1 (Standard Score Measures)**

An investigation was then undertaken to see whether time spent in specific classroom activities was related to performance on the three sets of outcome test measures. First, correlations of the Literacy Components with the Set 1 measures (the 3 standard score tests) were examined. Table 11 presents results of Pearson correlations between the time spent in Literacy Component (and Other) activities and performance on the standardized scoring of the Woodcock Word Attack, the WRAT Reading, and the WRAT Spelling tests. The Phonics component had a moderate correlation (\( r = .59, p < .01 \)) with the Woodcock Word Attack and WRAT Reading scores (\( r = .46, p < .05 \)), and the correlation was approaching significance with the Spelling scores (\( r = .43, .05 < p < .10 \)). This indicates that in all the classrooms examined, the time children spent involved in activities focusing on phonics-related elements (e.g., explicit letter-sound correspondences, word analysis) was associated with how well children performed on standardized measures of reading and spelling. The Letter Formation component also showed a moderate correlation (\( r = .52, p < .05 \)) with the Woodcock Word Attack test scores (a measure of nonword analysis), suggesting that something other than just learning letters may occur when direct guidance in the formation of letters takes place. It may be that the context in which this activity occurs underlies
the relation of the Letter Formation component with nonword decoding (see General Discussion for elaboration of this suggestion).

Literacy Components which are often considered to be important to reading acquisition (e.g., Concepts of Print, Vocabulary/Language Development) showed no association to the standardized measures of which Set 1 was comprised. As will become evident, these components also showed no significant (positive) relation to any of the Phase 2 outcome measures in this study.

Interestingly, the Auditory Phonemic Awareness component was also not significantly correlated with these or any of the other outcome measures. Possible reasons for this will be discussed later.

Table 11.

| Correlations Between Literacy Components (and Other) and Set 1 (Standard Scores Measures) (N = 20) |
|-------------------------------|------------------|------------------|
|                               | WaSS             | WrRdSS           | WrSpSS           |
| **Phonics**                   |                  |                  |
| *APA*                         | 0.158            | -0.050           | -0.027           |
| *StWd*                        | 0.415            | 0.268            | 0.289            |
| *RdGram*                      | 0.322            | 0.243            | 0.156            |
| *ConPrint*                    | 0.170            | 0.242            | 0.244            |
| *RIWrite*                     | 0.370            | 0.402            | 0.265            |
| *LForm*                       | 0.524*           | 0.396            | 0.348            |
| *LtNmL*                       | -0.277           | -0.294           | -0.313           |
| *Vocab*                       | -0.197           | -0.159           | -0.060           |
| *Other*                       | -0.333           | -0.152           | -0.179           |

Note: *p < .05; **p < .01; ***p < .001.

Key: Literacy Components

APA = Auditory Phonemic Awareness, StWd = Sight Word, RdGram = Real Reading and Grammar, ConPrint = Concepts of Print, RIWrite = Real Writing, LForm = Letter Formation, LtNmL = Letter Name Learning, Vocab = Vocabulary/Language Development, Other = Other non-literacy activities.

Literacy Components with Set 2 (Full Word Raw Score Measures)

Next, the correlation of the time spent on the various Literacy Components with Set 2 measures (full word raw scores) was examined. Both the Phonics and the Letter Formation categories had consistent, positive correlations with all of these measures (see Table 12). Significant correlations ranged from moderate to moderate/high ($r = .48, p < .05$ to $r = .62, p < .01$) for the Phonics component, and the correlations for the Letter Formation category were all moderate in strength, ranging from $r = .50, p < .05$ to $r = .55, p < .01$. These results support those found in the first set of correlations with the more conservative (standardized measures). Not only is time spent in phonics activities connected to performance on reading and spelling, but there seems to be something about letter formation activities which is also related to how well children read and spell whole real words and nonwords. No other Literacy Components showed any significant association with the Set 2 measures.
Table 12.

Correlations Between Literacy Components (and Other) and Set 2 (Full Word Measures) 
(N = 20)

<table>
<thead>
<tr>
<th></th>
<th>WrRdWds</th>
<th>B&amp;R</th>
<th>WrSpWds</th>
<th>WaRS</th>
<th>NwRd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonics</td>
<td>0.587**</td>
<td>0.477*</td>
<td>0.514*</td>
<td>0.624**</td>
<td>0.604**</td>
</tr>
<tr>
<td>APA</td>
<td>0.087</td>
<td>-0.111</td>
<td>0.129</td>
<td>0.187</td>
<td>0.148</td>
</tr>
<tr>
<td>StWd</td>
<td>0.269</td>
<td>0.218</td>
<td>0.180</td>
<td>0.311</td>
<td>0.295</td>
</tr>
<tr>
<td>RdGram</td>
<td>0.332</td>
<td>0.209</td>
<td>0.281</td>
<td>0.345</td>
<td>0.260</td>
</tr>
<tr>
<td>ConPrint</td>
<td>0.218</td>
<td>0.151</td>
<td>0.120</td>
<td>0.047</td>
<td>0.039</td>
</tr>
<tr>
<td>RIWrite</td>
<td>0.426</td>
<td>0.381</td>
<td>0.271</td>
<td>0.287</td>
<td>0.315</td>
</tr>
<tr>
<td>LForm</td>
<td>0.550*</td>
<td>0.495*</td>
<td>0.548*</td>
<td>0.518*</td>
<td>0.501*</td>
</tr>
<tr>
<td>LtNmL</td>
<td>-0.311</td>
<td>-0.343</td>
<td>-0.275</td>
<td>-0.260</td>
<td>-0.294</td>
</tr>
<tr>
<td>Vocab</td>
<td>-0.184</td>
<td>-0.179</td>
<td>-0.129</td>
<td>-0.272</td>
<td>-0.255</td>
</tr>
<tr>
<td>Other</td>
<td>-0.253</td>
<td>-0.175</td>
<td>-0.216</td>
<td>-0.301</td>
<td>-0.307</td>
</tr>
</tbody>
</table>

Note:  *p < .05;  **p < .01; ***p < .001.

Key:

**Literacy Components:**

APA = Auditory Phonemic Awareness, StWd = Sight Word, RdGram = Real Reading and Grammar, ConPrint = Concepts of Print, RIWrite = Real Writing, LForm = Letter Formation, LtNmL = Letter Name Learning, Vocab = Vocabulary/Language Development, Other = Other non-literacy activities.

**Set 2**

WrRdWds = WRAT Reading Full Words, B&R = Burns and Roe Word Recognition, WrSpWrds = WRAT Spelling Full Words, WaRS = Woodcock Word Attack Raw Score, NwRd = Nonword Reading Task

Literacy Components with Set 3

(Phonemic Awareness, Basic Skills, Coding, and Phoneme Analysis Measures)

Examination of the correlations between Literacy Components and the Set 3 outcome measures show some moderate/high and high positive correlations of the same two components (Phonics and Letter Formation), as well as positive and negative associations of some other components. Tables 13 and 14 depict these correlations by the three subsets of Set 3 described below.
Literacy Components with Set 3 -- Subset A (Phonemic Awareness and Basic Skills Measures)

The Phonics category was not significantly correlated with any of the subset A measures (see Table 13); however, the Letter Formation category showed a moderate correlation with Letter Names \( (r = .44, p < .05) \), and a moderate/high positive correlation with the phonemic awareness measure -- TAAS \( (r = .68, p < .001) \). This latter result suggests that activities emphasizing explicit letter formation practice may draw attention to the sound the particular letters represent. In the Jolly Phonics classrooms these activities often occurred within the context of alphabetic coding instruction, and thus the correlation of Letter Formation with the TAAS may help explain why the previous correlations between Letter Formation and the reading and spelling measures were found.

Literacy Components with Set 3 -- Subset B (Alphabetic Coding Measures)

Table 13 shows that the Phonics category was highly (positively) correlated with Letter-Sound Recall \( (r = .71, p < .001) \), and showed a moderate/high correlation with Letter-Sound Recognition \( (r = .61, p < .01) \). The Letter Formation category also showed strong positive correlations with alphabetic coding. These associations seemed to follow the same pattern as the Phonics correlations in that the significance of the Letter Formation's relation with Letter-Sound Recall \( (r = .62, p < .01) \) was at stricter level of probability than the level of the Letter-Sound Recognition correlation \( (r = .53, p < .05) \). These set of correlations demonstrate that the more time children spent in phonics activities and practicing letter formation, the better they performed on tests measuring knowledge of letter-sound correspondences.

It is interesting to note that the Vocabulary/Language Development category showed moderate negative correlations with both Letter-Sound Recall \( (r = -.48, p < .05) \) and Letter-Sound Recognition \( (r = -.52, p < .05) \) suggesting that for all classes, when more time was spent on activities encouraging vocabulary/language development, it might have been at the expense of phonics (letter-sound) instruction. A Pearson correlation of these two categories indeed showed that the Vocabulary/Language Literacy Component was moderately negatively correlated with the Phonics Literacy Component \( (r = -.47, p < .05) \). However, informal examination of the protocols
noted that observations of several types of "housekeeping" language activities (e.g., discussion of bringing in money for "pizza" day) were recorded in the Vocabulary/Language category; and it may be that these sorts of activities are the bases for the negative correlations. That is, classes that spent more time on these less language-oriented activities, may have had less time left for phonics instruction.

Table 13.

<table>
<thead>
<tr>
<th>Correlations Between Literacy Components and Set 3 Measures -- Subsets A and B (Phonemic Awareness, Literacy Skills, and Alphabetic Coding) (N = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subset A</strong>&lt;br&gt;<strong>Phonemic Awareness and Basic Skills</strong>&lt;br&gt;<strong>TAAS</strong></td>
</tr>
<tr>
<td>Phonics</td>
</tr>
<tr>
<td>APA</td>
</tr>
<tr>
<td>StWd</td>
</tr>
<tr>
<td>RdGram</td>
</tr>
<tr>
<td>ConPrint</td>
</tr>
<tr>
<td>RIWrite</td>
</tr>
<tr>
<td>LForm</td>
</tr>
<tr>
<td>LtNmL</td>
</tr>
<tr>
<td>Vocab</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

**Note:** *p < .05; **p < .01; ***p < .001.

**Key:**

**Literacy Components**
- **APA** = Auditory Phonemic Awareness,
- **StWd** = Sight Word,
- **RdGram** = Real Reading and Grammar,
- **ConPrint** = Concepts of Print,
- **RIWrite** = Real Writing,
- **LForm** = Letter Formation,
- **LtNmLn** = Letter Name Learning,
- **Vocab** = Vocabulary/Language Development,
- **Other** = Other non-literacy activities.

**Set 3**
- **Subset A** = TAAS = Rosner Test of Auditory Analysis Skills, **LtNm** = Letter Name Task, **WrtAlph** = Write Alphabet Task
- **Subset B** = L/SRecog = Letter-Sound Recognition Task, L/SRecall = Letter-Sound Recall Task
Literacy Components with Set 3 -- Subset C (Phonemic Analysis Measures)

In the set of six outcome measures representing a finer analysis of children's ability to apply their alphabetic coding skills, several powerful, positive associations were revealed (see Table 14). The Phonics category was correlated with all but one of the reading/spelling measures which were scored according to degree of phonemic analysis. The highest correlations were with the nonword measures, demonstrating that the more time children spent involved in Phonics activities during their school day, the better they were able to analyze the phonemes of nonwords with respect to both reading (decoding) and spelling (encoding). On the nonword task assessing ability to represent phonemes in writing (i.e., Nonword Spelling Task -- Phonemic Analysis) the correlation was $r = .68, p < .001$. For nonword reading measures, the correlations of the Phonics Literacy Component ran as high as $r = .76, p < .001$ with Woodcock Word Attack -- Phonemic Analysis, and $r = .82, p < .001$ with the experimentally-designed Nonword Reading Task -- Phonemic Analysis. The Phonics category was also significantly correlated with 2 of the 3 real word measures when they were scored phonemically -- WRAT Reading and WRAT Spelling ($r = .60, p < .01$, and $r = .63, p < .001$ respectively). Phonics was weakly correlated with the phonemic analysis of Burns and Roe Word Recognition which only approached significance ($r = .44, .05 < p < .10$).

The Literacy Component of Letter Formation was again correlated with many of the measures. Interestingly, however, this category seemed to be more closely linked to the real word categories since it had two significant correlations with these measures and only one with the nonword measure. The correlation with WRAT Reading -- Phonemic Analysis was moderate/high ($r = .63, p < .01$), and with WRAT Spelling -- Phonemic Analysis the association was moderate ($r = .544, p < .05$). The one correlation with the nonword measure (Nonword Reading Task -- Phonemic Analysis) was moderate ($r = .49, p < .05$).

The Sight Word category also had some significant positive correlations, but only with the nonword phonemic analysis measures. This category showed moderate correlations with the phonemic analysis scores on the Woodcock Word Attack measure ($r = .50, p < .05$), the Nonword Reading measure ($r = .50, p < .05$), and the Nonword Spelling measure ($r = .53, p < .05$). This
relation between the Sight Word category and nonword reading and spelling performance suggests activities which are aimed at the development of children's sightword vocabulary (e.g., word-wall activities) may draw attention to some spelling/sound relationships that children remember when attempting to decode/encode nonwords.

The combined category of Reading/Grammar showed a moderate significant correlation with only the Nonword Spelling Task -- Phonemic Analysis ($r = .47, p < .05$).

There were two negative associations between the Other category (representing nonliteracy activities) with the nonword measures. These were moderate negative correlations of $r = -.46, p < .05$ with the Nonword Reading Task -- Phonemic Analysis, and $r = -.49, p < .05$ with the Nonword Spelling Task -- Phonemic Analysis. That is, it appears that spending more time in class on non-literacy activities is associated with less successful phonemic analysis of the sounds in nonword reading, and less effective representations of the sounds in spelling nonwords (pseudowords).
Table 14.

Correlations Between Literacy Components and Set 3 Measures -- Subset C
(Phonemic Analysis of Real Words and Nonwords)
(N = 20)

<table>
<thead>
<tr>
<th>Literacy Components</th>
<th>Real Words</th>
<th>Nonwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>APA</td>
<td>0.596**</td>
<td>0.822***</td>
</tr>
<tr>
<td>StWd</td>
<td>0.153</td>
<td>0.373</td>
</tr>
<tr>
<td>RdGram</td>
<td>0.304</td>
<td>0.498*</td>
</tr>
<tr>
<td>ConPrint</td>
<td>0.265</td>
<td>0.349</td>
</tr>
<tr>
<td>RIWrite</td>
<td>0.333</td>
<td>0.145</td>
</tr>
<tr>
<td>LForm</td>
<td>0.433</td>
<td>0.218</td>
</tr>
<tr>
<td>LtNmL</td>
<td>0.627**</td>
<td>0.406</td>
</tr>
<tr>
<td>Vocab</td>
<td>-0.197</td>
<td>-0.118</td>
</tr>
<tr>
<td>Other</td>
<td>-0.313</td>
<td>-0.339</td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < .01; ***p < .001.

Key: Literacy Components
APA = Auditory Phonemic Awareness, StWd = Sight Word, RdGram = Real Reading and Grammar, ConPrint = Concepts of Print, RIWrite = Real Writing, LForm = Letter Formation, LtNmLn = Letter Name Learning, Vocab = Vocabulary/Language Development, Other = Other non-literacy activities.

Subset C: WRdPh = WRAT Reading -- Phonemic Analysis, B&RPh = Burns and Roe Word Recognition -- Phonemic Analysis, WrSpPh = WRAT Spelling -- Phonemic Analysis, WaPh = Woodcock Word Attack -- Phonemic Analysis, NwRdPh = Nonword Reading Task -- Phonemic Analysis, NwSpPh = Nonword Spelling Task Phonemic Analysis

To summarize, the correlational analyses reveal that for all of these kindergarten classes, there were many classroom literacy activities which were consistently related to end-of-year performance on various outcome measures of reading, spelling, and reading readiness. The more time children in the Jolly Phonics and control classrooms spent engaged in these activities, the higher (or lower, as in some cases of Vocabulary and nonliteracy activities) were their scores on various reading and spelling measures at the end of the school year. The Phonics component
showed the most consistent and strongest positive associations with the outcome scores, ranging from moderate correlations with quasi-significant levels ($r = .43, .95 < p < .10$) to high ($r = .82, p < .001$). Positive correlations for the Letter Formation were also evidenced often, and these ranged from $r = .44, p < .05$ to $r = .68, p < .001$. There were a few moderate positive correlations for the Sight Word component which were significant, and one for the combined Reading/Grammar category.

What was particularly surprising in the correlation analyses was the fact that the Auditory Phonemic Awareness was not significantly correlated with any of the outcome scores. Examination of the subcategories revealed that a very large proportion of the observations were scored in the Implicit Rhyme subcategory which includes mainly songs, rhyming chants, and so on, but no explicit attention drawn to phonological units of words. For the Jolly Phonics group, 42.91% (SD = 33.94%) of all Auditory Phonemic Awareness observations were scored in this subcategory, and for the control children this proportion was even greater, 65.89% (SD = 33.85%). Implications of these findings and further qualitative interpretations will be presented in the General Discussion section.

Two significant negative correlation were revealed for the Other (non-literacy activities) and the Vocabulary/Language Development category. In addition, time spent in other activities thought to be central to early kindergarten classroom programming such as learning about print concepts, or activities involving authentic writing, also showed no association to any of the Phase 2 outcome measures in this study, suggesting that any putative links these components have to reading/spelling achievement, are of an indirect nature.

**Multiple Regression Analyses**

The next set of analyses examined which combination of Literacy Components from the classroom observations best predicted reading and spelling success on outcome measures. As in the correlational analyses, it is the class means that are being examined, and so, there are 20
"subjects" in the regression analyses. The observations did not track individual participants; instead, they yielded information on each classroom as a whole. Therefore, a maximum of only four variables could be selected as the set of predictors for the regression analyses in order to maintain a minimum viable predictor to subject ratio (Tabachnick & Fidell, 1989).

Variable Selection

Although certain variables were highly correlated with outcome scores suggesting that they would be the first obvious choices for the regression analysis, choosing variables in this manner as predictors is not a preferred means of variable selection and may result in "sample specific" findings (Stevens, 1996). With a large number of correlations, there will be some that are significant purely by chance, and thus, would not appear in correlation analyses with a different sample of individuals. It is therefore preferable to choose predictors based on theoretical considerations (see Stevens, 1996 for further discussion of variable selection).

The choice of the Phonics category for the first variable in the regression was based on theoretical (and empirical) information, and as part of the hypothesis of this study which assesses the efficacy of a phonics-based reading program. Ordinarily, the Auditory Phonemic Awareness would have been the second predictor, but since it was very clear from the correlation analysis that there was no relation of this variable to any of the outcome measures (and for theoretical reasons to be discussed later) it was decided to drop this component from the analysis. The final three variables, although largely chosen by the process of elimination described below, also had some theoretical and practical support.

The Other category was eliminated as a predictor because, although it could perhaps contribute to prediction in a negative way, it was of no interest in terms of finding out what is important to the development of reading and spelling skills. It was felt that the Vocabulary category, as defined in the observation measures, would also have no direct influence on reading and writing acquisition, although it would likely contribute to language development. In addition, this category became somewhat "diluted," since, as described earlier, it appeared to be a catch-all category where any type of verbal discourse (including instructions, discipline, etc.) was recorded.
The category of Letter-Name Learning was not chosen as a potential predictor for the following reason. The activity of learning the names of letters will likely predict outcome measures of letter name knowledge, but not much else, and knowing how many letter names a child can generate is of no real interest as an outcome measure. It is of great interest as a pretest measure since letter-name knowledge at the beginning of kindergarten has been consistently linked to future reading success (e.g., Adams, 1990; Chall, 1967, 1983, 1996; Scarborough, 1998). However, this relation is probably not a direct result of letter-name knowledge per se, but instead, is likely an indirect result of the literacy environment to which a preschooler has access prior to formal schooling. Time spent on an activity focusing on pure letter-name knowledge (i.e., the conditions under which the Letter-Name Learn category would be scored) is of no particular interest. Real Writing was eliminated because, in kindergarten, there were very few instances of this activity. The category Concepts of Print, as defined by the observation measure, was also dropped from the list of potential predictors because it mainly covered children becoming familiar with what books were and how they work through teachers reading them stories, or by flipping through books on their own. Although this can be considered to be literacy-related activity, direct learning of reading or writing words was not taking place when this category was scored. There were also no correlations of this category with any of the outcome measures.

The above process of elimination resulted in the final 3 categories of Sight Word, Letter Formation, and the category which was a combination of the Real Reading and Grammar categories; and these three components are reasonable predictor candidates in their own right. The teaching of sight words should have some impact on at least the word recognition measures, and perhaps some of the spelling. As for the Real Reading portion of the combined Reading/Grammar category, in order for this category to be scored, the observer had to be reasonably certain that children were recognizing some simple words when reading, and this usually occurred with the teacher helping the child to read/decode. The reason for the inclusion of the Grammar portion as a reasonable predictor was not initially as evident. However, when examining the protocols for instances where Grammar was scored, it was noted that this mainly consisted of activities which either required children to put jumbled (simple) sentences into correct order, or visual explanations...
of when to use capital letters (which would highlight, at least in part, the spellings of words) and involve some cursory reading.

As can be seen by the correlation section, the category of Letter Formation is also a very likely predictor, but because theoretical support was also desired for including this as a predictor (for both empirical and statistical reasons), the protocols were again examined to ascertain some explanation for the strong link between this component and outcome scores. It was noted, at least for the experimental classes, that most letter formation observations were the end part of letter-sound learning exercises (a key component of the Jolly Phonics program). As mentioned earlier, it seems likely that practicing letters, particularly in the context of learning their sounds, may aid in drawing children's attention to the spellings of words. In addition, Evans and Carr had found some mild association between printing and reading achievement in their classroom observational study (1985).

**Regression Results**

The regression results are arranged by order of the 3 sets of measures. Hierarchical analyses were performed with the Phonics category entered as the first variable. After Phonics was forced into the equation, stepwise analysis was used to select the strongest predictor(s) of the remaining three variables: Letter Formation, Sight Word, and Reading/Grammar. In addition, for each regression a best subsets analysis (setwise) was also performed to confirm that Phonics was indeed the best predictor. In the few cases where Phonics was not the best predictor, alternative analyses (i.e., stepwise with Phonics not forced) are presented. The setwise analysis was also examined to determine if combined variables entered as a single step produced a better model that did the straight stepwise regression.

Results tables show $R^2$ (i.e., the proportion of variance that the regression explains) adjusted $R^2$, (this is $R^2$ adjusted to more closely reflect the goodness of fit of the model in the population - Norusis, 1994), the $F$ ratio and significance level (for Multiple $R^2$) for each step in the regression. The table also presents the change in $R^2 (R^2Ch)$, as well as the $F$ ratio ($FCh$) and significance level for this change for each variable at the step when it was entered into the equation.
In this way the significance of the proportion of explained variance can be assessed for each additional variable.

**Literacy Components Predicting Set 1 (Standard Score Measures)**

Table 15 presents the regression results for the Set 1 standard score variables. For the Woodcock Word Attack regression, the Phonics component alone explained 30.5% (adjusted $R^2$) of the variance in these scores, $F(1, 18) = 9.35, p < .01$. Stepwise regression produced no significant second predictor, but when entered as a group (as indicated by the setwise analysis), the Letter Formation and Sight Word components accounted for a significant increase of 23.6% (adjusted $R^2Ch = 19.4%$), of explained variance. The 3 predictors accounted for a total 49.9% (adjusted $R^2$) of the variance, $F(3, 16) = 7.3, p < .01$. Addition of the final variable (RdGram) increased $R^2$ by a negligible amount (1.1%) and adjusted $R^2$ actually decreased. The time children spent in activities involving phonics, the direct formation of letters, and the emphasis of sight words, explained almost half of the variance in their end-of-year scores on this nonword standardized measure.

The WRAT Reading and Spelling regressions showed less powerful predictions. Table 15 shows that only the Phonics variable significantly predicted performance on the WRAT Reading explaining just 16.6% (adjusted $R^2$) of the variance, $F(1, 18) = 4.78, p < .05$. For the Spelling, the Phonics component accounted for 14.25% (adjusted $R^2$) of the variance in spelling scores, which only approached significance at a probability level of $p < .10 \ (F[1, 18] = 4.13)$. The weaker prediction for these standard score measures is to be expected, because at the kindergarten level WRAT Reading and Spelling mainly taps letter knowledge as the first 15 points in the raw scores are achieved through naming letters/writing letters. However, as can be seen in the next sections, when regression analyses were performed using the other methods of scoring these measures (i.e., full-word raw scores and phonemic analysis), a different picture emerged.
Table 15.

Literacy Components Predicting Set 1 (Standard Scores Measures)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mult $R^2$</th>
<th>(adj) $R^2$</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>$R^2$ Ch</th>
<th>$F$ Ch</th>
<th>$p&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y = Word Attack</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>34.2</td>
<td>30.5</td>
<td>9.35</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm (setwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>StWd</td>
<td>57.8</td>
<td>49.9</td>
<td>7.30</td>
<td>.01</td>
<td>23.6</td>
<td>4.46</td>
<td>.05</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>58.9</td>
<td>47.9</td>
<td>5.37</td>
<td>.01</td>
<td>1.1</td>
<td>0.41</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = WRAT Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>21.0</td>
<td>16.6</td>
<td>4.78</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>25.9</td>
<td>17.2</td>
<td>2.97</td>
<td>.10</td>
<td>4.9</td>
<td>1.12</td>
<td>ns</td>
</tr>
<tr>
<td>3. StWd</td>
<td>30.3</td>
<td>17.3</td>
<td>2.32</td>
<td>ns</td>
<td>4.4</td>
<td>1.02</td>
<td>ns</td>
</tr>
<tr>
<td>4. RdGram</td>
<td>31.5</td>
<td>13.2</td>
<td>1.72</td>
<td>ns</td>
<td>1.2</td>
<td>0.24</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = WRAT Spelling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>18.7</td>
<td>14.2</td>
<td>4.13</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>22.0</td>
<td>12.8</td>
<td>2.39</td>
<td>ns</td>
<td>3.3</td>
<td>0.72</td>
<td>ns</td>
</tr>
<tr>
<td>3. StWd</td>
<td>27.2</td>
<td>13.6</td>
<td>2.00</td>
<td>ns</td>
<td>5.2</td>
<td>1.17</td>
<td>ns</td>
</tr>
<tr>
<td>4. RdGram</td>
<td>27.2</td>
<td>7.8</td>
<td>1.40</td>
<td>ns</td>
<td>0.0</td>
<td>0.00</td>
<td>ns</td>
</tr>
</tbody>
</table>

Step 1, df = 1, 18
Key: LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar

Literacy Components Predicting Set 2 (Full Word Raw Score Measures)

The next set of analyses looked at how well this set of variables predicted the actual number of words (i.e., full word raw scores) children got right on the reading and spelling measures. Table 16 presents the results of these 5 regression analyses, starting with the 3 measures of real words, followed by the 2 nonword tasks. The regression examining the word portion of the WRAT Reading once again showed Phonics to be the main predictor; however, this time it explained a much larger proportion of the variance of these full word scores. Adjusted $R^2$ was 30.8%, $F(1, 18) = 9.44, p < .01$. The Letter Formation component explained an additional 11%
which only approached significance at the level of $p < .10$. No other variables added significantly to the equation.

The Phonics variable was the only significant predictor for the Burns and Roe Word Recognition regression analysis explaining 18.4% (adjusted $R^2$), of the variance in scores $F(1, 18) = 5.30, p < .05$. This is a relatively small proportion compared with the explained variance for word recognition on the WRAT Reading (full word raw scores) described above. One reason for this may be that standard regression screening measures for this variable (e.g., Cook's Distance, residual plots), found it to be relatively unstable, and thus results should be viewed with caution.

For the measure of real word spelling (WRAT Spelling - word portion), Phonics accounted for 22.3% (adjusted $R^2$) of the variance, $F(1, 18) = 6.47, p < .05$, and an additional 13.3% (adjusted $R^2_{Ch} = 10.3\%$) was explained by the Letter Formation variable which only approached significance $p < .10$. None of the other variables was significant.

For the two nonword measures, only the Phonics predictor was significant. For the Woodcock Word Attack analysis, Phonics accounted for 35.6% (adjusted $R^2$), $F(1, 18) = 11.50, p < .01$, and it explained 32.9% (adjusted $R^2$) of the variance in performance for the Nonword Reading Task, $F(1, 18) = 10.32, p < .01$.

Thus, knowing how much time children spent in phonics and letter formation activities, helped to predict how well they were able to read and spell real words. However, only the Phonics variable aided prediction of how well children were able to decode nonwords.
Table 16.
Literacy Components Predicting Set 2 (Full Word Measures)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mult$R^2$</th>
<th>(adj)$R^2$</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>$R^2$Ch</th>
<th>FCh</th>
<th>$p&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y = WRAT Reading Words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>34.4</td>
<td>30.8</td>
<td>9.44</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>45.5</td>
<td>39.1</td>
<td>7.09</td>
<td>.01</td>
<td>11.1</td>
<td>3.46</td>
<td>.10</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>52.3</td>
<td>43.4</td>
<td>5.85</td>
<td>.01</td>
<td>6.8</td>
<td>2.31</td>
<td>ns</td>
</tr>
<tr>
<td>4. StWd</td>
<td>53.4</td>
<td>41.0</td>
<td>4.30</td>
<td>.05</td>
<td>1.1</td>
<td>0.36</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = Burns and Roe Word Recognition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>22.7</td>
<td>18.4</td>
<td>5.30</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>33.2</td>
<td>25.3</td>
<td>4.22</td>
<td>.05</td>
<td>10.5</td>
<td>2.66</td>
<td>ns</td>
</tr>
<tr>
<td>3. StWd</td>
<td>37.0</td>
<td>25.1</td>
<td>3.13</td>
<td>.10</td>
<td>3.8</td>
<td>0.96</td>
<td>ns</td>
</tr>
<tr>
<td>4. RdGram</td>
<td>37.8</td>
<td>21.2</td>
<td>2.28</td>
<td>ns</td>
<td>0.8</td>
<td>0.19</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = WRAT Spelling Words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>26.4</td>
<td>22.3</td>
<td>6.47</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>39.7</td>
<td>32.6</td>
<td>5.59</td>
<td>.05</td>
<td>13.3</td>
<td>3.72</td>
<td>.10</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>45.2</td>
<td>34.9</td>
<td>4.39</td>
<td>.05</td>
<td>5.5</td>
<td>1.59</td>
<td>ns</td>
</tr>
<tr>
<td>4. StWd</td>
<td>45.5</td>
<td>31.0</td>
<td>3.13</td>
<td>.05</td>
<td>0.5</td>
<td>0.10</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = Word Attack Raw Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>39.0</td>
<td>35.6</td>
<td>11.50</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>46.8</td>
<td>40.5</td>
<td>7.47</td>
<td>.01</td>
<td>7.8</td>
<td>2.50</td>
<td>ns</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>53.2</td>
<td>44.5</td>
<td>6.07</td>
<td>.01</td>
<td>6.4</td>
<td>2.22</td>
<td>ns</td>
</tr>
<tr>
<td>4. StWd</td>
<td>54.4</td>
<td>42.3</td>
<td>4.48</td>
<td>.05</td>
<td>1.2</td>
<td>0.38</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = Nonword Reading Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>36.4</td>
<td>32.9</td>
<td>10.32</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>43.7</td>
<td>37.1</td>
<td>6.60</td>
<td>.01</td>
<td>7.3</td>
<td>2.19</td>
<td>ns</td>
</tr>
<tr>
<td>3. StWd</td>
<td>47.2</td>
<td>37.3</td>
<td>4.76</td>
<td>.05</td>
<td>3.5</td>
<td>1.04</td>
<td>ns</td>
</tr>
<tr>
<td>4. RdGram</td>
<td>48.2</td>
<td>34.4</td>
<td>3.50</td>
<td>.05</td>
<td>1.0</td>
<td>0.31</td>
<td>ns</td>
</tr>
</tbody>
</table>

Step 1, df = 1, 18

Key:  
LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar
Literacy Components Predicting Set 3 Measures
(Phonemic Awareness, Basic Skills, Coding, and Phoneme Analysis)

The final regression analyses are presented according to the three groupings of subsets (defined by the types of skills tapped) for the Set 3 measures.

Literacy Components Predicting Set 3 -- Subset A
(Phonemic Awareness and Basic Skills Measures)

When entered first in the hierarchical regression analysis, the Phonics component did not predict any significant proportion of the variance in scores on the phonemic awareness variable (TAAS). Stepwise regression was performed to determine which variable(s) was the best predictor. Table 17 shows that the Letter Formation variable accounted for 43.2 % (adjusted $R^2$) of the variance in TAAS posttest scores, $F(1, 18) = 15.43, p < .001$. No other components added to this prediction. Consistent with the earlier finding of a moderate/high correlation of this variable with the TAAS, this regression analysis demonstrates that knowledge of time spent engaged in activities involving letter formation provided valuable information in predicting children's phonemic awareness ability at the end of the year. Possible reasons for this will be addressed in the General Discussion section.

The Phonics variable was also not significant for the Letter Naming task when entered first hierarchically, and so a stepwise regression with the 4 predictors was performed for this measure (see Table 17). Not surprisingly, the Letter Formation component was again the only significant predictor of performance, explaining 15.2% (adjusted $R^2$) of the variance, $F(1, 18) = 4.40, p < .05$.

For the Writing Alphabet Task, the final measure of basic literacy skills, stepwise analysis results showed that none of the predictors was significant. The last entry in Table 17 presents these results.
Table 17.

Literacy Components Predicting Set 3 - Subset A
(Phonemic Awareness and Basic Skills Measures)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mult$R^2$</th>
<th>(adj)$R^2$</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>$R^2$Ch</th>
<th>FCh</th>
<th>$p&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = Rosner Test of Auditory Analysis Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. LForm (stepwise)</td>
<td>46.1</td>
<td>43.2</td>
<td>15.43</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. RdGram</td>
<td>52.2</td>
<td>46.6</td>
<td>9.29</td>
<td>.01</td>
<td>6.1</td>
<td>2.16</td>
<td>ns</td>
</tr>
<tr>
<td>3. StWd</td>
<td>52.2</td>
<td>43.2</td>
<td>5.83</td>
<td>.01</td>
<td>0.0</td>
<td>0.00</td>
<td>ns</td>
</tr>
<tr>
<td>4. Phonics</td>
<td>52.9</td>
<td>40.3</td>
<td>4.20</td>
<td>.05</td>
<td>0.7</td>
<td>0.20</td>
<td>ns</td>
</tr>
</tbody>
</table>

Y = Name Letters

| 1. LForm (stepwise) | 19.7 | 15.2 | 4.40 | .05  |
| 2. StWd            | 27.0 | 18.4 | 3.15 | .10  |
| 3. Phonics         | 27.1 | 13.4 | 1.98 | ns   |
| 4. RdGram          | 27.1 | 7.6  | 1.39 | ns   |

Y = Write Alphabet

| 1. Phonics (stepwise) | 9.7  | 4.7  | 1.94 | ns   |
| 2. RdGram           | 21.3 | 12.1 | 2.31 | ns   |
| 3. LForm            | 31.6 | 18.8 | 2.47 | ns   |
| 4. StWd             | 31.7 | 13.5 | 1.74 | ns   |

Step 1, df = 1, 18
Key: LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar

Literacy Components Predicting Set 3 -- Subset B (Alphabetic Coding Measures)

The next section looked at how well the 4 literacy components predicted performance on the alphabetic coding measures. Table 18 shows that the Phonics category was a strong predictor of both the recognition and the generation of the sounds that letters make. For Letter-Sound Recognition, Phonics explained 34% (adjusted $R^2$) of the variance, $F(1, 18) = 10.77, p < .01$. Letter Formation accounted for an additional 9% (adjusted $R^2$) but this was not significant ($FCh = 2.82, p = .11$).

Both the Phonics and the Letter Formation components were more powerful predictors for the Letter-Sound Recall measure (see Table 18). Phonics accounted for 47.9% (adjusted $R^2$) of
the variance, $F(1, 18) = 18.50, p < .001$, and Letter Formation explained 12.4% (adjusted $R^2_{Ch} = 11.1\%$) of the variance beyond that of the Phonics variable ($F_{Ch} = 5.76, p < .05$).

It is not surprising that the amount of time children spent engaged in alphabetic coding instruction (and other phonic-related activities) strongly predicted how well they would perform on tests measuring these skills. However, it is interesting to note that knowledge of how much of the school day children spent learning/practicing how to form letters provided additional predictive data.

Table 18.

Literacy Components Predicting Set 3 -Subset B
(Alphabetic Coding Measures)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mult $R^2$</th>
<th>(adj)$R^2$</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>$R^2_{Ch}$</th>
<th>$F_{Ch}$</th>
<th>$p&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y =$ Letter/Sound Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>37.4</td>
<td>34.0</td>
<td>10.77</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>46.4</td>
<td>40.1</td>
<td>7.35</td>
<td>.01</td>
<td>9.0</td>
<td>2.82</td>
<td>ns</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>49.6</td>
<td>40.2</td>
<td>5.25</td>
<td>.01</td>
<td>3.2</td>
<td>1.04</td>
<td>ns</td>
</tr>
<tr>
<td>4. StWd</td>
<td>49.6</td>
<td>36.2</td>
<td>3.69</td>
<td>.05</td>
<td>0.0</td>
<td>0.00</td>
<td>ns</td>
</tr>
<tr>
<td>$Y =$ Letter/Sound Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>50.7</td>
<td>47.9</td>
<td>18.50</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>63.1</td>
<td>58.8</td>
<td>14.56</td>
<td>.001</td>
<td>12.4</td>
<td>5.76</td>
<td>.05</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>66.9</td>
<td>60.7</td>
<td>10.77</td>
<td>.001</td>
<td>3.8</td>
<td>1.80</td>
<td>ns</td>
</tr>
<tr>
<td>4. StWd</td>
<td>66.9</td>
<td>58.1</td>
<td>7.58</td>
<td>.01</td>
<td>0.0</td>
<td>0.01</td>
<td>ns</td>
</tr>
</tbody>
</table>

Step 1, df = 1, 18

Key: LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar

Literacy Components Predicting Set 3 - Subset C (Phonemic Analysis Measures)

The final set of analyses examined the prediction of the 4 Literacy Components for the measures of reading and spelling that were scored phonemically (i.e., taking into account the
child's phonemic attempts at decoding and encoding words). Table 19 presents the regressions results for the phonemic scoring of the real word measures, and Table 20 displays an alternate analysis for two of these measures where Phonics was not the best single predictor (explained in more detail below). The nonword tasks results are depicted in Table 21.

**Literacy Components Predicting Phonemic Analysis of Real Words**

When the hierarchical regression was performed on the phonemic analysis scores of the WRAT Reading test, Phonics accounted for 32% (adjusted $R^2$) of the variance, $F(1, 18) = 9.93$, $p < .01$, followed by the Letter Formation component which explained an additional 17% (adjusted $R^2_{Ch} = 15\%)$ of the variance, $F_{Ch} = 6.10$, $p < .05$, for a total of 49% adjusted $R^2$. The Sight Word variable accounted for a further 8.4% (adjusted $R^2_{Ch} = 6.7\%$), but this only approached significance at a .10 probability level.

The setwise regression (used as confirmation -- explained earlier) showed that Phonics was not the best predictor and so a full stepwise analysis was performed where Phonics was not forced into the equation in the initial position. Results showed that Letter Formation was the single best predictor (only slightly higher than Phonics) accounting for 36% (adjusted $R^2$) of the variance. However, when Letter Formation was followed by Sight Word, an additional 20.7% (adjusted $R^2_{Ch} = 19.4\%$) of the variance was explained. This total of 55.4% (adjusted $R^2$) explained variance for these two components indicates that although they share much of the variance explained by the Phonics predictor, they are also related to some unique aspect of performance on this test which accounts for better prediction (see Table 20 for this second, alternate analysis).

A similar situation occurred in the regression analysis of the phonemic scoring for the Burns and Roe Word Recognition test. Table 19 shows that when Phonics was the first predictor forced into the hierarchical analysis, its prediction was very weak 14.7% (adjusted $R^2$), $F(1, 18)$ $= 4.27$, $p < .10$; however, setwise analysis confirmed that Phonics was the best single predictor. The setwise analysis also indicated that the best combination of predictors was Letter Formation and Sight Word. In the alternate analysis, when entered first (setwise) as a group, these two components proved to be more powerful predictors explaining 28.4% of the variance (adjusted
\(R^2\), \(F(1, 18) = 4.77, p < .05\) (see Table 20 for alternate analysis), suggesting a relatively heavy influence of children's sightword lexicon on word recognition tasks involving real words. After the Letter Formation and Sight Word group entry, stepwise analysis was used to select the strongest predictor(s) of the remaining two variables. None was significant.

In both the Set 2 and Set 3 regression analyses (i.e., full word raw scores, and phonemic analyses) involving the two word recognition measures, the variance explained by the Phonics predictor was much greater for the WRAT than for Burns and Roe. Aside from the fact that the Burns and Roe Word Recognition measure appears to be a somewhat unstable variable (mentioned earlier), there may be another explanation for the discrepancy in size of Phonics prediction for these two apparently similar measures. It may occur because of more rigorous development (i.e., test construction) of the WRAT Reading standardized test. That is, the list of words used for this measure underwent item analysis, and thus it more keenly discriminates between good and poor readers than the list of (high frequency) words for the Burns and Roe Informal Reading Inventory. Thus, the higher discrimination value of the WRAT Reading word list, may provide a "tighter" measure for the prediction of word recognition.

The final regression analysis for the phonemic scoring of the real word measures, shows that Phonics proved to be a strong predictor of performance for WRAT Spelling -- Phonemic Analysis, accounting for 36% (adjusted \(R^2\)) of the variance, \(F(1, 18) = 11.68, p < .01\). The addition of 9.3% (adjusted \(R^2Ch = 6.7\%\)) to the prediction by the Letter Formation variable only approached significance at the .10 probability level (see Table 19). No other category added significantly to the prediction.
Table 19.

Literacy Components Predicting Set 3 - Subset C
(Phonemic Analysis - Real Word Measures)

<table>
<thead>
<tr>
<th>Variable</th>
<th>MultR²</th>
<th>(adj)R²</th>
<th>F</th>
<th>p&lt;</th>
<th>$R^2$Ch</th>
<th>FCh</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y = WRAT Reading - Phonemic Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>35.6</td>
<td>32.0</td>
<td>9.93</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>52.6</td>
<td>47.0</td>
<td>9.44</td>
<td>.01</td>
<td>17.0</td>
<td>6.10</td>
<td>.05</td>
</tr>
<tr>
<td>3. StWd</td>
<td>61.0</td>
<td>53.7</td>
<td>8.35</td>
<td>.001</td>
<td>8.4</td>
<td>3.46</td>
<td>.10</td>
</tr>
<tr>
<td>4. RdGram</td>
<td>62.0</td>
<td>51.9</td>
<td>6.12</td>
<td>.01</td>
<td>1.0</td>
<td>0.38</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = Burns and Roe - Phonemic Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>19.2</td>
<td>14.7</td>
<td>4.27</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>25.0</td>
<td>16.2</td>
<td>2.84</td>
<td>.10</td>
<td>5.8</td>
<td>1.32</td>
<td>ns</td>
</tr>
<tr>
<td>3. StWd</td>
<td>36.0</td>
<td>23.9</td>
<td>2.99</td>
<td>.10</td>
<td>11.0</td>
<td>2.72</td>
<td>ns</td>
</tr>
<tr>
<td>4. RdGram</td>
<td>39.3</td>
<td>23.2</td>
<td>2.43</td>
<td>.10</td>
<td>3.2</td>
<td>0.83</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Y = WRAT Spelling - Phonemic Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phonics (forced)</td>
<td>39.4</td>
<td>36.0</td>
<td>11.68</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Variables (stepwise)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LForm</td>
<td>48.7</td>
<td>42.7</td>
<td>8.08</td>
<td>.01</td>
<td>9.3</td>
<td>3.10</td>
<td>.10</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>54.2</td>
<td>45.6</td>
<td>6.31</td>
<td>.01</td>
<td>5.5</td>
<td>1.90</td>
<td>ns</td>
</tr>
<tr>
<td>4. StWd</td>
<td>56.2</td>
<td>44.5</td>
<td>4.81</td>
<td>.05</td>
<td>2.0</td>
<td>0.69</td>
<td>ns</td>
</tr>
</tbody>
</table>

Step 1, df = 1, 18

Key: LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar

Table 20 presents the "alternate" analyses for the phonemic scoring of the WRAT Reading and Burns and Roe Word Recognition tests. That is, these are the results of setwise and stepwise regressions for these two outcome variables when Phonics was not entered first (and therefore, its "shared" variance was explained by earlier entries).
Table 20.
Alternate Analysis for Literacy Components Predicting Phonemic Analysis of WRAT Reading and Burns and Roe Word Recognition

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mult$R^2$ (adj)$R^2$</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>$R^2$Ch</th>
<th>$F$Ch</th>
<th>$p&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y = WRAT Reading - Phonemic Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. LForm (stepwise)</td>
<td>39.4</td>
<td>36.0</td>
<td>11.68</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. StWd</td>
<td>60.1</td>
<td>55.4</td>
<td>12.78</td>
<td>.001</td>
<td>20.7</td>
<td>8.82</td>
</tr>
<tr>
<td>3. RdGram</td>
<td>61.1</td>
<td>53.7</td>
<td>8.36</td>
<td>.001</td>
<td>1.0</td>
<td>0.41</td>
</tr>
<tr>
<td>4. Phonics</td>
<td>62.0</td>
<td>51.9</td>
<td>6.12</td>
<td>.01</td>
<td>0.9</td>
<td>0.37</td>
</tr>
</tbody>
</table>

| **Y = Burns and Roe - Phonemic Analysis** |                      |      |       |         |       |       |
| 1. LForm (setwise) | 35.9                  | 28.4 | 4.77  | .05     |       |       |
| StWd              |                       |      |       |         |       |       |
| Subsequent Variables (stepwise) |                      |      |       |         |       |       |
| 2. RdGram         | 39.3                  | 27.9 | 3.46  | .05     | 3.4   | 0.90  | ns    |
| 3. Phonics        | 39.3                  | 23.2 | 2.43  | .10     | 0.0   | 0.01  | ns    |

Step 1, df = 1, 18

*Key:* LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar

It appears that for measures assessing the degree of phonemic analysis of real word reading, all three of the above variables (Phonics, Letter Formation, and Sight Word) play a role in predicting performance. Statistically speaking, how these variable are combined, and which ones are entered first into the regression equation governs how much relative predictive power they exert. For practical purposes, however, it may be sufficient to be aware that the amount of time children spent engaged in all three of these literacy activities was related to how many phonemes they recognized when reading real words. For spelling, time spent in Sight Word activities was not related to how many phonemes of real words children were able to produce, suggesting that lack of explicit word analysis in Sight Word activities may have contributed to lack of transfer. That is, the emphasis of learning words as a whole (with no internal analysis) to build a sight word vocabulary, may not be sufficient to consolidate the spellings in memory well enough to be able to recall them when needed for purposes of spelling.
Literacy Components Predicting Phonemic Analysis of Nonwords

Table 21 shows that on the three nonword tasks, predictions were more powerful, and Phonics was clearly the most important predictor (setwise analysis confirmed this). For the phonemic scoring of the Woodcock Word Attack test, the Phonics component was the only significant predictor explaining a full 55.7% (adjusted $R^2$) of the variation in performance, $F(1, 18) = 24.92, p < .001$. On the Nonword Reading Task, the Phonics variable was a very potent predictor, accounting for 65.7% (adjusted $R^2$) of the variance. Only an additional 8.8% (adjusted $R^2_{Ch} = 6.2\%$) was explained by the combination of the Letter Formation and Sight Word variables entered as a group (as indicated by the setwise analysis), and this effect only approached significance, $F_{Ch} = 2.97, p < .10$. On the Nonword Spelling Task -- Phonemic Analysis, Phonics again was the main predictor, accounting for 43.6% (adjusted $R^2$) of the variance, $F(1, 18) = 15.71, p < .001$. The Reading/Grammar variable provided a slight improvement to the prediction by explaining an additional 9.2% (adjusted $R^2$ change =7%) of the variance beyond that of the Phonics component, but this only approached significance $F_{Ch} = 3.53, p < .10$. 

113
Table 21.

**Literacy Components Predicting Set 3 - Subset C (Phonemic Analysis - Nonword Measures)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\text{Mult} R^2$</th>
<th>$\text{adj})^2 R^2$</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>$R^2_{\text{Ch}}$</th>
<th>$F_{\text{Ch}}$</th>
<th>$p&lt;$</th>
</tr>
</thead>
</table>

**Y = Word Attack -- Phonemic Analysis**

1. Phonics (forced) | 58.1 | 55.7 | 24.92 | .001 |

Subsequent Variables (stepwise)

2. RdGram | 62.0 | 57.5 | 13.86 | .001 | 3.9 | 1.74 | ns |
3. LForm | 63.8 | 57.0 | 9.38 | .001 | 1.8 | 0.79 | ns |
4. StWd | 66.7 | 57.9 | 7.52 | .01 | 2.9 | 1.35 | ns |

**Y = Nonword Reading -- Phonemic Analysis**

1. Phonics (forced) | 67.5 | 65.7 | 37.47 | .001 |

Subsequent Variables (stepwise)

2. LForm (setwise) StWd | 76.3 | 71.9 | 17.17 | .001 | 8.8 | 2.97 | .10 |
3. RdGram | 77.1 | 71.0 | 12.62 | .001 | 0.8 | 0.52 | ns |

**Y = Nonword Spelling -- Phonemic Analysis**

1. Phonics (forced) | 46.6 | 43.6 | 15.71 | .001 |

Subsequent Variables (stepwise)

2. RdGram | 55.8 | 50.6 | 10.74 | .001 | 9.2 | 3.53 | .10 |
3. LForm | 60.0 | 52.5 | 8.00 | .01 | 4.2 | 1.66 | ns |
4. StWd | 66.4 | 57.4 | 7.41 | .01 | 6.4 | 2.86 | ns |

Step 1, df = 1, 18

**Key:** LForm = Letter Formation, StWd = Sight Word, RdGram = Real Reading and Grammar

On these measures assessing children's attempts to decode and encode nonwords, it is clear that, by far, the most important predictor of how well children performed was knowledge of how much time they had spent in phonics-related activities. The influence of other categories in these predictions was negligible.
To summarize thus far, as expected, the Phonics category was the most powerful predictor of scores (as confirmed by the setwise analyses) on most of the posttest measures, with percentages of variance explained ranging from 15% to 66%. Phonics did not predict performance on the phonemic awareness measure (the Rosner Test of Auditory Analysis Skills), nor on tasks involving basic literacy skills (i.e., letter naming and letter writing); but on measures tapping the more complex tools needed for reading and spelling (i.e., alphabetic coding), as well as tasks reflecting the application of those tools (i.e., word recognition and analysis), the Phonics component proved to be a consistent predictor of performance (with the two exceptions described below). In other words, the larger the proportion of the school day that teachers in the experimental and control classes spent emphasizing letter-sound correspondences, word analysis, and word spellings, the better the children were able to learn and apply these literacy skills.

Explicit phonics instruction consistently stood out as the key factor related to children's success on almost every literacy measure, and was especially powerful in its connection to performance on the purer measures of word analysis (i.e., nonword measures, where sightword memory cannot enhance performance). The next most important element found to contribute to reading and writing success was the Literacy Component involving the explicit instruction and practice of forming letters.

Often, the Letter Formation component supplemented significantly the amount of explained variance (usually at the .10 probability level of significance), and in two instances (i.e., the Rosner Test of Auditory Analysis Skills and the Letter Name Task), it was the only predictor. That the Letter Formation component predicted letter naming performance is not unexpected. However, the strong prediction properties of the Letter Formation category with regard to children's phonemic awareness ability (explaining 43% [adjusted $R^2$] of the variance) was somewhat more surprising. This finding strongly suggests that there is some element(s) within the activities associated with Letter Formation which makes the phonemic makeup of words more salient (see General Discussion for elaboration).

For the phonemic analysis scoring of the real word reading measures (i.e., WRAT Reading and Burns and Roe Word Recognition), Letter Formation played a more significant role when the
Sight Word component was included. Together, these two variables were somewhat better predictors of scores than was the Phonics component alone (or along with any other predictor), suggesting that the practice of forming letters may aid in emphasizing real word spellings when combined with sight word activities. In any event, it appears that time can be profitably spent involving children in the types of activities represented by all three of these Literacy Components (Phonics, Letter Formation, and Sight Word).

In only one instance did the fourth variable (i.e., Reading/Grammar) in the set of predictors add information to regression predictions. On the Nonword Spelling -- Phonemic Analysis task, Reading/Grammar made a marginal contribution (i.e., quasi-significant, \( p < .10 \)) to the proportion of explained variance.

It is evident from the regression analyses that the Phonics category was a very potent predictor of achievement on literacy measures; but the term Phonics is a fairly broad categorization encompassing a number of activities represented by "subcategories" (as mentioned in the Method section, most main categories consisted of subcategories so that finer discriminations within activities could be made). So it was thought that a closer examination of these subcategories of the Phonics category would lend further insight into its prediction properties.

**Phonics Subcategories**

The Phonics main category was composed of 5 subcategories\(^{13}\) designed to capture all the phonics-related activities observed. These were: Oral Spelling (OralSpell -- involves only oral spelling without influence of print, e.g., children chanting a spelling of a word); Practice Spelling (PracSpell -- involves children repeating letter names while looking at a word, but no in-depth word analysis); Letter-Sound Correspondence (L/Sound -- involves any direct learning of letters sounds), Word Analysis (WordAn -- involves any direct word analysis when occurring with print,

---

\(^{13}\) Recall that although Letter Name Learning and Sight Words were listed on the observation protocol under the Phonics heading for easier recording, these Literacy Components were considered separate non-phonics main categories and treated as such in all analyses.
e.g., sounding out words, "making words", word families, etc.), and Jolly Phonics Actions (JPActions -- marked when any observation of the Jolly Phonics program's actions representing letter-sounds occurred).

The following analyses show some significant differences in the time Jolly Phonics and control classes spent on the various activities representing the subcomponents of the Phonics category. It is important to keep in mind that, as mentioned earlier, occasionally, the same observation (i.e., scan of 4 children -- see Method section for elaboration of procedure) could be scored in more than one subcategory (e.g., Jolly Phonics Actions and Letter-Sound Correspondence) of the main Literacy Component category. However, when calculating the total percentages for main Literacy Components, any dual (or multiple) recordings of the same observation within a main Literacy Component category were counted as only one observation. So, although the percentages of time spent on the various subcategory activities within a main category may overlap, they do not inflate the main Literacy Component category total, since they are accounted for only once.

Group Comparison of Time Spent in Phonics Subcategories

A MANOVA was performed to compare the Jolly Phonics and control classes (i.e., class means) with regard to the amount of time (i.e., percentage of the school day) spent on the various Phonics subcategories. One of these subcategories was not included in the MANOVA (i.e., JPActions) because it was unique to the Jolly Phonics program. The MANOVA was significant $F(4, 15) = 5.84, p < .01$, with two subcategory activities contributing to the results -- Letter-Sound Correspondence, $F(1, 18) = 15.75, p < .001$, effect size = 3.98, and Word Analysis, $F(1, 18) = 19.15, p < .001$, effect size = 8.88 (see Table 22). The Jolly Phonics classes spent an average of 5.15% (SD = 3.33) of the school day (i.e., approximately 8 minutes) participating in activities involving explicit learning of letter-sound correspondences, compared with 0.75% (SD = 1.12) of the day (i.e., approximately 1 minute) for the control classes. Direct word-analysis activities comprised 4.96% (SD = 3.18) of the day (i.e., approximately 8 minutes) for the Jolly Phonics group. The control classes averaged just 0.52% (SD = 0.50) of their day (i.e., < 1 minute).
involved in any type of word-analysis activity. There were no significant differences in time spent in oral or print-related spelling activities (which was minimal for both groups). The Jolly Phonics classes spent an average of 2.33% (SD = 2.29) of the day (i.e., approximately 4 minutes) learning and performing the actions (unique to the Jolly Phonics program) that go along with the letter-sound correspondences.

Table 22.
Percent of School Day Spent in Phonics Subcategory Activities:
Jolly Phonics Classes vs. Control Classes

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 10)</th>
<th>Control (n = 10)</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>OralSpell</td>
<td>0.18% 0.42</td>
<td>0.15% 0.32</td>
<td>0.03</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>PracSpell</td>
<td>1.00% 0.88</td>
<td>0.52% 0.72</td>
<td>1.81</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>L/Sound</td>
<td>5.15% 3.33</td>
<td>0.75% 1.12</td>
<td>15.75</td>
<td>.001</td>
<td>3.93</td>
</tr>
<tr>
<td>WordAn</td>
<td>4.96% 3.18</td>
<td>0.52% 0.50</td>
<td>19.15</td>
<td>.001</td>
<td>8.88</td>
</tr>
</tbody>
</table>

MANOVA, F(4, 15) = 5.84, p < .01.
Key: OralSpell = Oral Spelling, PracSpell = Practice Spelling, L/Sound = Letter-Sound Correspondence, WordAn = Word Analysis

Intracorrelations, and Correlations with Outcome Measures

To see how the Phonics subcategories were related to each other, and with Phase 2 posttest performance on outcome measures, Pearson correlational analyses were performed using class means as the unit of analysis -- N = 20 for all but the JPActions subcategory. Since the JPActions subcategory was unique to the Jolly Phonics group, all correlational analyses for this particular variable were performed on just the 10 (experimental) class means. There are certain difficulties associated with examining correlations with such a small N. For example, correlations need to be quite substantial (i.e., r = .632) in order to reach significance; in addition, caution should be taken when interpreting these correlations since the small N may render the variable unstable (Shavelson,
Nevertheless, as will be seen in the following analyses, the relative size\(^{14}\) and the consistency of the associations of the JPActions with various outcome measures, make this component a factor worth exploring.

**Intracorrelation Among Phonics Subcategories**

Table 23 shows the intracorrelations among the Phonics subcategories. The Letter-Sound Correspondence subcategory was strongly associated with the Word Analysis subcategory \((r = .75, p < .001)\). Jolly Phonics Actions (JPActions) and the Letter-Sound Correspondence subcategory were also correlated \((r = .52, p = .11)\), however, since this analysis had an \(N\) of only 10 (see previous paragraph for explanation), the correlation did not meet the stricter criterion for significance. These associations are not surprising since many phonics activities involved a combination of these 3 components. There was a weak positive correlation of the Word Analysis with the Practice Spelling subcategory; however, this only approached significance \((r = .44, .05 < p < .10)\).

**Table 23.**

<table>
<thead>
<tr>
<th></th>
<th>OralSpell</th>
<th>PracSpell</th>
<th>L/Sound</th>
<th>WordAn</th>
</tr>
</thead>
<tbody>
<tr>
<td>PracSpell</td>
<td>0.128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/Sound</td>
<td>-0.289</td>
<td>0.097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WordAn</td>
<td>-0.112</td>
<td>0.436</td>
<td>0.748***</td>
<td></td>
</tr>
<tr>
<td><strong>JPActions a</strong></td>
<td>-0.311</td>
<td>-0.512</td>
<td>0.523</td>
<td>-0.076</td>
</tr>
</tbody>
</table>

Note: *\(p < .05; \; **p < .01; \; ***p < .001.\)

**Key:** OralSpell = Oral Spelling, PracSpell = Practice Spelling, L/Sound = Letter-Sound Correspondence, WordAn = Word Analysis, JPActions = Jolly Phonics Actions

\(a\) JPActions is unique to the Jolly Phonics group and so for correlations with this variable, \(N = 10, df = 8,\) and critical \(r = .652,\) for \(p = .05;\) therefore, correlations must be higher to be significant.

---

\(^{14}\) Because of the issue associated with the small \(N\) for the JPActions correlations, descriptions are modified when discussing what would normally be considered high correlations (e.g., \(r = .762)\).
The next set of analyses was performed to examine how time spent in these various Phonics subcategories related to performance on the three sets of outcomes measures.

**Phonics Subcategories with Set 1 (Standard Score Measures)**

Correlations with the standard score measures reveal that JPActions was correlated with all 3 tests: Woodcock Word Attack \( r = .68, p < .05 \); WRAT Reading \( r = .78, p < .01 \); and WRAT Spelling, \( r = .79, p < .01 \) (see Table 24). Letter-Sound Correspondence was moderately associated with the Woodcock Word Attack test \( r = .50, p < .05 \), as was the Word Analysis subcategory \( r = .47, p < .05 \). The Letter-Sound Correspondence associations with the other two standard measures (WRAT Reading and Spelling) only approached significance \( r = .42, .05 < p < .10 \), and \( r = .42, .05 < p < .10 \), respectively). Thus, the more classroom time children spent learning letter-sound correspondences during the year, particularly when these correspondences were accompanied/emphasized by actions (for the Jolly Phonics children), the better children scored on measures of reading and spelling at the end of the school year.
Table 24.

Correlations Between Phonics Subcategories and Set 1 Measures (Standard Scores)
(N = 20, but for JPActions N = 10)

<table>
<thead>
<tr>
<th></th>
<th>WaSS</th>
<th>WrRdSS</th>
<th>WrSpSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OralSpell</td>
<td>-0.081</td>
<td>-0.057</td>
<td>-0.114</td>
</tr>
<tr>
<td>PracSpell</td>
<td>0.109</td>
<td>-0.091</td>
<td>-0.168</td>
</tr>
<tr>
<td>L/Sound</td>
<td>0.495*</td>
<td>0.418</td>
<td>0.421</td>
</tr>
<tr>
<td>WordAn</td>
<td>0.470*</td>
<td>0.307</td>
<td>0.295</td>
</tr>
<tr>
<td><strong>JPActions</strong></td>
<td>0.683*</td>
<td>0.783**</td>
<td>0.787**</td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < .01; ***p < .001.

Key:
Phonics Subcategories
OralSpell = Oral Spelling, PracSpell = Practice Spelling, L/Sound = Letter-Sound Correspondence, WordAn = Word Analysis, JPActions = Jolly Phonics Actions
a JPActions is unique to the Jolly Phonics group and so for correlations with this variable, N = 10, df = 8, and critical r = .652, for p = .05; therefore, correlations must be higher to be significant.

Phonics Subcategories with Set 2 (Full Word Raw Score Measures)

Correlational analysis of the Phonics subcategories with the Set 2 (Full Word Raw Score) measures again revealed associations of JPActions to all the outcome scores of this set (see Table 25). This analysis examined the associations of the subcomponents to the "purer" measures of real word and nonword reading and spelling (that is, measures that reflect just how many words were correct -- the letter-name portions of these tests were not included). Results showed positive associations for JPActions ranging from $r = .71, p < .05$ with the three real word measures (i.e., WRAT Reading - Full Words, Burns and Roe Word Recognition - Full Words, and WRAT Spelling - Full Words) to $r = .80, p < .01$ with the raw scores of the Woodcock Word Attack test. All of the Letter-Sound Correspondence correlations with the Set 2 measures were significant (ranging from $r = .45, p < .05$ to $r = .55, p < .05$). The only significant correlations for the Word Analysis subcategory were with the two nonword measures -- Woodcock Word Attack - Raw Scores $r = .48, p < .05$, and the Nonword Spelling Task, $r = .47, p < .05$. The association of
Word Analysis with the WRAT Reading - Full Words measure only approached significance ($r = .43$, $0.05 < p < 0.10$).

As mentioned earlier, the three standardized tests of Set 1, and these 5 measures in Set 2, are rather conservative in their scoring, since there are no "part marks" for close attempts at decoding and encoding. Nevertheless, many of the kindergarten children were able to decode and encode some of these words, and the more time experimental and control classes spent in activities explicitly emphasizing letter-sounds, the more successful were children's performances on these conservative measures. Moreover, activities which involved a physical means of highlighting letter-sound relationships (as found in the Jolly Phonics classrooms) were consistently related to how well the children were able to perform.

**Table 25.**

<table>
<thead>
<tr>
<th>Phonic Subcategories</th>
<th>WrDrdWs</th>
<th>B&amp;R</th>
<th>WrSpWds</th>
<th>WaRS</th>
<th>NwRd</th>
</tr>
</thead>
<tbody>
<tr>
<td>OralSpell</td>
<td>-0.063</td>
<td>-0.152</td>
<td>-0.118</td>
<td>-0.176</td>
<td>-0.159</td>
</tr>
<tr>
<td>PracSpell</td>
<td>0.153</td>
<td>-0.093</td>
<td>0.089</td>
<td>0.083</td>
<td>0.124</td>
</tr>
<tr>
<td>L/Sound</td>
<td>0.505*</td>
<td>0.467*</td>
<td>0.449*</td>
<td>0.551*</td>
<td>0.505*</td>
</tr>
<tr>
<td>WordAn</td>
<td>0.430</td>
<td>0.351</td>
<td>0.365</td>
<td>0.483*</td>
<td>0.474*</td>
</tr>
<tr>
<td>JPActions*</td>
<td>0.714*</td>
<td>0.716*</td>
<td>0.713*</td>
<td>0.798**</td>
<td>0.734*</td>
</tr>
</tbody>
</table>

Note: *$p < .05$; **$p < .01$; ***$p < .001$.

**Key:**

Phonics Subcategories
- OralSpell = Oral Spelling, PracSpell = Practice Spelling, L/Sound = Letter-Sound Correspondence, WordAn = Word Analysis, JPActions = Jolly Phonics Actions

Set 2
- WrDrdWs = WRAT Reading Full Words, B&R = Burns and Roe Word Recognition, WrSpWds = WRAT Spelling Full Words, WaRS = Woodcock Word Attack Raw Score, NwRd = Nonword Reading Task

JPActions is unique to the Jolly Phonics group and so for correlations with this variable, $N = 10$, df = 8, and critical $r = .652$, for $p = .05$; therefore, correlations must be higher to be significant.
Phonics Subcategories with Set 3
(Phonemic Awareness, Basic Skills, Coding, and Phonemic Analysis Measures)

For the most part, results similar to those found in the Set 1 and 2 measures were evident when the correlations among Phonics subcategories and the Set 3 outcome measures were examined. However, some interesting new information came to light as well (see Table 26).

Phonics Subcategories with Set 3 -- Subset A (Phonemic Awareness and Basic Skills Measures)

In the previous correlational analyses of the main Literacy Components with the TAAS (Rosner Test of Auditory Analysis Skills) and basic skills measures, no significant correlations had been found for the Phonics category as a whole. However, when associations of the subcategories of the Phonics component with these measures were analyzed, a different pattern of results emerged. The JPActions subcategory showed a significant correlation with the TAAS ($r = .68, p < .01$) suggesting that the JPActions subcomponent is an active "phonics" ingredient in the facilitation of phonemic awareness, and that this association was masked by the inclusion of all the other subcategories in the broader analyses of main Literacy Component categories. JPActions was also significantly correlated with the Letter Name Task ($r = .81, p < .01$), indicating that although the Jolly Phonics program does not spend much time teaching letter names, the children pick these up just the same. None of the other subcomponents was significantly correlated with the measures of phonemic awareness and basic literacy skills.

Phonics Subcategories with Set 3 -- Subset B (Alphabetic Coding Measures)

Both the Letter-Sound Correspondence and JPActions subcategories demonstrated powerful associations with the measures assessing ability to recognize and generate letter-sound relationships (more complex literacy skills). However, JPActions correlations were the strongest (the issues regarding small sample size for this variable notwithstanding). In fact, its correlation with Letter-Sound Recall (the more difficult of the two tasks) was the highest of all the correlations of any measures and components examined ($r = .93$) reaching a significance level of .001 for this small $N$. JPActions' correlation with Letter-Sound Recognition was $r = .89, p < .001$. For the
Letter-Sound Correspondence subcategory (which focuses specifically on just this type of activity), correlations with Letter-Sound Recall were moderate/high ($r = .66, p < .01$), and with Letter-Sound Recognition they were moderate ($r = .56, p < .01$). The Word Analysis subcategory was also significantly correlated with Letter-Sound Recall ($r = .46, p < .05$). It is not surprising that when children spent more time learning letter-sound relationships, they performed more successfully on measures assessing this knowledge. However, the strong associations of the physical actions (i.e., JPActions) with these alphabetic coding scores, suggests an additional benefit of highlighting letter-sound correspondences in a multi-modal way. It may be that the more ways available to children for processing information, the better this information will be learned and remembered when needed.

**Phonics Subcategories with Set 3 -- Subset C (Phonemic Analysis Measures)**

Table 26 shows that for the phonemic analysis of the outcome reading and spelling measures, the Letter-Sound Correspondence subcategory showed the most consistent significant positive associations (with five of the six measures). Correlations with the highest significance were with the nonword measures. Although most of the correlations of the JPActions subcategory with these phonemic analysis measures were at a higher value than those demonstrated by the Letter-Sound Correspondence subcategory, they were not all significant (for the reasons of small $N$ explained earlier), and, therefore, cannot be considered as strong. Correlations for the Letter-Sound Correspondence subcategory ranged from $r = .52, p < .05$ (WRAT Reading -- Phonemic Analysis), to $r = .72, p < .001$, (Nonword Reading Task -- Phonemic Analysis). The JPActions' correlations ranged from $r = .50, ns$ (Burns and Roe Word Recognition -- Phonemic Analysis), to $r = .78, p < .01$, (WRAT Spelling -- Phonemic Analysis); however, only three of these can be considered significant. Two of these three significant correlations were with the nonword measures.

Not surprisingly, Word Analysis was also significantly correlated with many of the measures, and these coefficients ranged from $r = .48, p < .05$ with WRAT Spelling -- Phonemic Analysis, to $r = .68, p < .001$ with Nonword Reading Task -- Phonemic Analysis. Correlations
only approached significance with the phonemic analysis scoring on the WRAT Reading and the Burns and Roe Word Recognition measures -- \( r = .41, .05 < p < .10 \), and \( r = .39, .05 < p < .10 \), respectively). Three of the four significant Word Analysis correlations were also with the nonword measures.

Results of this set of correlational analyses examining the degree to which children could analyze and generate spellings, indicate that time spent in the above three literacy activities (which focused on phoneme/grapheme instruction, explicit analysis of words, and, for the Jolly Phonics group, accompanying actions reinforcing the letter-sound mappings) was closely connected to how well children were able to apply alphabetic coding skills, particularly to the analysis of nonwords.
Table 26.

Correlations Between Phonics Subcategories and Set 3 Measures
(Phonemic Awareness, Literacy Skills and Phonemic Analysis)
(N = 20, but for JPActions N = 10)

<table>
<thead>
<tr>
<th>Subset A: Phonemic Awareness and Basic Skills</th>
<th>TAAS</th>
<th>LtNm</th>
<th>WrtAlph</th>
</tr>
</thead>
<tbody>
<tr>
<td>OralSpell</td>
<td>-0.074</td>
<td>-0.107</td>
<td>0.044</td>
</tr>
<tr>
<td>PracSpell</td>
<td>0.150</td>
<td>-0.100</td>
<td>-0.018</td>
</tr>
<tr>
<td>L/Sound</td>
<td>0.240</td>
<td>0.292</td>
<td>0.315</td>
</tr>
<tr>
<td>WordAn</td>
<td>0.105</td>
<td>0.167</td>
<td>0.282</td>
</tr>
<tr>
<td>JPActions(^a)</td>
<td>0.675(^*)</td>
<td>0.806(^**)</td>
<td>0.230</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subset B: Alphabetic Coding</th>
<th>L/SRecog</th>
<th>L/Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>OralSpell</td>
<td>-0.058</td>
<td>-0.047</td>
</tr>
<tr>
<td>PracSpell</td>
<td>0.045</td>
<td>0.077</td>
</tr>
<tr>
<td>L/Sound</td>
<td>0.562(^**)</td>
<td>0.655(^**)</td>
</tr>
<tr>
<td>WordAn</td>
<td>0.376</td>
<td>0.456(^*)</td>
</tr>
<tr>
<td>JPActions(^a)</td>
<td>0.892(^***)</td>
<td>0.925(^***)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subset C: Phonemic Analysis</th>
<th>WrRdPh</th>
<th>B&amp;RPh</th>
<th>WrSpPh</th>
<th>WaPh</th>
<th>Nonwords</th>
<th>NwRdPh</th>
<th>NwSpPh</th>
</tr>
</thead>
<tbody>
<tr>
<td>OralSpell</td>
<td>-0.010</td>
<td>-0.108</td>
<td>-0.110</td>
<td>-0.003</td>
<td>-0.155</td>
<td>-0.146</td>
<td></td>
</tr>
<tr>
<td>PracSpell</td>
<td>0.236</td>
<td>0.243</td>
<td>0.188</td>
<td>0.312</td>
<td>0.308</td>
<td>0.284</td>
<td></td>
</tr>
<tr>
<td>L/Sound</td>
<td>0.520(^*)</td>
<td>0.333</td>
<td>0.536(^*)</td>
<td>0.646(^**)</td>
<td>0.720(^***)</td>
<td>0.556(^*)</td>
<td></td>
</tr>
<tr>
<td>WordAn</td>
<td>0.406</td>
<td>0.393</td>
<td>0.480(^*)</td>
<td>0.642(^**)</td>
<td>0.681(^***)</td>
<td>0.596(^**)</td>
<td></td>
</tr>
<tr>
<td>JPActions(^a)</td>
<td>0.618</td>
<td>0.495</td>
<td>0.776(^**)</td>
<td>0.595</td>
<td>0.696(^*)</td>
<td>0.660(^*)</td>
<td></td>
</tr>
</tbody>
</table>

Note: \(^*\)p < .05; \(^**\)p < .01; \(^***\)p < .001.

Key: Phonics Subcategories
OralSpell = Oral Spelling, PracSpell = Practice Spelling, L/Sound = Letter-Sound Correspondence, WordAn = Word Analysis, JPActions = Jolly Phonics Actions

Set 3
Subset A: TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, WrtAlph = Write Alphabet Task
Subset B: L/SRecog = Letter-Sound Recognition Task, L/Recall = Letter-Sound Recall Task
Subset C: WrRdPh = WRAT Reading -- Phonemic Analysis, B&RPh = Burns and Roe Word Recognition -- Phonemic Analysis, WrSpPh = WRAT Spelling -- Phonemic Analysis, WaPh = Woodcock Word Attack -- Phonemic Analysis, NwRdPh = Nonword Reading Task -- Phonemic Analysis, NwSpPh = Nonword Spelling Task -- Phonemic Analysis

JPActions is unique to the Jolly Phonics group and so for correlations with this variable, N = 10, df = 8, and critical r = .652, for p = .05; therefore, correlations must be higher to be significant.
In sum, findings from the correlational analyses of the Phonics subcategories suggest that there are two main specific components, as well as a third, less influential factor, which are very closely linked to the acquisition of reading and spelling skills, as well as to their application. The component demonstrating the most and the strongest connections with the outcome measures was the JPActions subcategory (which occurred only in the experimental classes). As mentioned in the beginning of this section (i.e., "Phonics Subcategories") due to the small N of the JPActions correlations, there is some concern about the stability of this variable, and so discussion of high correlations should be qualified. Nevertheless, the consistently "high" associations of the JPActions with various outcome measures (especially for the alphabetic coding measures), warrant further investigation of this factor's relation to reading acquisition. For the experimental and control classes, explicit letter-sound instruction was also highly correlated with several performance scores. The associations of the third factor, Word Analysis, were apparent mainly with the phonemic analysis of the nonwords tasks.

Anecdotal perceptions from the classroom observations noted that when systematic alphabetic coding instruction (Letter-Sound Correspondence) was delivered in the context of a story accompanied by relevant physical actions (JPActions) to represent specific letter-sounds, children appeared enthused and motivated. This combination of activities not only emphasized and reinforced the "lesson" (i.e., the particular letter-sound being introduced), but it provided a more meaningful experience for the children than just learning which letter makes a particular sound. The actions, in a sense, made what would otherwise have been essentially meaningless (i.e., the sound of a phoneme) into something that was meaningful. In most cases, the actions could be described as both promoting the correct articulation of the phoneme, as well as providing a "meaning" that could act as a mediator, reinforcing the symbol-sound association. When seen in this light, the strong associations of these activities with reading and spelling performance is not entirely unexpected. The more time children spent in these activities, the more apt they were able to remember the letters and their sound, read whole words and nonwords, as well as spell a number of words. The fact that many of these children were able to make good phonemic attempts when trying to read and spell pseudowords, attests to their ability to apply what they had learned.
The time spent learning and performing the Jolly Phonics letter-sound actions was also related to phonemic awareness scores (on the TAAS), suggesting that this activity may not only aid in remembering phoneme/grapheme correspondences, but it may also draw attention to the phonemic makeup of words. Speculations aside, it was clear that the more time children were exposed to all forms of explicit phoneme/grapheme instruction with related reinforcing activities, the more likely they were able to transfer the knowledge they gleaned from these experiences to tasks requiring them to apply these skills.

PART 4
At-Risk Subsample

This final section examines a subset of the sample consisting of children considered to be "at risk" for future reading difficulties. Generally, programming for this population consists of individual tutoring, aimed at raising children's skill level to more closely match that of children considered to be average. The current analysis was undertaken to determine whether the at-risk children in the Jolly Phonics group benefited sufficiently (i.e., to meet an "average" criterion) from a teaching method designed to be part of regular classroom programming for all children. After Phase 1 (pretest) comparisons to confirm at-risk designation, two main comparisons of outcome achievement were made. The performance of the at-risk (AR) children in the Jolly Phonics group was first compared with the AR control children's performance. The second comparison examined outcome scores of the AR Jolly Phonics group with control children who had been designated as "average" at Phase 1 pretesting.

Phase 1 (Pretest)
At-Risk Subsample Selection

Whether or not a child was included in the at-risk (AR) subsample depended upon the participant's pretest score on the Letter Naming task (see Participant Characteristics -- Method section), since the literature has repeatedly shown that the best single predictor of future reading
performance is a child's preschool knowledge of letter names (Adams, 1990; Chall, 1998; Scanlon & Vellutino, 1996; Scarborough, 1998).

All of the children who scored 12 or below (i.e., in the first quartile) on the Phase 1 (pretest) Letter Name task comprised the AR subsample. This resulted in a total of 73 participants, 29 out of 120 control and 44 out of 145 children from the Jolly Phonics group. A Chi-Square test ($\chi^2 [1, N = 265] = 0.23, p < .63$) revealed that there was no significant difference between the experimental and control groups in the proportion of AR children; however, there was a difference in the proportion of children in each AR group for whom English was a second language ($\chi^2 [1, N = 73] = 6.10, p < .05$). There were 18 ESL children (out of 44) in Jolly Phonics AR group, and 4 (out of 29) in the control AR group. To address this issue of a differential split of ESL and nonESL children in the experimental and control at-risk groups, a second set of analyses (parallel to the one described in this section) was performed (see Appendix F). All ESL children in both at-risk groups were removed from this second set of analyses, thus eliminating a potential confound from the disproportionate number of ESL children in the Jolly Phonics at-risk group. Even though this resulted in a smaller $N$ (and therefore a more conservative test), results of this set of analyses were essentially the same as the findings reported here.

Variable Screening

To determine if these children were of similar ability level on all Phase 1 pretest measures, performance of the AR children on the remainder of the pretest measures (i.e., minus the Letter-Name Task used for AR designation) was compared by group (i.e., Jolly Phonics vs. control) Prior to this comparison, these data were screened to meet the assumptions for the MANOVA and univariate ANOVA's. Only the Burns and Roe Word Recognition task did not meet the assumptions and could not be included in the MANOVA analyses because all but one of the at-risk children scored 0 on this measure.
Phase 1 (Pretest) Comparisons of Jolly Phonics and Control At-Risk Children

The MANOVA performed on the 4 pretest measures indicated there was no significant difference between the two groups, \(F(4, 68) = 1.64, p = .174\). The only univariate ANOVA which showed a pretest difference was in favor of the control AR children, indicating that they knew a few more letter names, \(F(1, 71) = 4.8, p < .05\), (Jolly Phonics AR mean = 4.59, SD = 4.20, control AR mean = 6.79, SD = 4.20). However, since the MANOVA was not significant, the ANOVA effect cannot be considered significant (particularly at only \(p < .05\)). Table 27 presents these comparisons, as well as the mean and standard deviation for the Burns and Roe Word Recognition test which was not included in the analysis.

Table 27.

Phase 1 Comparisons (Pretest Measures)
Jolly Phonics AR vs. Control AR Children

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics AR</th>
<th>Control AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>((n = 44))</td>
<td>((n = 29))</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>TAAS</td>
<td>1.30</td>
<td>1.76</td>
</tr>
<tr>
<td>LtNm</td>
<td>4.59</td>
<td>6.79</td>
</tr>
<tr>
<td>L/SRecog</td>
<td>1.11</td>
<td>1.41</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>1.00</td>
<td>1.31</td>
</tr>
<tr>
<td>B&amp;R*</td>
<td>0.00</td>
<td>0.04</td>
</tr>
</tbody>
</table>

MANOVA, \(F(4, 68) = 1.64, p = .174\).

**Key:** TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, L/SRecog = Letter-Sound Recognition Task, L/SRecall = Letter-Sound Recall Task, B&R = Burns and Roe Word Recognition

**Note:** B&R was not included in analysis due to lack of sufficient variance in both AR groups.
To see whether the AR designation (based on letter-name scores) for this subsample of children was consistent with their performance across all pretest measures, comparisons were made with the performance of non-at-risk (NonAR) children on the remainder of these tests.

Non-at Risk (NonAR) Subsample Selection

The NonAR group was comprised of all the children who scored 14 or above on the Letter Naming pretest (recall the AR children scored 12 and below). This eliminated 8 participants who scored 13 on this task, leaving a gap between the AR and NonAR children so that the "cut-off" point was less arbitrary. Furthermore, when the original AR/NonAR grouping had been determined by scores of 12 and below for the AR group, and 13 and above for the NonAR group, a discriminant analysis (using the Letter Name task as a first predictor) considered these 8 children (who scored 13) to be misclassified, and so it was decided to drop them altogether.

Variable Screening

The data were screened to meet the assumptions for the pretest MANOVA and ANOVA for the AR and NonAR groupings of children. Transformations for severe negative skewness were undertaken on three of the remaining four pretest variables (Tabachnick & Fidell, 1989). The fourth variable (i.e., Burns and Roe Word Recognition) could not be included in the MANOVA analyses because all but one of the at-risk children scored 0 on this measure, thus performance could not be compared. MANOVA with univariate ANOVA's was performed on both transformed and nontransformed variables. Results showed that there were no differences between the analyses using transformed and nontransformed variables with regard to multivariate or univariate significance levels. Thus, analysis using nontransformed variables will be presented, since interpretation of these results is more meaningful.

Phase 1 Comparisons of At-Risk Children and Non-At-Risk Children

The MANOVA comparing AR and NonAR performance on the remainder of the Phase 1 (pretest) measures (i.e., excluding the Letter Name Task and the Burns and Roe Word
Recognition) was highly significant, \( F(3, 253) = 26.74, p < .001 \), as were each of the univariate ANOVA's, thus supporting the validity of the AR grouping method. Table 28 shows the means, standard deviations, \( F \) values, and significant levels of each of the pretest measures for the AR and NonAR groups. In addition the mean and standard deviation for the Burns and Roe Word Recognition test is presented, although no test of significance could be performed (see "variable screening" above).

Table 28.
Phase 1 Comparisons (Pretest Measures): At-Risk vs. Non-At-Risk Children

<table>
<thead>
<tr>
<th></th>
<th>At-Risk ((n = 73))</th>
<th>Non-At-Risk ((n = 184))</th>
<th>(F)</th>
<th>(p&lt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAAS</td>
<td>1.48  1.28</td>
<td>2.87  2.41</td>
<td>21.71</td>
<td>.001</td>
</tr>
<tr>
<td>L/SRecog</td>
<td>1.23  1.88</td>
<td>8.43  6.76</td>
<td>80.22</td>
<td>.001</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>1.12  1.36</td>
<td>7.02  6.80</td>
<td>53.91</td>
<td>.001</td>
</tr>
<tr>
<td>B&amp;R*</td>
<td>0.01  0.12</td>
<td>1.19  2.41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MANOVA, \( F(3, 253) = 26.74, p < .001 \).**
**Key:**
- **TAAS** = Rosner Test of Auditory Analysis Skills,
- **LtNm** = Letter Name Task,
- **L/SRecog** = Letter-Sound Recognition Task,
- **L/SRecall** = Letter-Sound Recall Task,
- **B&R** = Burns and Roe Word Recognition

*Note:* B&R was not included in analysis due to lack of variance in AR group.

Phase 2 (Posttest)

The previous analyses demonstrated, that the Jolly Phonics and control children designated to be at risk performed similarly on the Phase 1 pretest measures conducted midway through the school year, but both AR groups' performance was far below that of the NonAR children. The next set of analyses compared the performance of the Jolly Phonics AR children and control AR children on the three sets of Phase 2 (posttest) measures, most of which were obtained in June. For the Jolly Phonics children, the posttest equivalent of the pretest measures (i.e., Rosner Test of Auditory Analysis Skills, Letter Name Task, Letter-Sound Recognition Task, Letter-Sound Recall
Task, Burns and Roe Word Recognition, plus the Writing Alphabet Task) were obtained in April, due to their earlier pretesting date. All other Phase 2 data were collected in June.

**Variable Screening**

When the data were screened to meet the assumptions for posttest MANOVA's and ANOVA's for this subset of children at posttest, various transformations were required on some of the variables in an attempt to normalize the distributions. Set 1 variables (i.e. Standard Score measures) did not require transformation. In the MANOVA for this set, the Burns and Roe Word Recognition test from the Set 2 measures was included, because this variable (which was negatively skewed) was the only Set 2 variable which could be successfully transformed (via logarithmic transformation). The remainder of the Set 2 measures could not be transformed or compared because of lack of variance in the scores for control AR group. These children scored 0 on all but one measure (i.e., WRAT Reading - Full Word Raw Scores where only 2 children scored above 0).

The variables for Subsets B and C of the Set 3 measures (i.e., Alphabetic Coding Measures and Phonemic Analysis Measures, respectively) had to be transformed as well (e.g., square root or logarithmic). Analyses for the Set 1 (plus the Burns and Roe Word Recognition measure), and the Set 3 measures were run first using the transformed variables, and then performed on the same sets using nontransformed variables. Differences in these analyses were negligible, with virtually no change in multivariate or univariate significance levels. Therefore, analyses using nontransformed variables for Set 1 and Set 3 measures will be discussed, since interpretation of these results is more meaningful. No formal analyses were performed on the Set 2 (Full Word Raw Scores) measures.

**Phase 2 Comparisons of Jolly Phonics At-Risk and Control At-Risk Children**

As in the larger analyses for the complete sample, omnibus MANOVA's were performed on the outcome measures comparing the two AR groups. However, for the reasons described in
the "variable screening" section, only two MANOVA's were performed (on Set 1 and Set 3 measures) for the AR comparisons.

Set 1 (Standard Score Measures plus Burns and Roe Word Recognition)

A MANOVA showed that the Jolly Phonics and control groups differed significantly on these gross measures of word recognition $F(4, 68) = 3.57, p < .05$. As seen in Table 29, the means for the Jolly Phonics group on all measures were higher than those for the control group; however, only three of the four were significant -- WRAT Reading, WRAT Spelling, and Burns and Roe Word Recognition.

Table 29.

**Phase 2 Comparisons (Outcome Measures)**

Jolly Phonics AR vs. Control AR Children:
Set 1 (Standard Score Measures plus Burns and Roe Word Recognition)

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics AR ($n = 44$)</th>
<th>Control AR ($n = 29$)</th>
<th>$F$</th>
<th>$p&lt;$</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>WaSS</td>
<td>89.93</td>
<td>20.07</td>
<td>85.17</td>
<td>16.34</td>
<td>1.13</td>
</tr>
<tr>
<td>WrRdSS</td>
<td>96.68</td>
<td>12.80</td>
<td>87.59</td>
<td>10.75</td>
<td>9.97</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>93.59</td>
<td>17.16</td>
<td>84.76</td>
<td>12.63</td>
<td>5.65</td>
</tr>
<tr>
<td>B&amp;R</td>
<td>1.41</td>
<td>2.42</td>
<td>0.21</td>
<td>0.41</td>
<td>6.97</td>
</tr>
</tbody>
</table>

MANOVA, $F(4, 68) = 3.57, p<.05$.

Key:


Set 2 (Full Word Raw Score Measures)

Formal analyses could not be completed for Set 2 measures (see "variable screening" above) except for the Burn and Roe Word Recognition task which was discussed along with the Set 1 measures. Informal observation indicates that the Jolly Phonics AR children's scores were higher than those of their control counterparts on every measure. Figure 7 graphically represent
these results and, for purposes of consistency, includes performance on the Burns and Roe Word Recognition as well. On the word portion of WRAT Reading, the Jolly Phonics group averaged 0.76 (SD = 1.20) words, and two of the 29 control AR children were able to achieve a score on this test, for a control group average of .07 (SD = .26). On the WRAT Spelling measure and the two nonword measures, the at-risk children in the control group were unable to decipher (or encode) a single word and so their mean for these three measures was 0. On the word portion of the WRAT Spelling, the Jolly Phonics AR group mean was 0.323 (SD = 0.71). Their mean for Word Attack (raw scores) was 0.945 (SD = 2.22), and for the (informal) Nonword Reading task the mean was 0.73 (SD = 1.92).

Figure 7. Performance of Jolly Phonics at-risk and control at-risk children on the number of full words correct for the Phase 2 outcome measures.

Key:  
- **WrRdWrds** = WRAT Reading Full Words, **B&R** = Burns and Roe Word Recognition, **WrSpWrds** = WRAT Spelling Full Words, **WaRS** = Woodcock Word Attack Raw Score, **NwRd** = Nonword Reading Task
Set 3 (Phonemic Awareness, Basic Skills, Coding, and Phonemic Analysis)

The result of the MANOVA performed on the 11 measures which made up Set 3 was highly significant, $F(11, 61) = 7.101, p < .001$, with all but two measures contributing to this outcome. These were two of the same three Set 3 measure (i.e., Subset A, phonemic awareness and basic literacy skills) which were not significant in the comparison of posttest outcomes for the entire sample.

Set 3 -- Subset A (Phonemic Awareness and Basic Skills)

ANOVA's for two of the 3 measures in this set were not significant. The Jolly Phonics and control AR groups did not differ on the posttest measure of phonemic awareness (TAAS), univariate $F(1, 71) = 0.10 p < .758$. The Jolly Phonics AR group raw score mean was 2.75 (SD = 1.93) compared with the control mean of 2.62 (SD = 1.43). An examination of the change scores for this subgroup of children on the TAAS revealed that there were also no differences between the experimental and control AR groups with regard to improvement on this measure, $F(1, 71) = 2.22, p = .141$. The Jolly Phonics mean increase was 1.46 (SD = 1.73), compared with a mean increase of 0.862 (SD = 1.56) for the controls. On this measure children can receive up to 3 points by segmenting just syllables. Deleting phonemes begins at the fourth point. Since the mean on this measure for both groups was below 4, it would appear that both groups of children had difficulty auditorally analyzing words at the phoneme level. Possible reasons for this are explored in the General Discussion section.

There were no group differences on the Writing Alphabet task where both groups performed comparably $F(1, 71) = 1.66, p < .202$. The Jolly Phonics group averaged 8.36 letters (SD = 5.06), and the control group averaged 6.97 (SD = 3.60). On the Letter Name task, however, the Jolly Phonics AR group knew significantly more letters (mean = 15.32, SD = 8.26) than did the control AR children (mean = 11.66, SD = 5.65), $F(1,71) = 4.35, p < .05$, effect size = .65. As mentioned earlier, higher scores in letter-name knowledge, in this study, is probably a by-product of superior alphabetic coding skills.
Set 3 – Subset B (Alphabetic Coding)

When the Alphabetic Coding measures were examined, striking differences between the two groups were evident (see Figure 8). On the Letter-Sound Recognition task the Jolly Phonics AR group averaged 11.5 (SD = 7.39) letter-sound correspondences compared with only 4.06 (SD = 4.47) for the control, $F(1, 71) = 23.65, p < .001$, effect size = 1.67. An even larger discrepancy was apparent on the Letter-Sound Recall task where the experimental AR children could generate almost 6 times as many letter-sound correspondences as the control children. The Jolly Phonics mean was 12.93 (SD = 8.12) and the control mean was 2.21 (SD = 3.42), $F(1, 71) = 45.12, p < .001$, effect size = 3.14. Thus, the greater amount of time that the Jolly Phonics classes spent on phonics-related activities (as determined by the observation analyses) appeared to also benefit the at-risk subsample of this group.

**Alphabetic Coding**

![Graph showing performance of Jolly Phonics at-risk and control at-risk children on Letter-Sound Recognition (LSRecog) and Letter-Sound Recall (LSRecall) tasks.]

**Key:**  
LSRecog = Letter-Sound Recognition Task, LSRecall = Letter-Sound Recall Task

**Figure 8.** Performance of Jolly Phonics at-risk and control at-risk children on the Phase 2 outcome measures of alphabetic coding.
Set 3 -- Subset C (Phonemic Analysis)

Analyses of the last set of measures examined application of the literacy skills the children had acquired in their respective classrooms. Results demonstrated that the performance gap between the Jolly Phonics and control at-risk children was even larger on the phonemic scoring of the real word and nonword measures. Figure 9 illustrates the highly significant differences in the AR children's ability to analyze words for the purposes of both reading and spelling. On the phonemic scoring of the WRAT Reading test, the average score of the Jolly Phonics at risk group was 7.13 (SD = 8.12), compared to 1.17 (SD = 2.51) for the controls, $F(1, 71) = 14.63$, $p < .001$, effect size = 2.37. For the Burns and Roe Word Recognition (phonemic analysis) the Jolly Phonics group mean was 10.54 (SD = 16.78), and the control mean was 1.66 (SD = 2.44), $F(1, 71) = 7.98$, $p < .01$, effect size = 3.64. On the real word spelling measure (WRAT Spelling -- Phonemic Analysis), the mean for the experimental group was 14.12 (SD = 13.01) and for the control it was 4.28 (SD = 5.01) $F(1, 71) = 15.06$, $p < .001$, effect size = 1.97.

The differences on the nonword tasks indicated that the control at-risk children were virtually unable to decode words when there was not an opportunity to draw upon any sight word lexicon they may have developed. On the phonemic analysis of the Word Attack measure, the Jolly Phonics group's average score was 7.13 (SD = 7.31), whereas the control group was only able to generate an average score of 0.69 (SD = 1.78), $F(1, 71) = 21.56$, $p < .001$, effect size = 3.62. For the Nonword Reading Task, the experimental AR children averaged a phonemic score of 14.80 (SD = 12.40) compared with the control children's average of 1.31 (SD = 1.80), $F(1, 71) = 33.70$, $p < .001$, effect size = 7.49. For the Nonword Spelling Task -- Phonemic Analysis, the mean scores were 15.55 (SD = 14.20) and 1.86 (SD = 2.49) for the Jolly Phonics and control groups respectively, $F(1, 71) = 26.29$, $p < .001$, effect size = 5.50. These children who, midway through their kindergarten year, scored in the bottom quartile on a literacy task as basic as simple letter-name knowledge, were able, by the end of the year, to not only recognize letters and their sounds, but were also able to analyze words (if they had been exposed to the explicit instruction of the Jolly Phonics literacy program).
In sum, comparisons of the Jolly Phonics and control at-risk children point to very important differences between these two groups with regard to end-of-year skills. The first difference is one of degree. That is, the mean skill level of the Jolly Phonics AR children was higher than that of the control AR children as demonstrated by significantly higher performances on almost every outcome measure of reading and spelling, and a far larger number of letter-sound correspondences on the alphabetic coding measures. The mean scores for the Jolly Phonics group on Full Word measures (i.e., Set 2 tasks) barely reached 1, or at the most, 1.5 words correct which, practically speaking, are not very high. However, the fact that the Jolly Phonics at-risk children were able to achieve at least some correct full word responses, especially on the nonword measures (where the mean for the AR control group was 0), may reflect a difference in kind of
skill rather than just one of degree. That is, the Jolly Phonics AR children showed a tentative beginning grasp of the alphabetic principle which appears to be lacking in the performance of the control AR children. When the outcome measures were more finely examined by determining the degree of phonemic analyses these at-risk children were able to engage in, this interpretation is more strongly supported. Findings from the at-risk subsample investigation clearly demonstrated that the Jolly Phonics AR children's superior scores on the alphabetic coding measures translated into superior performance on the tasks representing application of these skills. As mentioned earlier, the correlation analyses (Table 7) of phonemic awareness, literacy skills and outcome reading/spelling tests for the entire sample revealed strong correlations of alphabetic coding scores with all but one posttest measures of reading and spelling, suggesting a definite link between knowing letter-sound mappings, and comprehension of the alphabetic principle.

Correlations Between Phases 1 and Phase 2 Performance:
Phonemic Awareness and Alphabetic Coding Skills for AR Children

The results of the AR comparisons lead to a question of practical interest: Are the at-risk children who scored poorly on the Phase 1 (pretest) measures doomed to remain low at the end of the year (i.e., Phase 2 posttest)? To answer this question, a correlational analysis (using individual scores, not class means) was undertaken with a particular interest in the relation of Phase 1 and Phase 2 alphabetic coding scores, since these measures had been shown to be so strongly linked to reading and spelling performance. It must be noted that significance levels in this analysis are of less import than the actual size of the correlations, since larger sample sizes (as is the case for the experimental at-risk group relative to the control) will exhibit significance for even weak correlations.

Results showed that for the control group, there was a relation between Phase 1 and Phase 2 alphabetic coding measures (see Table 30). For this group, Phase 1 and Phase 2 versions of the Letter-Sound Recall task were moderately correlated ($r = .42, p < .05$), and for the Letter-Sound Recognition task the association was moderate/high ($r = .63, p < .001$). There was also a high
Correlation \((r = .73, p < .001)\) between Phase 1 and Phase 2 knowledge of letter names for the control AR children.

On the other hand, examination of the experimental group's correlations showed that there was only a moderate association \((r = .47, p < .01)\) between Phase 1 and Phase 2 scores on the basic-skill measure (Letter-Name Task). For the complex-skill measures, there was no significant correlation \((r = .29, \text{ns})\) between Phase 1 and Phase 2 versions of the Letter-Sound Recall task, and only a weak relation between the two sets of scores for Letter-Sound Recognition \((r = .36, p < .05\), see Table 31). There was, however, a moderate correlation for the Jolly Phonics group between Phase 1 and Phase 2 TAAS (phonemic awareness) scores \((r = .48, p < .01)\).

Findings from the correlational analysis suggest that for the skills most highly related to reading and spelling measures (i.e., the complex literacy skills of alphabetic coding), it appears that end-of-year levels for some at-risk children do not necessarily depend upon the skill level these children brought with them at kindergarten entry. This observation carries notable implications considering how many children come to school from literacy-impoverished backgrounds (Adams, 1990).

### Table 30

**Correlations Between Phase 1 (Pretest) and Phase 2 (Posttest) Scores on Measures of Phonemic Awareness and Literacy Skills For Control At-Risk Children \((n = 29)\)**

<table>
<thead>
<tr>
<th></th>
<th>TAAS</th>
<th>LtNm</th>
<th>L/SRecog</th>
<th>L/SRecall</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostTAAS</td>
<td>0.354</td>
<td>0.076</td>
<td>0.152</td>
<td>-0.042</td>
</tr>
<tr>
<td>PostLtNm</td>
<td>-0.206</td>
<td><strong>0.731</strong>*</td>
<td>0.240</td>
<td>0.288</td>
</tr>
<tr>
<td>PostL/SRecog</td>
<td>0.124</td>
<td>0.348</td>
<td><strong>0.625</strong>*</td>
<td>0.487**</td>
</tr>
<tr>
<td>PostL/SRecall</td>
<td>-0.037</td>
<td>0.341</td>
<td>0.447*</td>
<td><strong>0.422</strong>*</td>
</tr>
</tbody>
</table>

*Note: *\(p < .05; **p < .01; ***p < .001\).*

**Key:**
- **Pretest** PostTAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, L/SRecog = Letter-Sound Recognition Task, L/SRecall = Letter-Sound Recall Task.
- **Posttest** PostL/SRecog = Letter-Sound Recognition Task, PostL/SRecall = Letter-Sound Recall Task.
Table 31

Correlations Between Phase 1 (Pretest) and Phase 2 (Posttest) Scores on Measures of Phonemic Awareness and Literacy Skills For Jolly Phonics At-Risk Children (n = 44)

<table>
<thead>
<tr>
<th></th>
<th>TAAS</th>
<th>LtNm</th>
<th>L/SRecog</th>
<th>L/SRecall</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAAS</td>
<td>0.475**</td>
<td>0.222</td>
<td>0.464**</td>
<td>0.301*</td>
</tr>
<tr>
<td>PostLtNm</td>
<td>0.225</td>
<td>0.470**</td>
<td>0.366*</td>
<td>0.332*</td>
</tr>
<tr>
<td>PostL/SRecog</td>
<td>0.200</td>
<td>0.504***</td>
<td>0.357*</td>
<td>0.290</td>
</tr>
<tr>
<td>PostL/SRecall</td>
<td>0.105</td>
<td>0.380*</td>
<td>0.231</td>
<td>0.092</td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < .01; ***p < .001.

Key:
Pretest \( \text{TAAS} = \) Rosner Test of Auditory Analysis Skills, \( \text{LtNm} = \) Letter Name Task, \( \text{L/SRecog} = \) Letter-Sound Recognition Task, \( \text{L/SRecall} = \) Letter-Sound Recall Task.
Posttest \( \text{PostTAAS} = \) Rosner Test of Auditory Analysis Skills, \( \text{PostLtNm} = \) Letter Name Task, \( \text{PostL/SRecog} = \) Letter-Sound Recognition Task, \( \text{PostL/SRecall} = \) Letter-Sound Recall Task.

The overall findings of the at-risk group comparisons show that the reading and spelling performance for the AR children in the Jolly Phonics classrooms far surpassed that of the AR controls. Given the substantial gains these children made, it is reasonable to ask, how their performance compares with control children who had average reading readiness skills at the time of pretest.

At-Risk Jolly Phonics Group versus "Average" Control Group

Phase 1 (Pretest)

"Average" Subsample Selection

Performance of the Jolly Phonics AR children was compared to a subsample of the control children who were considered to be "average" at Phase 1 testing -- who had scored in the mid-range (between the 25th and 75th percentile = scores of 14 to 23) on the Letter Name Task. There were 54 control Average children who scored in this mid-range. These groups were first compared on Phase 1 (pretest) measures, then on the outcome measures of Phase 2 testing.
**Variable Screening**

As with all sample and subsample comparisons, the data were first screened to be sure that assumptions for MANOVA and ANOVA were met. For Phase 1 screening, two variables were skewed (TAAS, and Letter-Sound Recognition), thus square root transformations were undertaken. Because all of the Jolly Phonics at-risk children scored 0 on the Phase 1 Burns and Roe Word recognition, comparisons on this measure could not be made. MANOVA's with univariate ANOVA's were performed on both transformed and nontransformed remaining variables. Results showed that there were no differences in multivariate or univariate levels of significance.

For Phase 2 posttest analyses, only one variable was positively skewed (a ceiling effect on the Letter Name task) so that assumptions were not met. Reflection and transformation for severe positive skewness (Tabachnick & Fidell, 1989) was undertaken, and ANOVA's were performed on both the transformed and nontransformed Letter-Name Task variable. In both analyses the effect was significant in favor of the control "average" group; however, the ANOVA using the transformed variable was significant at only the .05 level, whereas for the nontransformed variable, the significance level was $p < .001$. Never-the-less, analyses using nontransformed variables will be presented, since interpretation of these results is more meaningful, and there were no differences with regard to which variables were significant.

**Phase 1 (Pretest) Comparisons:**

**At-Risk Jolly Phonics Group versus Average Control Group**

A MANOVA was performed to determine if the scores of the two groups under investigation were significantly different on all of the Phase 1 (pretest) measures. The Letter Name task is not part of this analysis because scores on this measure were used to determine group designation (i.e., at-risk and average). The Burns and Roe Word Recognition is also not included since, as mentioned above, the Jolly Phonics at-risk group scored a pretest mean of 0, and therefore, there was no variance to compare. As expected, results of the MANOVA (presented in
Table 32) were significant, $F(3, 94) = 17.29, p < .001$, with all three measures contributing to this finding.

**Table 32.**

**Phase 1 Comparisons (Pretest Measures):**
**Jolly Phonics At-Risk Children vs. Control Average Children**

<table>
<thead>
<tr>
<th>Jolly Phonics AR (n = 44)</th>
<th>Control Average (n = 54)</th>
<th>$F$</th>
<th>$p&lt;  $</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>TAAS</td>
<td>1.30</td>
<td>2.88</td>
<td>2.26</td>
</tr>
<tr>
<td>L/SRecog</td>
<td>1.11</td>
<td>7.35</td>
<td>5.70</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>1.00</td>
<td>5.93</td>
<td>5.62</td>
</tr>
</tbody>
</table>

MANOVA, $F(3, 94) = 17.29, p < .001$.

**Phase 2 (Posttest) Comparisons:**
**At-Risk Jolly Phonics Group versus Average Control Group**

A series of ANOVA's comparing these two groups demonstrated that on several outcome measures, the Jolly Phonics at-risk children had caught up to Average controls, and on one measure even surpassed these children. Tables 33 and 34 show the means, standard deviations, $F$ values, and significance levels of all posttest comparisons. A brief summary of these findings is presented here.

Generally speaking, the Jolly Phonics at-risk children did not perform as well as the Average control children only on the basic-skill measures (i.e., Letter Name Task and Writing Letters), or on those standardized tests which mainly tap (at the kindergarten level) just these abilities (i.e., WRAT Reading/Spelling). Differences on these tasks were significant. However, on 14 out of 15 remaining measures, there were no significant differences between the two groups. On the 15th task, the at-risk children in the Jolly Phonics group scored significantly higher than the average children in the control group.
The Jolly Phonics AR group scored comparably on the measure of phonemic awareness (TAAS), and on the alphabetic coding measures. Moreover, the at-risk children in the Jolly Phonics group were able to use their phonemic and alphabetic coding skills to recognize real words and analyze nonwords just as well as the control children who had been designated as having average literacy ability. In one instance of nonword analysis (i.e., Nonword Reading Task -- Phonemic Analysis), the AR group significantly outperformed the control average group, $F(1, 96) = 8.81, p < .01$, effect size = .72, with a mean of 14.80 (SD = 12.40) for the Jolly Phonics AR children compared with 8.39 (SD = 8.94) for the Average controls. In addition, on the nonword standardized measure (Word Attack) the difference (favoring the Jolly Phonics AR children) between the two groups' means approached significance, $F(1, 94) = 3.47, .05 < p < .10$. The AR Jolly Phonics mean was 89.93 (SD = 20.07) on this measure and the control Average group's mean was 82.56 (SD = 18.87).
Table 33.

Phase 2 Comparisons (Outcome Measures):
Jolly Phonics At-Risk Children vs. Control Average Children
Sets 1 and 2 (Standard Score and Full Word Raw Score Measures)

<table>
<thead>
<tr>
<th>Set 1 (Standard Score Measures)</th>
<th>Jolly Phonics AR (n = 44)</th>
<th>Control Average (n = 54)</th>
<th>F</th>
<th>p&lt;</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>WaSS</td>
<td>89.93</td>
<td>20.07</td>
<td>82.56</td>
<td>18.87</td>
<td>3.49*</td>
</tr>
<tr>
<td>WrRdSS</td>
<td>96.68</td>
<td>12.80</td>
<td>103.35</td>
<td>7.46</td>
<td>10.38</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>93.59</td>
<td>17.16</td>
<td>100.48</td>
<td>9.12</td>
<td>6.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 2 (Full Word Raw Score Measures)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>WrRdWrds</td>
<td>0.76</td>
<td>1.20</td>
<td>1.09</td>
<td>1.52</td>
<td>1.36</td>
</tr>
<tr>
<td>B&amp;R</td>
<td>1.41</td>
<td>2.42</td>
<td>2.04</td>
<td>2.96</td>
<td>1.28</td>
</tr>
<tr>
<td>WrSpWrds</td>
<td>0.32</td>
<td>0.71</td>
<td>0.33</td>
<td>0.70</td>
<td>0.01</td>
</tr>
<tr>
<td>WaRS</td>
<td>0.95</td>
<td>2.22</td>
<td>0.56</td>
<td>1.37</td>
<td>1.14</td>
</tr>
<tr>
<td>NwRd</td>
<td>0.73</td>
<td>1.92</td>
<td>0.41</td>
<td>1.02</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Note: *Favoring Jolly Phonics AR group.
Set 2 WrRdWrds = WRAT Reading Full Words, B&R = Burns and Roe Word Recognition, WrSpWrds = WRAT Spelling Full Words, WaRS = Woodcock Word Attack Raw Score, NwRd = Nonword Reading Task,
Table 34

Phase 2 Comparisons (Outcome Measures):
Jolly Phonics At-Risk Children vs. Control Average Children
Set 3 (Phonemic Awareness, Basic Skills, Coding, and Phonemic Analysis)

<table>
<thead>
<tr>
<th>Jolly Phonics AR (n = 44)</th>
<th>Control Average (n = 54)</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>---</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Set 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subset A (Phonemic Awareness &amp; Basic Literacy Skills)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAAS</td>
<td>2.75</td>
<td>1.93</td>
<td>3.44</td>
<td>2.83</td>
</tr>
<tr>
<td>LtNm</td>
<td>15.32</td>
<td>8.26</td>
<td>22.19</td>
<td>2.90</td>
</tr>
<tr>
<td>WrtAlph</td>
<td>8.36</td>
<td>5.06</td>
<td>12.62</td>
<td>6.14</td>
</tr>
<tr>
<td>Subset B (Alphabetic Coding)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/SRecog</td>
<td>11.50</td>
<td>7.39</td>
<td>12.72</td>
<td>6.90</td>
</tr>
<tr>
<td>L/SEncall</td>
<td>12.93</td>
<td>8.12</td>
<td>10.87</td>
<td>7.38</td>
</tr>
<tr>
<td>Subset C (Phonemic Analysis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WtrRdPh</td>
<td>7.13</td>
<td>8.12</td>
<td>9.58</td>
<td>10.21</td>
</tr>
<tr>
<td>B&amp;RPh</td>
<td>10.54</td>
<td>16.78</td>
<td>14.57</td>
<td>19.34</td>
</tr>
<tr>
<td>WtrSpPh</td>
<td>14.12</td>
<td>13.01</td>
<td>15.46</td>
<td>11.34</td>
</tr>
<tr>
<td>WaPh</td>
<td>7.13</td>
<td>7.31</td>
<td>4.98</td>
<td>6.24</td>
</tr>
<tr>
<td>NwRdPh</td>
<td>14.80</td>
<td>12.40</td>
<td>8.39</td>
<td>8.94</td>
</tr>
<tr>
<td>NwSpPh</td>
<td>15.55</td>
<td>14.20</td>
<td>13.83</td>
<td>11.05</td>
</tr>
</tbody>
</table>

Note: * Favoring Jolly Phonics AR group.
Key:
Set 3 Measures:
Subset A: TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, WrtAlph = Write Alphabet Task
Subset B: L/SRecog = Letter-Sound Recognition Task, L/SEncall = Letter-Sound Recall Task
Subset C: WtrRdPh = WRAT Reading -- Phonemic Analysis, B&RPh = Burns and Roe Word Recognition -- Phonemic Analysis, WtrSpPh = WRAT Spelling -- Phonemic Analysis, WaPh = Woodcock Word Attack -- Phonemic Analysis, NwRdPh = Nonword Reading Task -- Phonemic Analysis, NwSpPh = Nonword Spelling Task -- Phonemic Analysis

Findings from the Jolly Phonics AR/control Average analyses showed that children in the Jolly Phonics classrooms who, as a group, were considered midway through the year (Phase 1 pretest) to be at-risk for future reading difficulties, were, at the end of the year, performing at a
level comparable to that of the control children who were in the average range at Phase 1 pretesting. On one measure, the AR Jolly Phonics group even performed significantly better than the control Average group. These results suggest far-reaching implications of early preventative programming aimed at reducing the large numbers of children with later reading difficulties.

Summary

The summary of results is organized as answers to the research questions presented.

Research Questions and Answers

1. *Do participants who received the Jolly Phonics teaching approach perform significantly better than do controls on measures of prereading skill (letter recognition and letter writing), phonological awareness (phoneme deletion), and phoneme/grapheme knowledge (letter-sound recognition, letter-sound recall)? If so, does this superiority in performance translate into significantly higher scores on measures reflecting application of these skills -- word recognition, spelling, and word analysis (nonword reading and spelling)?*

On almost all of the Phase 2 outcome measures which assessed reading, spelling and reading readiness, the children who were in Jolly Phonics classrooms demonstrated significantly superior performance to the children in the control classrooms. There were no significant differences on the two measures representing basic (readiness) skills, and on the measure of phonemic awareness (the Rosner Test of Auditory Analysis Skills -- TAAS). However, in comparison of Phase 1 pretest and Phase 2 posttest scores on the TAAS, the Jolly Phonics group demonstrated significantly greater improvement than did the control group.

2. *Do the Jolly Phonics classrooms spend more time involved in activities which focus on the key components found in the research to be strongly associated with reading acquisition*
(alphabetic coding, phonics, phonemic awareness)? If so, is this accomplished at the expense of time spent on literacy components that are thought to be important elements in kindergarten programming (e.g., language experience, concepts of print, writing experiences)?

In the examination of how the children in each group spent their school day, significant differences were noted. The children in the Jolly Phonics classrooms spent considerably more time involved in certain literacy-based activities, than their control classroom counterparts. For the experimental group, significantly higher proportions of the school day were spent participating in activities which focused on the following main Literacy Components: Phonics, Auditory Phonemic Awareness, Sight Words, and Grammar. In fact, the amount of time that the Jolly Phonics group was involved in phonics-based activities was more than 5 times that of the control group. They also passed significantly less time than controls engaging in non-literacy activities.

Time spent on Phonics activities did not compromise time spent in other areas for the Jolly Phonics group as compared with the controls. These children spent a comparable amount of time to the control children in activities which focused on reading (and being read to), writing, and language development.

3. For all classes, does the time spent involved in the various literacy activities (auditory phonemic awareness training, phonics, reading, writing, etc.) predict scores on the outcome measures? Specifically:
   
a) How much variance in the outcome scores does the phonics component explain?

b) After phonics, what is the next most significant predictor(s)?

c) How much additional variance does classroom time spent on development of auditory phonological awareness explain?
Correlational analyses revealed that for the whole sample of kindergarten classes, several of the Literacy Components were consistently related to end-of-year performance on the various outcome measures. Regression analyses demonstrated that the Phonics Literacy Component was the most potent predictor of scores on most of the posttest measures with proportions of explained variance ranging from 15% to 66%. The Letter Formation component often enhanced prediction, and on occasion, the Sight Word component was involved as well. The Literacy Component of Auditory Phonemic Awareness did not show any association with reading, spelling, or phonemic awareness performance.

Examination of the Phonics subcategories revealed strong correlations of some subcomponents with outcome measures. The subcategory termed "JPActions" (physical actions which accompanied alphabetic coding instruction and related activities only for the Jolly Phonics group) showed the most consistent and, with qualification,15 most often the highest correlations with outcome performance. Correlations of this subcomponent were especially high with the alphabetic coding measures. There were also strong correlations of the Letter-Sound subcategory particularly with the alphabetic coding tasks, and with the reading and spelling measures which were scored by degree of phonemic analysis and representation. These findings were consistent with results of a correlational analysis which examined the relation of literacy skills (i.e., tests measuring basic and complex skills) and phonemic awareness (TAAS) to performance on the outcome measures. In this analysis, the correlations with reading and spelling outcome performance, were much stronger, and more consistent, for the alphabetic coding measures (complex skills), than they were for the phonemic awareness test (TAAS) used in this study.

The last Phonics subcomponent of note was Word Analysis. This was mainly correlated with phonemic analysis scores for the nonwords tasks. Overall, nonword tasks (as

15 Since only the Jolly Phonics group was involved in the correlations of the JPActions subcategory, the N was small which could result in less stability of this factor in analyses.
opposed to real word tasks), seemed to be a slightly better indicator of skill learning and importance of time spent.

4. *Do at-risk children who received the Jolly Phonics teaching approach perform significantly better than their at-risk counterparts in the control classes on measures of prereading skill (letter recognition and letter writing), phonological awareness (phoneme deletion), and phoneme/grapheme knowledge (letter-sound recognition, letter-sound recall)?* If so, does this superiority in performance translate into significantly higher scores on measures reflecting application of these skills -- word recognition, spelling, word analysis (nonword reading and spelling)?

Investigations of the subsample of children who were considered to be at risk at Phase 1 pretesting, resulted in very positive findings for the Jolly phonics program, as the children who were exposed to this teaching approach clearly benefited. Although experimental and control at-risk groups scored comparably on all the Phase 1 pretest measures, their Phase 2 performance on all but three of the outcome measures analyzed (i.e., Writing Alphabet, Word Attack standard scores, and the Rosner Test of Auditory Analysis Skills), significantly differentiated the two groups. The most dramatic differences between these two groups were on the alphabetic coding measures, and on the tests reflecting application of these skills. Midway through the school year (Phase 1 pretest) all of the at-risk children scored poorly on the measures of alphabetic coding, however, only for the controls were substantial correlations seen between Phase 1 pretest and Phase 2 posttest performance on these measures.

A subsequent analysis comparing the Jolly Phonics at-risk group with children who had average reading readiness skills at the time of pretest -- those who scored in the mid-range on the pretest Letter Name Task -- indicated that the status of the Jolly Phonics at-risk children appeared to have changed by the end-of-year testing. That is, on almost all but the
most conservative (or most basic) outcome measures of reading and spelling, these at-risk children performed comparably to the "Average" controls. The Jolly Phonics AR children performed at least as well as the Average controls on the measures assessing alphabetic coding and its application (phonemic analysis scoring of the reading and spelling tasks), and in one instance, significantly surpassed the controls (Nonword Reading Task -- Phonemic Analysis).
GENERAL DISCUSSION
Overview

The effectiveness of the Jolly Phonics program for the full sample of kindergarten children in general will first be discussed, along with explanation and implications of unique findings regarding the importance of various literacy components. Next, results specific to the at-risk subsample of children will be explored. Then the contributions and limitations of this study, future considerations, and conclusions will be presented.

Entire Sample

This current study falls into a somewhat different category than the three groupings put forth in the Literature Review section. In that section, studies were categorized as scientific/laboratory experiments, research examining experimenter-designed "training" programs for classroom implementation or one-to-one tutoring, or as classroom research which compared either existing programs by theoretical approach, or good and poor readers within the same approach. This study does entail classroom research, but it evaluates a new, teacher-developed program based more on empirically and theoretically supported principles than on general "philosophical orientation" (i.e., a belief system). Thus, the current research differs from many previous studies that had examined programs labeled simply either "whole language" or "phonics." Moreover, it utilizes classroom observations (time sampling) to determine not only what was actually occurring in the classrooms, but also how important these activities may be to the desired end result of successfully providing kindergarten children with the foundation for reading and writing, thus preparing them for the curriculum demands of Grade one.

It was through these observations that some interesting findings came to light, around which most of this discussion section revolves. It is clear that the program under investigation was far superior to those used by the control classrooms with regard to successful preparation for reading and spelling. As expected, the children in the Jolly Phonics program outperformed their control counterparts on several different types of reading and spelling measures. However, it is
the examination (through the classroom observation process) of program components and their relative contributions to reading success that is of particular theoretical interest and practical importance.

*Phonics and JPActions*

Consistent with other research findings (e.g., Ball & Blachman, 1991; Foorman et. al., 1998), observation results from this study demonstrated that direct, systematic phonics instruction is a very important teaching component, and in the current study, phonics was strongly predictive of reading and spelling performance. However, there was one aspect (termed JPActions on the observation measure) unique to the experimental program's method of teaching phonics which was correlated with almost all indices of reading and spelling (particularly with measures of alphabetic coding knowledge). This subcomponent of the Phonics observation category is comprised of the association of physical "actions" to each new letter(s)-sound (i.e., grapheme/phoneme) being taught. In the Jolly Phonics program, this introduction of a logical and meaningful action (that is clearly fun for the children) to represent a letter(s) and its sound occurs within the context of an interesting story that uses many words containing the sound of interest. As demonstrated by the example in the Method section, the story topic is suggested by the program (to coincide with a matching picture representing the story, printed words, and actions), but elaboration of the topic is left to the imagination of the teacher. This allows for personal adaptation (e.g., including the names of children in the classroom), making the story all the more meaningful to the children involved; the importance of meaningfulness in literacy instruction is agreed upon by supporters of all educational/theoretical perspectives (Adams, 1990). This manner of teaching takes phonics instruction out of the realm of the "dull drill" of skills in isolation, into an atmosphere of fun and enthusiasm, linking it to words, meaning, and to children's own experiences.

The importance, in this study, of connecting actions to letters and their sounds cannot be overstated. This link seems to have served as a mnemonic device, giving the children an additional means of storing and accessing this alphabetic coding information. Indeed, examiners often witnessed the enthusiastic accompaniment of these actions with verbal responses during testing.
The classroom observations revealed that time spent in activities involving the action/letter/sound connections (i.e., JPActions) was significantly correlated with 15 of the 19 measures of reading, spelling, and phonemic awareness. Significant correlations\(^{16}\) ranged from \(r = .66, p < .05\) to \(r = .93, p < .001\) (the latter correlation was with a measure of alphabetic coding ability). The observations showed that instances of just letter-sound instruction (minus the actions) for all classes also highly correlated with most of the outcome measures (though not quite as many -- 13 of 19). The correlations of this subcategory with the measures of alphabetic coding, were not as strong as JPActions', suggesting an added benefit of JPActions' mnemonic properties.

The strong correlations of the JPActions subcomponent with all the spelling measures is a particularly important finding, because although much research has supported the connection of alphabetic coding instruction to reading outcomes, the effects on spelling performance have been inconsistent. Sometimes there have been no significant results on spelling performance for children who received just code instruction (e.g., Ball & Blachman, 1991). When explicit code instruction has been combined with an emphasis on phonological awareness development, spelling effects have been found more often (e.g., Ball & Blachman, 1991), but they usually have not been as strong and/or durable as reading effects (e.g., Hatcher, et al., 1994); and some researchers have not found any transfer to achievement on spelling tasks (e.g., Foorman et al., 1998). Since spelling places greater demands on recall memory (as well as phonological processes) than does reading,\(^{17}\) these findings suggest that some type of mnemonic aide may be the missing ingredient for facilitating more consistent generalization of explicit phonics instruction to spelling. The added emphasis that the Jolly Phonics actions places on letter-sound correspondences (by way of another mode of sensory input) may allow for more thorough consolidation of this information, as well as

---

\(^{16}\) Since the JPActions' correlations were only for the Jolly Phonics classes, the smaller \(N\) resulted in a higher significance criteria (i.e., critical \(r = .632\)) than for all the other correlational analyses (i.e., critical \(r = .443\)) which included both experimental and control classes.

\(^{17}\) Reading a given word requires seeing letters that are present and recalling the sound(s) that they represent; whereas spelling a given word requires segmenting the word into phonemes, recalling which letters go with those sounds, recalling how those letters look (there are no visual cues present), and recalling how to motorically print them.
provide another route of access, thus better accommodating the greater demands that spelling imposes on recall memory. This notion is further supported by the fact that JPAotions' highest correlation value was with performance on the Letter-Sound Recall Task ($r = .93, p < .001$). That is, the more time children spent participating in activities involving JPAotions, the better they were able to recall the sounds that went with visually presented letters.

The Jolly Phonics group's superior results on spelling tasks also run contrary to beliefs of some whole language advocates who contend that the best way to develop spelling skills is through having ample opportunity to engage in functional and "authentic" writing experiences (e.g., Goodman, 1986). It appears that simply being involved in writing activities is not sufficient to facilitate significant development of spelling skills (at the kindergarten level examined here), and that additional explicit phoneme/grapheme instruction is needed to help provide the tools for written language expression. This does not, however, have to be at the expense of writing experiences. In the current study, there was no difference between groups with regard to time spent in writing activities (but there were considerable differences between the two groups' ability to spell).

There is another possible benefit of the Jolly Phonics actions. The physical explicitness of these activities may have also led to increased phonemic awareness, as evidenced by this subcomponent's significant correlation with the Rosner Test of Auditory Analysis Skills ($r = .68, p < .05$). This may be a result of the fact that over-articulation of the sounds accompanied the actions, thus adding an extra emphasis to the phoneme portion of the phoneme/grapheme association. These actions were often part of a game involving the class "acting" out an entire word the teacher had written on the board; and it was not unusual to see children acting out their own names, sometimes even during free-play time. These activities would require segmenting the word into phonemes in order to determine the appropriate actions; thus it is not surprising that some increase in phonemic awareness skills would likely accrue.

The enhancement of skill acquisition that the Jolly Phonics actions appear to provide may also act at a more basic physical level of children's development. That is, simply being active rather than passive may foster better learning, especially among young children who have a hard
time sitting still. The benefits of active/dynamic learning is agreed upon by several disciplines (e.g., educational and cognitive psychology).

**Letter Formation**

Results showed that activities involving the specific writing of letters also proved to be strongly associated with reading (and spelling). Next to Phonics (and the subcategories of JPActions and Letter-Sound Correspondence therein), the main observation category of Letter Formation was the Literacy Component most highly correlated with the outcome measures of reading and spelling. As well, after the main category of Phonics, it was the next most significant predictor of performance on these measures. It was also the single highest predictor of performance on the Rosner Test of Auditory Analysis Skills (adjusted $R^2 = 43\%$). With the exception of Share et al. (1984) who found that kindergarten letter-copying skill added 8% explained variance, after phonological awareness, to Grade 1 reading achievement, there is little research\(^{18}\) that discusses this activity as being particularly important to the reading process much less to that of phonemic awareness development.

Results of the classroom observations suggest that the prediction value of letter formation activities may contribute, in some ways, as does the prediction value of letter-name knowledge to future reading success. It may not be the process of forming letters itself that is the causal factor underlying the correlation, but it may be the context in which this activity occurs. It would seem that emphasizing, in isolation, how to form a letter would be of little value in learning to read; however, when connecting this instruction to what the letter is, and more importantly, to what sound it makes, the practice of letter formation takes on new properties which make it not only more relevant to the acquisition of reading skills, but also possibly to the enhancement of phonemic awareness. Within this alphabetic coding instruction context, the physical practice of letter formation in and of itself probably serves some reinforcing purpose via consolidation of information through a different mode of processing (i.e., motoric); furthermore, it may also make

---

\(^{18}\) Although Evans and Carr (1988) reported a weak correlation of printing with reading performance.
letter sounds more salient (as noted by the strong correlations with the alphabetic coding outcome measures). Actually "drawing" the sound that is being introduced may help to tune the ear to the phonemic make-up of words (as reflected by the Letter Formation component's predictions of performance on the TAAS), rendering the phoneme more "visible" (Perfetti, 1995), both literally and figuratively. Letter formation promotes active engagement with learning letters and their sounds, rather than passive watching without processing.

Letter formation practice does not, however, appear to translate directly to the facilitation of phonemic analysis when decoding or encoding (as indicated by low, or lack of correlations with nonwords); but it does seem to highlight, or draw attention to, the correct spelling patterns of words. This notion is suggested by the moderate/high correlations of the Letter Formation observation category with several of the real word measures.

**Auditory Phonemic Awareness**

One very important point that has not yet been raised is the fact that significant findings for one notable Literacy Component were conspicuously absent. Contrary to expectations, the Auditory Phonemic Awareness (APA) category of classroom activities was not significantly correlated with any of the outcome measures, not even the measure of phonemic awareness. This category reflects the proportion of the school day children spent in activities which involved the development of phonological awareness at strictly an auditory level. As noted in the Method section, for observation scoring purposes instances of phonological awareness activities could be recorded in one of two mutually exclusive main categories. If an activity emphasizing the phonological make-up of words occurred without involving any verbal or visual (i.e., print) reference to letters or actual spellings, then this was deemed to be an auditory phonological awareness activity and scored accordingly. However, if print or letter names were involved (e.g., such as letter-sound correspondences), the activity would be marked in the Phonics category (within the appropriate subcategory). In this way, the effects of any involvement of connecting sounds to words or letters could be separated from emphasizing strictly the phonological make-up of words. This scoring strategy resulted in many (and perhaps most) phonological activities being
marked under Phonics, because although children's attention was often drawn to the phonemes in words, these instances usually included print or verbally expressed letter referents. This explanation aside however, it was still surprising to see no correlations of the APA category whatsoever.

On closer examination of the APA category, it was discovered that a large proportion of the observations involved strictly implicit rhyming activities, such as chants and songs. Any development in phonological awareness that might take place from these activities (which would most likely be at the level of syllable or rime units of speech) may not show up on the more "phoneme-sensitive" TAAS (see Method section for fuller description of this test). However, if there had been any untapped increase in phonological awareness ability for these children, this increase did not appear to facilitate reading skill acquisition as indicated by the lack of correlations of APA (i.e., time spent in auditory phonological awareness activities) with any of the posttest reading (or spelling) measures. This finding is consistent with other research regarding negligible direct effects of activities involving implicit rhyming. Lundberg (1988) noted that "Rhyming seems to require less conscious and deliberate manipulation of [word] segments" (p. 282), and therefore would seem to be less useful in the facilitation of reading, since it is the conscious awareness of segments, and phonemes in particular (i.e., phonemic awareness) which has the strong relationship with word identification and reading.

Taken together, the findings from the current study suggest that initial explicit spelling-sound instruction works very well when undertaken at the phoneme/grapheme level, even though some authors contend that targeting this level of phonological awareness is not appropriate for early readers (e.g., Goswami 1994). Moreover, when explicit enough, prereaders (even at-risk prereaders) appear to have little difficulty distinguishing the phonemes onto which the letter(s) are mapped. It may be that the explicitness of the alphabetic coding instruction, rather then the end-step of successful word reading, is the element that facilitates phonemic awareness. The strong correlations of the direct alphabetic coding subcategory (i.e., Letter-Sound Correspondences) to outcome measures, along with the even more impressive link of the JPActions subcategory to reading, spelling and phonemic awareness performance appear to support this notion.
Although the results clearly point to the added benefits for kindergarten children of explicit alphabetic coding instruction (over just emphasizing development of phonological awareness), some researchers still suggest that many kindergarten children may not be "ready" to receive such instruction (e.g., Stahl & Miller, 1989). In their comprehensive review of the reading research literature to date (Snow, et al., 1998), the members of the Committee on the Prevention of Reading Difficulties in Young Children put forth a number of recommendations regarding reading instruction in kindergarten (and higher grades). Although most of these suggestions involve emphasizing the same literacy components as found in the Jolly Phonics program (e.g., "to provide practice with the sound structure of words, and the recognition and production of letters", [p. 322]), specific attention to explicit letter-sound and spelling-sound correspondences is not mentioned until they discuss first- and second-grade recommendations. Moreover, there are researchers who believe that whole language methods with "literacy rich" programs focusing on mainly the concepts of print and development of readiness skills would be more appropriate kindergarten fare, especially for at-risk individuals (Sacks & Mergendoller, 1997). Indeed, these are undoubtedly appropriate components for all children in kindergarten; however, they should not be the only components.

As the results of the current study demonstrate, the addition to kindergarten curricula of early explicit instruction in phonics (coupled with strong emphasis on phonemic awareness) may not only be desirable as a general policy for kindergarten programming, but may also be imperative as a preventative measure for future reading failure. However, it must be noted that this does not suggest that implementation of any kind of phonics program at this age level would necessarily be beneficial. Rigid, routinized, and less motivating types of phonics programs may be completely inappropriate for young children. Findings from the current study suggest that the type of phonics programming for children at the kindergarten (or junior kindergarten) level should have certain motivating characteristics (e.g., personalized stories, physical actions, playful delivery) which make it well suited to young children. The program (which contained these features) evaluated in the current study also appeared to be highly beneficial for at-risk children, -- children who had minimum pre-literacy skills at kindergarten entry. These are the children who often go on to have
considerable difficulty achieving an adequate level of reading ability in later grades (Hammill & McNutt, 1980; Scarborough, 1998; Snow etc. al., 1998).

**At-Risk Subsample**

Children who are at risk of literacy failure may be in particular need of explicit alphabetic coding instruction along with phonological awareness development. They are already behind their non-at-risk school mates, and any process which directly enhances their beginning reading skills will go far to reduce the gap. Based on the current study, explicit phonics instruction does not have to wait until these children acquire readiness skills. Instead, as this study suggests, children can become familiar with letters, learn print concepts, and develop phonemic awareness at the same time as they learn about phoneme/grapheme correspondences. In fact, basic readiness skills (such as letter-name familiarity) appear to be a by-product of direct alphabetic coding instruction.

Findings from this study demonstrate that when all of these activities are successfully combined in a viable kindergarten program, remarkable results can ensue.

At-risk children in the Jolly Phonics group spent no less time learning about concepts of print or being involved in language development activities than did the control children, yet they were able to learn many letter-sound associations, and appeared to have little difficulty applying this knowledge to reading and spelling tasks consisting of both real and nonwords. The Jolly Phonics at-risk (AR) children significantly outperformed their control AR counterparts on almost all of the posttest measures of reading and spelling. Moreover, the AR Jolly Phonics group performed comparably on most measures to the children designated as "average" in the control group (i.e., 25th to 75th percentile on pretest letter-name knowledge), suggesting that the at-risk students had "caught-up" in many respects to their average-performing peers in reading and spelling skills. The at-risk Jolly Phonics group even significantly outperformed the average control group on one measure.

It is interesting to note, however, that for the AR subsample, no difference between the two groups (i.e., experimental and control) in phonemic awareness improvement (as measured by the TAAS) was found, even though the AR children in the Jolly Phonics group did so well on
measures of reading and spelling. This may be an artifact of the sophistication of the phonemic awareness measure (TAAS). That is, the TAAS is a measure of more complex phonological awareness, mainly tapping phoneme knowledge and manipulation (Stahl & Murray, 1994; Yopp, 1988) and therefore, as mentioned earlier, might not be sensitive to any increase in more basic phonological ability that may have been acquired by the Jolly Phonics AR group. However, since their performance on reading and spelling measures, when scored according to degree of phonemic analysis and representation, indicated that these children were aware of the phonemes in words, another explanation for the lack of effect on the TAAS is proposed.

**Memory**

It may be that various types of memory play a key role in reading acquisition for some at-risk children. The literature shows that verbal memory for sentences is a strong predictor of future reading performance (Scarborough, 1998), and other research has indicated more generalized memory deficits for children with reading difficulties, such as working memory (e.g., Baddely, 1979, 1986) and short-term memory (e.g., Siegel & Linder, 1984). So poorer memory (or less memory capacity) may be part of the reason the at-risk children score low on initial readiness measures, and memory may also be involved in the apparent lack of differential improvement for the Jolly Phonics at-risk children in phonemic awareness as measured by the TAAS. It would seem that the phoneme manipulation required for success on TAAS test items (e.g., splitting blends, or deleting initial, medial, and final sounds) places considerable demands on working memory. It may be that the Jolly Phonics AR children were aware of these more complex phonemes (as suggested by their test scores mentioned earlier), but they were just not able to demonstrate this awareness without the aid of some sort of visual cues (e.g., as were available in the reading and spelling tests). The following example of what is required for successful performance on a nonword task (where sight-word memory would not be a confound) may help clarify this point. When children attempt to read a nonword, they see the letters and, thus, have the opportunity to use these visual cues to access the sound-mappings they have learned. For spelling tasks, however, children do not have any visual cues initially, but as they generate letters for the
first sounds they hear (and write them down), they no longer have to hold that particular piece of information in memory, and can move onto the next part (sound) of the task. Both types of tasks (i.e., reading and spelling) would require fewer demands on working memory than would the purely auditory TAAS. So, although the TAAS scores did not show significant phonemic awareness improvement for the Jolly Phonics at-risk children, their superior scores on the reading and spelling tasks, especially those involving phonemic decoding and encoding of nonwords, suggests that these children were indeed cognizant of many of the phonemes within words (more so than the control at-risk children), thereby indirectly indicating greater phonemic awareness.

Theorizing aside, whether or not the Jolly Phonics at-risk children did improve significantly in their phonemic awareness, may not matter so much as the fact that in spite of no test-measured differential phonemic awareness improvement, these children were still able to read and spell far better than the control at-risk children, and in most cases, they performed as well as the control children designated as average. As suggested by some research, although initial phonemic awareness is a predictor, and often associated with reading success, it may not be as important for children if they are going to be taught by direct code letter-sound correspondence methods (e.g., Foorman et al., 1998; Perfetti, 1987). Perfetti found that for children who were taught by direct code methods, "entering" (i.e., beginning) phonemic awareness abilities (i.e., synthesis and analysis) were less important to their skill in decoding words than it was for children who were not taught by direct code methods (p. 318). The current study provides converging evidence for this finding. It appears, that even if children are at a disadvantage with regard to their basic literacy and phonological skill level when beginning school, this does not necessarily translate into a disadvantage for successful reading acquisition. It seems that with the appropriate instruction, most children's at-risk status can be overcome, and that successful reading and spelling can be achieved.
Motivation

One vital issue which has not yet been addressed involves the apparent importance of motivation in literacy acquisition. The current study was not designed to study motivation, and that is one of its limitations. However, in spite of the lack of systematic examination of this factor, anecdotal observations made it clear that the approach used by the Jolly Phonics program to provide direct code instruction was highly motivating to both children and their teachers. The children in this study seemed to enjoy very much all aspects of this teaching approach. As described in the Methods section, Jolly Phonics delivers lessons in a fun and playful way, which include many related games and activities, along with letter-practice sheets, and personal "sound books" that the children appear to treasure. Classroom observations of the program in use showed clearly that it did not fit the commonly criticized stereotype of some older phonics programs described as dull and unimaginative. Far from the antiquated view of children being passive recipients of boring phonics lessons, the children in the Jolly Phonics program were enthusiastic, active participants. Prior to the start of the school day, they were peeking in the classroom windows in order to catch a glimpse of the new picture associated with introduction of the day's new sound. Often during free play, children would be seen imitating sounds and actions while playing "teacher" with each other; and on occasion they would protest having to end a phonics session in order to go for recess/outdoor play.

Contributions, Limitations, and Future Directions

Findings are unequivocal with regard to the success of the features found in the Jolly Phonics program in preparing these kindergarten children for the Grade one curriculum. For practical purposes alone, these results are an important contribution to early education since according to the Committee on the Prevention for Reading Difficulties in Young Children (Snow et. al., 1998), there has been no empirically sound evaluations of current kindergarten reading
programs. It must be noted that this comment was in reference to basal reading programs,\footnote{\textit{Jolly Phonics} is not considered a basal reading program. Basal programs are much more extensive in the amount of materials and activities they include, and are designed to provide a complete classroom curriculum for teachers to follow; whereas the Jolly Phonics program is an approach, with many suggested activities and some materials, which is integrated into a classroom's existing program. Most importantly, it provides theoretical explanations for the various instructional suggestions so that the teachers can understand why certain elements are so important, and thus teachers can adapt them accordingly to the unique characteristics of their individual classes (and the ability levels of the children therein).} and \textit{Jolly Phonics} is not a basal program (see footnote). Nevertheless, sound empirical research is needed to evaluate any commercial offerings.

For both practical and theoretical reasons, this research is significant for providing converging support for current knowledge about early reading acquisition, extending this knowledge as to the relative importance of various literacy components, as well as adding a potential candidate to the list of determinants associated with literacy acquisition.

The results of the very successful performance for children in the classes using the \textit{Jolly Phonics} teaching method are important in demonstrating its efficacy. However, it is the information gleaned from the classroom observations which provided the more important knowledge with regard to specific notable literacy components. Unfortunately, in this study, much individual information was lost, since observations were focused on class activity and not individual participants (and therefore class means were used in Literacy Component analyses). Future research might consider tracking the individual children so that specific correlations between their time spent in literacy activities, and their scores on outcome measures can be assessed. As well, using individual participant literacy-time scores, would enormously increase the subject-to-variable ratio for purposes of statistical analysis, thus allowing finer-grained examination of all the specific literacy subcomponents which may be contributing to the effects. Furthermore, this larger \textit{N} would mitigate the problem of the possible instability of the JPActions correlations (which were conducted with only the experimental class means). In addition to recording the amount of time children spend in various literacy activities, future observational research could also rate the quality

\footnote{\textit{Jolly Phonics} is not considered a basal reading program. Basal programs are much more extensive in the amount of materials and activities they include, and are designed to provide a complete classroom curriculum for teachers to follow; whereas the Jolly Phonics program is an approach, with many suggested activities and some materials, which is integrated into a classroom's existing program. Most importantly, it provides theoretical explanations for the various instructional suggestions so that the teachers can understand why certain elements are so important, and thus teachers can adapt them accordingly to the unique characteristics of their individual classes (and the ability levels of the children therein).}
of these activities. A composite quality/quantity score would more closely represent the true caliber of the instructional approaches (and literacy components) observed.

A further limitation of the current study is that the measure used for assessing phonological awareness was not sensitive enough to assess possible subtle effects, and was geared primarily to a more sophisticated level of this ability. As discussed in the Method section, a task designed to assess "simple" phonological awareness (i.e., rhyming) was administered, but was discarded due to unstable measurement. Future studies should use a number of measures to tap the varying levels of phonological awareness thought to exist (Adams, 1990; Stahl and Murray, 1994; Yopp, 1988) so as to more closely track the development of this process and to examine possible links to activities targeting specific literacy components. In addition, the development of phonological awareness in the ESL population of children (i.e., children for whom English is a second language) and its relation to reading and spelling achievement in their second language, need to be more closely examined as little research has been conducted in this area (e.g., Geva, Mack, Sidhu, & Wade-Woolley, 1997; Kwan & Willows, 1998; Morgan & Willows, 1996).

There is another possible limitation in the current study with regard to the issue of English as a second language. The fact that the experimental group had a significantly higher proportion of ESL children than the controls may be a limitation, since results may not be easily generalizable to classrooms with no, or few ESL children. However, studies which have specifically examined differential effects of the Jolly Phonics program on ESL and nonESL populations (or have investigated basic reading and spelling abilities of ESL vs. nonESL children in schools where this program happened to be part of the curriculum), have consistently found no differences in outcome reading and spelling performance between the two groups (e.g., Geva, in press; Kwan & Willows, 1998; Morgan & Willows, 1996). Although the focus of this study was not on second language issues, it appears (from the at-risk analyses where the Jolly Phonics group had a large number of ESL children) that ESL children in the experimental group benefitted from the substantial amount of time they spent involved in direct systematic reading (phonics) instructional activities. This is consistent with ESL studies which have repeatedly pointed to the importance of "instructional intensity" for reading acquisition in both first and second language (Geva, in press).
The findings from the current study suggests that the inclusion of memory measures (particularly working memory) in future research examining programs which utilize a mnemonic device (such as the one used in the Jolly Phonics program), may be useful in determining correlations among memory, the memory-focused literacy component (e.g., JPActions), and outcome scores on reading and spelling measures. The memory measure(s) may also be important as a pretest indicator of children at risk. In fact, it may be useful to include a battery of measures (e.g., Badian, 1994) to more finely discriminate between various types of at-risk and non-at-risk children at pretest, so as to be able to make more confident conclusions about benefits of various program components with this diverse population of prereaders.

A final suggestion for future research would be to conduct longitudinal studies investigating the long-term effects of the Jolly Phonics program, or similar programs which include the various features found, in this study, to be so effective. Longitudinal research could answer questions such as: Is superior performance still evident for the children in the experimental group on measures requiring more sophisticated application of the skills learned in kindergarten? Does their superior alphabetic coding ability transfer to superior reading and more prolific writing in higher primary grades? Do children who have appeared to change their at-risk status, remain within the "average range" at upper grades; and are they more motivated to read and write?

It appears that research in early reading processes and acquisition is beginning to make a momentous shift, moving from old comparisons, to new, more profitable domains. Instead of continuing to compare phonics methods of instruction with those of whole language and meaning emphasis, the importance of meaningful literacy experiences should be a "given", and the focus of future research should now be on the investigation of various types of phonics programs, with and without specific phonemic development emphases. For example, examining a program geared to explicit phonics instruction with a heavy emphasis on phonemic awareness development compared with a similar program minus the phonemic awareness component would further help to elucidate in what ways the development of this element enhances reading skills. As well, if more of these types of studies included classroom observations, more direct claims could be made as to how the
mechanisms of mediation may work; and, if they work differently for differently-skilled individuals.

Conclusions

The findings of this study provide support for the following conclusions: 1) there is no need to wait until kindergarten children have developed "readiness" skills or a certain level of phonological awareness before starting formal reading instruction, and that such instruction can easily occur within the context of learning concepts of print, language and phonological development, and rich literacy experiences; 2) spelling/sound instruction at the level of phonemes is easily acquired by kindergarten children (even those at risk) when it is explicit, systematic and taught within meaningful contexts; 3) the addition of physical actions linked to the letters and sounds appears to serve as an effective mnemonic to help children better consolidate alphabetic coding knowledge; 4) as found in previous studies, alphabetic coding skills taught explicitly, readily transfer to reading tasks; 5) but unique to the Jolly Phonics method of teaching, the physical-action mnemonic may be a key factor in the generalization of alphabetic coding skills to spelling tasks (which has been far less commonly found in the literature); 6) a mnemonic accompanying alphabetic coding instruction (such as the physical-actions in the current study) may facilitate phonemic awareness skills; 7) implicit rhyming activities do not appear to be associated with the enhancement of reading, spelling, or phonemic awareness; 8) all these components combined into a single teaching approach designed for regular kindergarten (and Grade 1) classes are also beneficial for children considered to be at-risk, and with this type of regular-class programming, at-risk children may be able to catch up to the skill level of their not-at-risk peers, thus reducing the need for intensive (and expensive) individual interventions.

It is evident from the present research that children of kindergarten age are not too young to engage in formal reading instruction; in other countries, they have been doing so for years (e.g., New Zealand and Great Britain). The current research shows that children can benefit greatly from an approach which involves systematic (and motivating) phonics instruction for about 15 minutes a
day within the context of a balanced literacy program. With this type of approach, there appears to be no reason why early reading instruction at the kindergarten level should not be implemented in North America. The findings from this study point to a high degree of success for the Jolly Phonics teaching method which is one example of this type of approach. It is a program which combines components that research to date suggests are important for the successful development of reading and spelling (i.e., phonemic awareness and alphabetic coding) with components that have been less explored (e.g., letter formation and actions as mnemonics); one that the children appear to thoroughly enjoy; one that is commercially available; one that requires minimal investment in terms of training, time, and money; and one that is easily implemented by every-day teachers in every-day classrooms composed of a broad range of typical children.
References


APPENDIX A

Nonstandardized Task Protocols
  Phase 1 (Pretest) Protocols
  Additional Phase 2 (Posttest) Informal Protocols

Examiner Instructions for Nonstandardized Tasks
  Phase 1 Instructions
  Phase 2 Instructions

Alternate Scoring Procedures (Phonemic Analysis)
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NAME</td>
<td>SEX</td>
<td>DATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>School</td>
<td>Grade</td>
<td>Tchr</td>
<td>BD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>RECITING ALPHABET</td>
<td>a b c d e f g h i j k l m n o p q r s t u v w x y z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Code 1, 2, or 3 for above: 1 = &quot;say the alphabet&quot;; 2 = &quot;say your ABC's&quot;; 3 = coaxing by starting them off.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>WRITING NAME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>WRITING ALPHABET (time 60&quot;) &quot;write as many letters as you can until I tell you to stop&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Checklist for above: Code 1 if child responds, Code 2 if needs coaxing (&quot;What comes after A?&quot;)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>For the following tasks:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>if right, mark a check or slash: if wrong, record errors and circle incorrect letters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>TELL ME THE NAME OF THIS LETTER (CARD A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Allow 4 seconds, then move on. After 5 consec. errors, reveal whole card: &quot;Are there any more you know?&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>s</td>
<td>a</td>
<td>t</td>
<td>i</td>
<td>l</td>
<td>p</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>c</td>
<td>e</td>
<td>h</td>
<td>r</td>
<td>m</td>
<td>d</td>
<td>k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>g</td>
<td>o</td>
<td>u</td>
<td>f</td>
<td>b</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>z</td>
<td>w</td>
<td>v</td>
<td>y</td>
<td>x</td>
<td>q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>READING WORDS (CARD B) - reveal whole card after 7 errors.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>SCORING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>If Right:</td>
<td>at</td>
<td>in 1 sec.</td>
<td>at</td>
<td>sounds out or silent, but right by 5 sec.</td>
<td>at</td>
<td>self-corrects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>If Wrong:</td>
<td>at</td>
<td>no response at</td>
<td>&quot;don't know at&quot;</td>
<td>what said</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>a</td>
<td>at</td>
<td>back</td>
<td>big</td>
<td>can</td>
<td>do</td>
<td>for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>go</td>
<td>have</td>
<td>help</td>
<td>l</td>
<td>in</td>
<td>jump</td>
<td>of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>one</td>
<td>play</td>
<td>said</td>
<td>see</td>
<td>she</td>
<td>that</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>one</td>
<td>play</td>
<td>said</td>
<td>see</td>
<td>she</td>
<td>that</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Record phonetic approximations of errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Indicate if the 1st word (a) is said as long, short or as &quot;uh&quot; (mark as right)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. WHICH LETTER SAYS "S" (CARD C - alphabet card)
   Have them look at you. Be sure they point to the letter. Repeat sound if necessary.
   Stop after 6 consec. errors. Circle letter if wrong, and write in errors.
   If they give "g" for "j" sound, ask for "another letter that says "j" - only "j" is right
   If they give "c" for "sss", ask for another sound - only "s" is right
   If they say a right letter but not able to point to it, write (VB) - mark wrong
   If they say a right letter (e.g., 'b'), but point to a wrong one (e.g., 'd'), write b(VB)d - mark wrong

2. WHAT SOUND DOES THIS LETTER MAKE? (CARD D1 & D2 - with phonemes)
   Record all answers given
   Allow 5 seconds and then move on.
   STOP after 6 consec. errors - reveal whole card "Any other sounds you know?"
   If child gives long vowels, ask for another sound. Mark only SHORT sound right.
   ** If child gives soft c, ask for another sound. Same with soft g
   * Ask for the second sound for each of these digraphs (book, room & this, thin)

3. s a t i l p n c get "c" & "k"
   e h r m d g o u
   f b j z w v y x q
Phase 2 (Posttest)

Additional Informal Protocols
**NONWORD TASK SCORE SHEET**

Nonword Reading Task #2 (CHILD READS WORDS)

**START WITH PRACTISE WORDS "IP" AND "BAP".**

Allow 10 seconds (if needed) before moving on.

Check mark if right, circle word if wrong and write in phonetic approximation of what child said.

STOP testing after 10 errors (they don't have to be consecutive)

<table>
<thead>
<tr>
<th>lod</th>
<th>sim</th>
<th>jat</th>
<th>kug</th>
<th>tep</th>
<th>yub</th>
<th>wid</th>
<th>ven</th>
</tr>
</thead>
<tbody>
<tr>
<td>gack</td>
<td>meed</td>
<td>hosh</td>
<td>saip</td>
<td>foap</td>
<td>corf</td>
<td>roop</td>
<td>rax</td>
</tr>
<tr>
<td>quim</td>
<td>zoit</td>
<td>shang</td>
<td>chig</td>
<td>thap</td>
<td>parn</td>
<td>vout</td>
<td>sarker</td>
</tr>
</tbody>
</table>
Nonword Spelling Task -- Phonemic Analysis

Administer after the WRAT spelling test (to the same group of children). Hand out the test sheets and ask them to put their names on the top. Then say:

"Now, I am going to ask you to try to spell some more words... but these words are not real words. They are silly words that don't make any sense...like NOP or LERP. I would like you to try to spell them just the way they sound."

"Some may be a little easy, and some might be kind of hard. Don't worry if you can't do them all...just try your best."

"I would like you to put the first word on the top line beside the number 1, the next word below it beside number 2,... and so on". Hold up your sheet and point to the spot, and run your finger down the page.

"Are there any question? Okay, let's begin."

"Now listen carefully, and try to write these as quickly as you can".

1. Allow approximately **15 seconds per word** (less if everyone is finished), but if someone is still obviously working on a word (somewhat successfully) give a little more time to finish.
2. Repeat words if necessary.
3. Give all 22 words.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>sim (sym, cym)</td>
</tr>
<tr>
<td>2.</td>
<td>gad</td>
</tr>
<tr>
<td>3.</td>
<td>fob (fawb)</td>
</tr>
<tr>
<td>4.</td>
<td>nep</td>
</tr>
<tr>
<td>5.</td>
<td>tud</td>
</tr>
<tr>
<td>6.</td>
<td>rin (ryn)</td>
</tr>
<tr>
<td>7.</td>
<td>lem</td>
</tr>
<tr>
<td>8.</td>
<td>huck (huc, huk)</td>
</tr>
<tr>
<td>9.</td>
<td>veet (veat, vete)</td>
</tr>
<tr>
<td>10.</td>
<td>cosh (kosh, cawsh, kawsh)</td>
</tr>
<tr>
<td>11.</td>
<td>paht (pate, payt)</td>
</tr>
<tr>
<td>12.</td>
<td>joap (jope, goap, gope)</td>
</tr>
<tr>
<td>13.</td>
<td>kie (ky, ki, cie, cy, ci)</td>
</tr>
<tr>
<td>14.</td>
<td>zong (awng)</td>
</tr>
<tr>
<td>15.</td>
<td>woop (wupe) (pronounce like soup)</td>
</tr>
<tr>
<td>16.</td>
<td>yift</td>
</tr>
<tr>
<td>17.</td>
<td>chax (chacs, chaks, chacks)</td>
</tr>
<tr>
<td>18.</td>
<td>thorp</td>
</tr>
<tr>
<td>19.</td>
<td>moit (moyt)</td>
</tr>
<tr>
<td>20.</td>
<td>quat (kwat, cwat)</td>
</tr>
<tr>
<td>21.</td>
<td>sout (sowt, cout, cowl) (pronounce like shout)</td>
</tr>
<tr>
<td>22.</td>
<td>narde</td>
</tr>
</tbody>
</table>
Examiner Instructions Phase 1 Session A

1. **Rhyme Task** (dropped from study)
2. **TAAS** - instructions on protocol

Examiner Instructions Phase 1 Session B

1. **Recite Alphabet:** (mark letters skipped or out of place on score sheet)
   “Can you say the alphabet?”
   Code as 1 if they respond to the above statement.
   Code as 2 if you have to say “your ABC’s”
   Code as 3 if you really have to prompt by saying “A...B...C...now what comes next?”

2. **Writing Name:** (Give child lined paper)
   “Can you write your name? Write it on the top of this paper for me please.” (Show the child where).

3. **Writing Alphabet** -or Some Letters: (On same sheet of paper) TIME - 60 sec.
   “I would like you to try to write the alphabet or as many letters as you can until I tell you to stop”
   Code as 1 if child responds to above statement, code as 2 or 3 if you had to prompt like in the first task.

FOR ALL THE FOLLOWING TASKS (except reading words) put a check mark or slash in the box if right. If wrong, circle the typed letter and write in the errors.

4. **Letter Name Task:** (Card A - mixed letter-order card) - Allow 4 seconds before moving on.
   Slide marker down revealing one row at a time. Mark down response and circle the score-sheet letter only if incorrect. If correct, mark a slash, or check mark in the box.
   “I’m going to show you some letters and I would like you to tell me their names”.

   *When child makes 5 consecutive errors, reveal the whole card and ask” “Are there any other letters that you know?”

5. **This is Pretest version of this measure.**

   **Reading Words (Burns & Roe):** (Card B - word list) Allow 5 seconds per word, then move on.
   Slide marker down revealing one word at a time.
   Say:
   “Can you read this word for me?”

   *After 7 errors (they do not have to be consecutive errors), reveal the whole card and ask:
   “Are there any other words on here that you can read?”

   **Indicate with a squiggly vertical line on the right side of the box when you have done this.
   ***For the first word “a”, please indicate whether the kid said the long vowel, the short sound or “uh” (as some people pronounce it like “uh”). Mark all of these answers right for now.

SCORING:
Correct Responses:
1. If the child reads it correctly right away (within approximately 1 second), put a diagonal slash in top left corner of score sheet.
2. If the child is sounding it out (or is silent) but gets it right before 5 seconds is up, (but longer than 1 second), put a diagonal slash in top right corner of score sheet.
3. If the child mispronounces the word, but corrects it before the 5 seconds are up, put the incorrect word (or phonetic approximation of the mispronunciation) in the score sheet box and write a “C”.

Errors:
1. If no response, circle the word. If the child says “I don’t know” write DK in box.
2. If mispronounces or substitutes another word, circle the typed word in the box and write the phonetic approximations.

*5. This is Posttest version of this measure.

Reading Words (Burns & Roe): (Card B1, B2, B3 etc. - word lists) Allow 5 seconds per word, then move on.
Starting with card B1, slide marker down revealing one word at a time.
Say:
“Can you read this word for me?”
Continue with the increasingly difficult word lists on Cards B2, B3, etc. until child makes 7 errors out of the 20 words on the card you are showing. These do not have to be consecutive errors, however, they have to be all on the same card. For example, if the child makes 6 errors out of the 20 words on card B1, you would show the Card B2 word list and count the next error as the 1st one on this list. If the child made 7 errors on this list (Card B2), stop testing.

*IF the child makes 7 errors on the first list, reveal the rest of the card and ask:
“Are there any other words on here that you can read?”
Indicate with a squiggly vertical line on the right side of the box when you have done this.
REVEAL THE REST OF THE CARD ONLY ON THE FIRST LIST. If the child makes it to the 2nd list (Card B2) and then gets 7 errors, simply stop testing.

Second Side of Protocol:

6. Sound Recognition:
Finding Letters When Given the Sounds: (Card C - alphabet card)
Place Card B in front of child and read from your list. Say:
“I’m going to say some sounds, and the letters that make these sounds are going to be somewhere on this card”. (Run your finger from left to right under the alphabet letters). “I want you to point to the letter that makes the sound I say, O.K? Can you show me the letter that says “s”? Which letter makes this sound: “s”?

STOP after 6 consecutive errors.

*Make sure the children point to the letter. If they give a verbal response, say: “Show me the letter.” If they can’t find it but say the right one, put (VB) in the box to indicate verbal, BUT circle the typed letter indicating an error. If they say the right letter name (e.g., b) but point to a wrong one (e.g., d) mark in the box b(VB)d and circle the typed letter (b) as wrong.

**Ask for both letters for the “Kuh” sound. If child gives one right answer (e.g., “c”) ask if they know another letter that says "Kuh"?

***If child gives the answer “g” when we give the sound for “j”, ask them what other letter makes that sound. If they then don’t give the letter “j”, or they give some other letter, mark it as wrong. Put down the letters they did give like this...g/c,...or if right, g/j.

****If they give the answer ‘c’ for the “sss” sound, ask for another letter that says “sss”. Score it right only if they give the letter ‘s’.
7. **Sound Recall:**
   Giving the Sounds When Shown Letters andDigraphs: (Cards D1 and D2 - letters and
digraphs cards)
   Allow 5 seconds before moving on.
   Point to the letter and say:
   "Can you tell me what sound this letter makes? What does this letter
   say?"
   When you come to a digraph say:
   "Sometimes when two letters are stuck together like this (point) they
   make one sound. Do you know what sound these two letters make
   when they are stuck together like this?"

*When child makes 6 consecutive errors, reveal the whole card and ask” “Are there any other
sounds that you know?” Show both cards.

**Be sure to write down phonetic approximations of answers if wrong and circle typed letter
(Phoneme)

***If they give the soft “g” for “g”, ask for another sound it makes. If they don’t give the hard
sound, or a wrong sound, score it wrong and mark phonetically what they do give. (for soft “g”
mark a “j”). If they give no second sound mark it j/-).
Same thing with soft “c”.

****If they give a long vowel sound mark the vowel in the square with a long vowel sign (straight
line) over it...and ask for another sound the letter gives. If they don’t give the short sound, or
another wrong sound, mark it wrong. Please put a slash beside the letter to indicate that you asked
for the second sound -- example, e/ - if no second response; e/e if right second response; and e/u
(or whatever sound they give) if wrong second response. Also, please be sure to put short or long
vowel signs over all vowel answers.

*****Saying “ex” for “x” is considered wrong. Accept only “ks”.
   Saying “en” for “n”, or “em” for “m” is also wrong...BUT if they say ennnnnnn, or
   emmmmm, then it is OK.
   Saying “er” for “r” is wrong, but “errrr” is OK.

******Ask for both sounds for the “oo” (book, moon) and for the “th” (this/thin)
PHASE 2 (POSTTEST)
Additional Nonstandardized Test Instructions

Nonword Reading Tasks Instructions

1. Nonword Recognition Task #1 *(dropped from study)*
2. Nonword Reading Task #2

This is the same type of task as WORD ATTACK

Introduce the task by saying "We are going to do the same kind of thing like we did a little while ago. I want you to try to read the nonsense words that I am going to show you. Remember, these aren't real words,...they're silly words that don't make any sense...words, like snap...or lurp. BUT remember, that letters make sounds, and when you put the sounds together, you can make words...even silly nonsense words"

"I'm going to show you a practice silly word, and I will sound it out for you."

Show the child the first word (ip) on the top of list #1a, keeping all the other words covered. Sound it out "i-p" and then put it together..."ip". Say:

"This word says ip. See, it doesn't make any sense, but that's what I want you to do... sound out these silly nonsense words. We will try the next one together."

Help the child sound out the next practice word (bap), and then say:

"Now I want you to try the next silly words by yourself. Some may be kind of hard, but don't worry,... just try your best."

i) Show the child list 1a (and 1b if the discontinue criterion has not been met) ONE WORD AT A TIME, keeping the remaining words covered.

ii) Allow 10 seconds before moving on to the next word (as the child may be silently or orally sounding out the word).

iii) DISCONTINUE AFTER 10 ERRORS...THEY DO NOT HAVE TO BE CONSECUTIVE.

iv) BE SURE to write a phonetic approximation of what the child says if the word is read incorrectly as we may score the number of phonemes children get right.

Alternate Scoring Instructions
for
Phonemic Analysis Scoring

Reading Scoring

All the reading will be scored the same way.

1. WRAT Reading (score only the 1st 10 words)
2. Burns & Roe (score only the 1st 19 words after the "a") *We are not including the first word "a"* for anything as so many children just read that as the name of a letter.

3. Word Attack (score only the 1st 6 words)
4. Nonword Reading Task (score only the 1st 10 words)

General Reading Issues
- Accept "errrr" as a correct phoneme for "r" as many teachers teach the sound that way.
- Do not accept letter names (e.g., "are" for "r") - except if it's a long vowel within an appropriate sounding-out situation (and not just naming letters).
- There will be **two kinds** of "correct" for our scoring:
  1. "approximately correct" (AC)
  2. "completely correct" (CC)

See "Criteria for "Correct" for Reading Pronunciation" (next page) for explanation.

Specific Scoring Instructions:

a) 1 pt. - only initial phoneme completely correct.

b) 2 pts. - only 2 or more (but not all) phonemes/letters CC (completely correct), AA, or a mix of AC and CC. They have to be in proper sequence.

c) 3 pt. - all phonemes/letter sounded out AC or a mix of AC and CC, blended or not blended.
   Example: the word "spell" would receive 3 points for:
   
   /sl/ /p/ /eel/ /ll/ or "speel".

   d) 4 pts. - all phonemes sounded out CC (for the word in question) but not blended

e) 5 pts. - word read correctly.

Note, a response can never be considered fully correct if it has even 1 AC phoneme. The best it can get is 3 points.

**Mark as "0" if children saying letter names for words.
***If a child says all the phonemes correctly, but adds more, it is only worth 2 points (e.g., "booklet" for "book", or "treat" for "tree").

Criteria for "Correct" for Reading Pronunciation:
There will be **two kinds** of "correct":

1. "approximately correct" (AC)
2. "completely correct" (CC)
When children are sounding out words, they are breaking them into chunks, and so when these chunks (or phonemes) are in isolation, there are sometimes 2 (or more) different correct pronunciations.

1) This is mostly true of the vowels (e.g., short and long "a"), but also relates to digraphs. For example, "sh" is really correctly pronounced as "shhhhh" (CC), but a child who is not yet familiar with the combination of "sh" may break these further into 2 separate letter-sounds, and will say "suh-huh", which we would consider is "approximately correct" (AC), since, on their own, these letters do indeed say "suh-huh". We don't consider this as CC because when these are together, they always say "shhh".

2) We will also accept b/d reversals (in pronunciation) as AC. For example, "dig" for the real word "big" would be: 1 AC phoneme, and 2 RC phonemes, and so that would rate a score of 3 points - (see ratings above a-e).

3) Another time we won't consider a sound as CC is when it is not CC for the word in question. For example, even though "ai" is usually pronounced as a long "a", it would be considered as only AC if pronounced that way when the child was trying to read the word "said", because in that word, the "ai" is pronounced differently.

The last instance of AC is an "appropriate" pronunciation which would be consistent with "early" letter sound knowledge, in the words we are scoring for the purposes of this study, this will only be an issue on the Nonword task, accept - "jack" for the nonword "gack" as AC.

In order to be considered CC or AC, the phonemes must be in proper sequence. For example, in the nonword "ift", you would give 2 points for "foot" because the "f" is before the "t" (and the child has said 2 correct phonemes) but no points for a response of "tuff" (because not in the right order). That is, the AC or CC phonemes "in question" must be in the right order with regard to each other...there may, however, be a few extraneous phonemes in-between (as in the "foot" example)

SUMMARY FOR AC'S
A pronunciation will be considered AC in the following situations:
1. an appropriate pronunciation for that letter (or phoneme) as an isolated letter (phoneme):
   - long or short vowels
   - broken-up digraphs ("suh"- "huh" for "sh")

2. b/d reversals in pronunciation (dig for big)

3. a usually correct pronunciation for a phoneme, but not right for the current word (sade for said)

4. a soft "g" when it should be hard ("jack" for the nonword "gack")

Examples Of Partial and Whole Words That are AC and CC:

<table>
<thead>
<tr>
<th>Real Word</th>
<th>Pronunciations</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>in</td>
<td>ine, or ie-nuh</td>
<td>3 (AC/CC mix)</td>
</tr>
<tr>
<td></td>
<td>on</td>
<td>0</td>
</tr>
<tr>
<td>cat</td>
<td>cate, cuh-ai-tuh</td>
<td>3 (AC/CC mix)</td>
</tr>
<tr>
<td></td>
<td>cuh-a-tuh (short &quot;a&quot;)</td>
<td>4 (CC)</td>
</tr>
<tr>
<td>book</td>
<td>boat</td>
<td>2 (AC/CC mix)</td>
</tr>
</tbody>
</table>
how  hoe, huh-oe  3  (AC/CC mix)  
(because the "ow" also makes an "oe" sound as in the word "show")
size  s-ie-zee  3  (AC/CC mix)
       seeze  2  (AC/CC mix)

**Word Attack**
dee  duh-ee-ee  3  (AC/CC mix)
duh-ee  4  (CC)
*bee  3  (AC/CC mix)

*Remember, we are accepting b/d reversals as AC.

raff  errr  1  (AC/CC mix)
(remember, we are accepting "errr" as well as "ruh" as correct...BUT NOT the letter name "r")

**Burns & Roe**
at  ate  3  (AC/CC mix)
back  bake  3  (AC/CC mix)
       buh-a-cuh-cuh  3  (AC/CC mix)
do  doe  3  (AC/CC mix)
of  off, oaf  3  (AC/CC mix)
       uh-vuh  4  (CC)
       if  0
one  ohnee  3  (AC/CC mix)

she  suh-he  3  (AC/CC mix)
     sh-ee  4  (CC)
that  tuh-ha-tuh  3  (AC/CC mix)
      this  1  (1st phoneme CC)

**Nonword Task**
tep  feep  2  (AC/CC mix)
wid  wib  3  (AC/CC mix)

**Spelling Scoring**

**Nonword Spelling Task** - described in methods section
**WRAT3 Spelling** -- Phonemic Analysis -- Haas' method - see next pages
### SPREADING SCALE SCHEMA

<table>
<thead>
<tr>
<th>SCORE</th>
<th>CRITERIA</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>OPOP</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>E / Y</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>I or N</td>
</tr>
<tr>
<td>3</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>EN</td>
</tr>
<tr>
<td>4</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>NI</td>
</tr>
<tr>
<td>5</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>IEN</td>
</tr>
<tr>
<td>6</td>
<td>Phonetically correct representation.</td>
<td>YN</td>
</tr>
</tbody>
</table>

**Single letter reversals accepted unless reversal represents another alphabetic character (eg. 'f' for 'd').**

---

### SPREADING SCALE SCHEMA

<table>
<thead>
<tr>
<th>SCORE</th>
<th>CRITERIA</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>OPOP</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>ND</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>ENC</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>ADN</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>ANDB</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation.</td>
<td>AHND</td>
</tr>
</tbody>
</table>

**Single letter reversals accepted unless reversal represents another alphabetic character (eg. 'f' for 'd').**

---

**WRAT3 Spelling -- Phonemic Analysis**

---
### SPelling Scale Schema

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>OP</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>MK</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>MAG</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>MKA</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>MAKG</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation</td>
<td>MAC or MAK</td>
</tr>
<tr>
<td>8</td>
<td>Phonetically correct representation including recognition of long vowel.</td>
<td>MAIC or MAYC MACE</td>
</tr>
<tr>
<td>9</td>
<td>Above including correct representation /k/ with a K.</td>
<td>MAIK or MAYK</td>
</tr>
<tr>
<td>10</td>
<td>Conventional spelling.</td>
<td>MAKE</td>
</tr>
</tbody>
</table>

Single letter reversals accepted unless reversal represents another alphabetic character (e.g., "g" for "d").

### SPelling Scale Schema

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>OP</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>HM</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>HEM</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>HMI</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>HIML</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation.</td>
<td>HIMM or HYMN</td>
</tr>
<tr>
<td>10</td>
<td>Conventional spelling.</td>
<td>HIM</td>
</tr>
<tr>
<td>Score</td>
<td>Criteria</td>
<td>Example: enter</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>OP</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>IN</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>MTR</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>NRT</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>MNTR</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation</td>
<td>NTR</td>
</tr>
<tr>
<td>8</td>
<td>Phonetically correct representation including recognition of n- or r- influenced vowel.</td>
<td>ENTR or NTUR</td>
</tr>
<tr>
<td>9</td>
<td>Above including correct representation of r- influenced vowel as UR or IR and correct representation of initial phoneme.</td>
<td>ENTR or ENTUR</td>
</tr>
<tr>
<td>10</td>
<td>Conventional spelling.</td>
<td>ENTER</td>
</tr>
</tbody>
</table>

**Score**

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
<th>Example: cook</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>MM</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>U</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>KU</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>CU</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>CUG</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>COKO</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>COUKEW</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation.</td>
<td>KOUC</td>
</tr>
<tr>
<td>8</td>
<td>Phonetically correct representation of /l/ phoneme with OU or conventional OO.</td>
<td>COUC or KOUK</td>
</tr>
<tr>
<td>9</td>
<td>Above with the representation of initial /l/ with C or final /l/ with K.</td>
<td>KOOK or COUC</td>
</tr>
<tr>
<td>10</td>
<td>Conventional spelling.</td>
<td>COOK</td>
</tr>
</tbody>
</table>
### SPELLING SCALE SCHEMA

<table>
<thead>
<tr>
<th>SCORE</th>
<th>CRITERIA</th>
<th>EXAMPLE: reach</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>QPOP</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>T</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>RU</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>RCH</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>RICH</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>RAECH</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>REACKH</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation</td>
<td>RECH</td>
</tr>
<tr>
<td>8</td>
<td>Phonetically correct representation including recognition of stressed vowel 'E'.</td>
<td>REECH or RECHE or REYCH</td>
</tr>
<tr>
<td>10</td>
<td>Conventional spelling.</td>
<td>REACH</td>
</tr>
</tbody>
</table>

**Additional extraneous letters disregarded.**

---

### SPELLING SCALE SCHEMA

<table>
<thead>
<tr>
<th>SCORE</th>
<th>CRITERIA</th>
<th>EXAMPLE: light</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Random letter string.</td>
<td>MN</td>
</tr>
<tr>
<td>1</td>
<td>One phonetically related letter/phoneme</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>One correct letter/phoneme or two phonetically related letters/phonemes.</td>
<td>L</td>
</tr>
<tr>
<td>3</td>
<td>Multiple phonetically related letters/phonemes or two or more correct letters or a minimum combination of one correct and one related letter/phoneme.</td>
<td>LT</td>
</tr>
<tr>
<td>4</td>
<td>Each list word phoneme is represented by a correct or related letter/phoneme.</td>
<td>LET</td>
</tr>
<tr>
<td>5</td>
<td>Transposition error in otherwise phonetically correct representation.</td>
<td>LINGT</td>
</tr>
<tr>
<td>6</td>
<td>Additional letter(s) in otherwise phonetically correct representation.</td>
<td>LITES</td>
</tr>
<tr>
<td>7</td>
<td>Phonetically correct representation.</td>
<td>LIT or LYT</td>
</tr>
<tr>
<td>8</td>
<td>Phonetically correct representation including recognition of short vowel as word requires.</td>
<td>LIET or LITE</td>
</tr>
<tr>
<td>9</td>
<td>Above with one but not all orthographic features recognized in the conventional way.</td>
<td>LIGHT</td>
</tr>
<tr>
<td>10</td>
<td>Conventional spelling.</td>
<td>LIGHT</td>
</tr>
</tbody>
</table>
Jolly Phonics Program:

Teacher Instructions

Timing of Component Introduction

Key Points
Teacher Instructions

INTRODUCING LETTERS AND SOUNDS

Please stick with the actions and words suggested by the sound sheets and frieze as the whole programme is developed around these (books, etc.) We will change the few English/Canadian differences in pronunciation. If you check all the sound sheets in the envelopes containing the masters, you will see some of the changes already made (e.g. stress Icky-Inky mouse, and for the further phonics for this sound, the 'y' at the end of the word - sunny - has been moved to the 'ea' sound sheet).

The Following is a rough guideline:

1. Show sound sheet (Envelope 1) with picture and tell the story about it (on the back)
   e.g. going for walk with dog...starts barking......I hear a SSSSS and see a big sssssssnake ssssssslither away.

2. Demonstrate the sound with the action and have kids join in.
   "Can you make the sound the snake makes?"
   -the more exaggerated the action, the better.

3. Show the letter on the flash card (envelope 2)
   -have kids do the action while making the sssss sound

4. Show them how to make the letter (on the board).
   -have them trace it in the air
   -write it on the black board
   -they can practise in their work books and on their copies of the sound sheets.

SEQUENCE OF LETTER LEARNING AND TIMING

1. Please stick to the order shown in the book:

   week  1 s, a, t, p, i, n
   2    ck, e, h, r, m, d
   3    g, o, u, l, f, b
   4    ai, j, oa, ie, ee, or
   5    z, w, ng, v, short oo, long oo,
   6    y, x, ch, sh, voiced th, unvoiced th
   7    qu, ou, oi, ue, er, ar

   -In the beginning it is important that the children learn only one way of writing the vowel sounds...e.g. “ai” for long “a” ("ay" and “a-e” come much later...after they really know “ai”).

2. Try to introduce the letters at the same rate as the book says. (We realise this may not be possible)
   -the book suggests 6 a week.
   -2 on Monday, Wednesday, and Friday.
   -it says it will take an hour each day.
   Tuesday and Thursday for review and reinforcing...letter games, other activities.
Accompanying Activities for Reinforcement:

1. have them look through books and find examples of the letter (this way they learn the connection among sounds, letters, words and books).

2. -use flash cards as often as possible (e.g., kids come in -show them the flash card, they all know to do the sound and action) 
   -small groups -kids take turn "being teacher" and showing cards.

3. -kids take home sound sheets..parents help them to practise.

4. Matching Letters, Words, and Pictures (envelope 6) after kids learn first group of letters (6 letters) 
   Easier version - put out pictures, ask kids to sound out the initial sound and put appropriate letter under the picture....or letter first, and match picture.

5. Sound Books (sheets and parent instructions in envelope 3) -as each letter is taught, paste the letter on one page of the child's sound book (a small half/size work book).
   -the child can also write the letter on the page BUT it is suggested that they do not draw a picture in their Sound Books, as the picture cue will distract from the strict sound/letter correspondence.
   - they use this to practise the sounds..at school and at home (use stars for rewards)

6. Games: these cards (Sound Book Sheets and Pictures, Letters and Words Sheets) can be used for games as well.
   e.g. ...pairs (game instructions -envelope 3), lotto (game boards and parent instructions -envelope 6)
   Both of these games can be played as soon as the first letter group has been introduced.

AUDITORY TRAINING

Right from the beginning you can start teaching the children how to listen for sounds in words. Ask them to listen carefully as you say some sounds and see if they can hear a word. If they hear the word tell them to call it out.

   e.g. D-O-G  -do a little of this every day.

BLENDING

1. After kids know the first six letters, demonstrate words on the board.
   e.g. tip pin etc.
   - you could even start earlier start earlier with 'is, as'.

   Be sure to sound out T-I-P as you are writing....then Blend the word....Have children say the sounds and then blend the word by saying the sounds quickly.
IDENTIFYING SOUNDS

1. Have the letters they have learned on display.
   - Choose a word. TIP.
   - Ask the children what sound they can hear at the beginning.
   - Then ask them the next two sounds.
   - Have one child choose each letter (from the flash cards) as it is figured out by the class and put it in the correct order on the blackboard ledge.
   - Then have them read the word.

2. Say the sounds of the word -- P-E-T
   - Have kids say the sounds quickly after you.
   - As they say each sound, they should hold up a finger for each sound.
   - Then say the words normally and see if they can hear the sounds.
   - Do several of these a day... hat, wet, leg, in.
   - You can use small words with letters that haven't been introduced yet... JUST to hear the sounds... not to read or to write yet.

3. Consonant Blends (for hearing, not reading in the early stages)
   - Important to hear the individual sounds in the blends.
   - Say the blend “CR” and the children should say “C” “R” and put up 2 fingers.
   - Do a few examples each day.
   - Also do a few words with blends... bran, flap, from, etc.... there's a list in the book under the 'Auditory Training' section (chapter 2) close to the front of the book (8th or 9th page)

OTHER AUDITORY TRAINING ACTIVITIES:

1. Rhyming words in nursery rhymes

2. Hearing the word after the initial sound is gone
   - Teacher says “pink... now take away the first sound”... class says... “ink”
   - Mice... ice
   - Bus... us

3. Finish the word
   - Teacher says the whole word, and then breaks it down getting the children to finish the word.
   - FOR EXAMPLE:

     split
     teacher says “s”... class says “plit”
     “sp” “lit”
     “spl” “it”
     “spli” “t”

4. Hear the number of syllables
   - Teacher says a word and the kids clap out the number of syllables... VACATION

FOR WRITING TOO

5. Word Families on the Letter Clue Picture (envelope 7)
   - After the appropriate sounds have been introduced, you can use these sheets (after the 3rd or 4th letter groups have been taught)

WRITING:
After the children seem to understand blending to read these words, they can try writing.
FOR EXAMPLE, they could copy the words after the above exercise of 'identifying the sounds or identifying the words from being given the sounds'.

**OTHER REINFORCING ACTIVITIES**
Using the Matching Letters, Words, and Pictures again (envelope 6)
1. This time put the word under picture

2. Put words down, kids sound out and blend words, and put picture under.
3. -put out pictures, have kids write the words... maybe start with just writing the appropriate letter, later moving on to words.
   - they should say the word, listen for the sounds, and write.

4. Missing Sound Sheets (envelope 8). When they have learned the appropriate letters (after the first 3 letter groups have been taught)

5. Word and Picture Joining (string joining - envelope 9 -after 3rd group)
   Sentence Sticking (envelope 10 - after 5th group)

**BLENDING PRACTICE**
After the 1st group of letters have been taught, introduce Word Boxes (envelope 4) Be sure children can hear words they have sounded out (T-I-P). (These are stepping-stones from letters to words to books).

Copy sheets, cut up and put in boxes or bags.

- have the kids, sound out the words, and blend
- kids take turns taking the word boxes home and practising with parents
- can also do it in class, with small groups

Timing:
- Give word box 1 to the children who are the best blenders.
  -When they read the words well, give them word box 2, and give word box 1 to the next group...and so on

  -Lloyd suggests that most kids can be given a new box a day.

In the 12th envelope with the reproducible materials, there are Parent Information Sheets which should be sent home. There is also a sheet with suggestion of how to briefly describe the programme and what they can do to help.
Timing of Component Introduction

These can be found in envelopes sorted in the following order. On the front of each envelope is the list of what is inside, what it is appropriate for, and approximately when they should be introduced into the programme.

<table>
<thead>
<tr>
<th>Envelope #</th>
<th>Materials Inside</th>
<th>When to Introduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Sound Sheets</td>
<td>right away</td>
</tr>
<tr>
<td>2.</td>
<td>Flash Cards</td>
<td>right away (after each letter)</td>
</tr>
<tr>
<td>3.</td>
<td>Sound Book Sheets</td>
<td>right away</td>
</tr>
<tr>
<td></td>
<td>Pairs Game Instructions for Parents</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Word Box Sheets</td>
<td>after 1st letter group</td>
</tr>
<tr>
<td>5.</td>
<td>Homework Writing Sheets</td>
<td>after 3rd letter group</td>
</tr>
<tr>
<td></td>
<td>with Parent Instructions</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Matching Letters, Words &amp; Pictures</td>
<td>after 1st letter group</td>
</tr>
<tr>
<td></td>
<td>Lotto game and Parent Instructions</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Letter Clue Pictures</td>
<td>after 3rd or 4th letter group</td>
</tr>
<tr>
<td>8.</td>
<td>Missing Sound Sheets</td>
<td>after 3rd letter group</td>
</tr>
<tr>
<td>9.</td>
<td>Joining Words &amp; Pictures (String Joining)</td>
<td>after 3rd letter group</td>
</tr>
<tr>
<td>10.</td>
<td>Sentence Sticking (cutting and pasting)</td>
<td>after 5th letter group</td>
</tr>
<tr>
<td>11.</td>
<td>Reading Games and Parent Instructions</td>
<td>after 5th letter group</td>
</tr>
<tr>
<td>12.</td>
<td>Information for Parents (to be sent home to enlist their help)</td>
<td>right away</td>
</tr>
</tbody>
</table>
Jolly Phonics Programme

KEY POINTS TO REMEMBER

For the experimental classes participating in the study, there are a few main activities, or areas, that are essential to this particular programme. We ask that they be included in your implementation.

1. **The order of sounds introduced.** Even though the vowels (and vowel phonemes -e.g., ai) can be pronounced a number of different ways, some research has shown that emphasising just one sound at first (until it becomes somewhat consolidated) helps to reduce confusion about these. The sequencing of the Phonics Handbook programme is based on this research. (We have also seen some anecdotal success of this sequence in schools that have been using the programme for a while)

2. **Try to emphasise the letter sound mostly, rather than the letter name,** any time you see an opportunity to reinforce the letter-sound correspondence. For example, when referring to an object (e.g., ball), or someone’s name, ask “What sound does this start with?” or “What object starts with a ‘buh’ sound?”, or “Whose name has an ‘SSSS’ sound in it?”, rather than “What object starts with “b”, or “Whose name has an “s” it?”

3. **The ‘Sound Books’ are important** for the children to practise associating just the sound with the letter. (No pictures in these books please)

4. **Proper letter formation.** When you introduce a new letter and sound, demonstrate how to properly make the letter with verbal instructions (e.g., “the circle starts here, goes up and touches the top line, comes around and rests its stomach on the bottom line, etc...”) and hand motions. Get the children to draw the letter in the air as you are demonstrating. This will help them get a feel of how the letter is supposed to be formed.

***5. **Auditory training (or Phonemic Awareness).** This is very important.

Look at the auditory training section in the Phonics Handbook for sample activities or the 4-page handout I gave you with the “Reproducible Materials” description sheet on the top. This does not have to take a lot of time. Just a few minutes here and there whenever you see an opportunity. But try to do at least a little every day (even 5 minutes...perhaps in morning circle).

**Example of oral blending:**
“T’m going to say some sounds and see if you can hear a word”
Then sound out some simple words “c-a-t, p-e-n, s-i-t”.
Eventually the words can get harder, -- “t-r-e-e, s-t-e-p, c-l-a-p”.

**A couple of teachers have been making a game of miming the actions quickly to produce a word. The children try to see who can be the first to figure out the word. They verbalise each sound of the demonstrated action and keep repeating the sounds until they ‘hear’ the word and shout it out.

These children love this ‘game’ and are totally engaged, poised for the next challenge. The children also take turns being the ‘pantomimer’. This means, they are not only learning to blend the sounds into words (as above), but they are also analysing a known word (by segmenting) in order to ‘spell’ it through the mimed actions for the class.
APPENDIX C

Classroom Observations:

Instruction Manual with Protocol Sample Pages
Observation Manual
Table of Contents

* Keep referring to sample sheets pp.228 & 229

Time Sampling Procedure.................................................................205
Overview
Data Sheets.......................................................................................206
Overview
Procedure Checklist.........................................................................207

TIME...............................................................................................208
GROUP TYPE....................................................................................208
How to Mark a Scan..........................................................................209

MAIN CATEGORIES (Headings)..........................................................211
#1 SUPERVISION.................................................................211
T led ..............................................................................................211
A/V led .........................................................................................211
ST led ............................................................................................211
Non-supervised................................................................................211

#2 ENGAGEMENT .....................................................................213
a) fully engaged: active.................................................................213
b) fully engaged: passive.............................................................213
c) somewhat eng:active...............................................................213
d) somewhat eng: passive..........................................................213
e) not at all.................................................................................213
Prepare/Finish Off........................................................................213

#3 ORAL VOCAB/LANG ......................................................215
T: date/weather/attend...................................................................215
T: disc. stry/pict/Sh & tell.............................................................215
T: Other .......................................................................................216
high quality...................................................................................216
ch: date/weather/attend...............................................................216
ch: disc stry/pict/Sh & tell...........................................................216
ch: other.......................................................................................216

#4 AUDITORY PHONEMIC AWARENESS .............................217
Overview
blending.........................................................................................217
segmenting words/sent...............................................................218
detect sounds in wds.................................................................218

203
Explicit rhyme practise.................................................................218
rhymes/poems/songs.......................................................................218
tongue twisters................................................................................218

#5 PRINT-RELATED ORAL PHONICS: orally = (o), with print = (wp)
Overview............................................................................................219
letter-name learn (o/wp).................................................................219
oral spelling (o)................................................................................219
practise “spelling” (wp).................................................................219
letter-sound corres (o/wp).............................................................220
JPAcations (o/wp).............................................................................220
word analysis (wp)...........................................................................220
read / find (sight) words..............................................................222
spelling dictation (wp).....................................................................222

#6 WRITING (Letter Formation Practise)............................................223
specific description / prac...............................................................223
copy: letters/wds/sent.....................................................................223
printing own name / letters..........................................................223

#7 WRITING (Real).............................................................................223
guided..............................................................................................223
self-generated...................................................................................223

#8 GRAMMAR / SYNTAX ..................................................................223

#9 READING (Concepts of Print)......................................................224
Overview
T/A/V/ST reads to ch........................................................................224
“Read” (chart/rep/mem)....................................................................224
ch. being “read to” by ch/ST............................................................224
T reads wds/sent on chart..............................................................224

#10 READING (Real)...........................................................................225
oral (real).........................................................................................225
silent (real)......................................................................................225
Using Context/pict...........................................................................225

#11 PHONICS HANDBOOK (Jolly Phonics Materials)......................226
#12 OTHER........................................................................................227
#13 MATERIALS................................................................................227

OBSERVATION DATA SHEETS EXAMPLES...............................228, 229
TIME SAMPLING OBSERVATION PROCEDURE

Overview

Observation Procedure Using the Pupil Activity Scan*

This observation process involves a time-sampling procedure by which a sample of the children’s behaviour over a number of 10-second intervals will be taken. Try to select a group of 4 children (3-5 if necessary) to observe for each 10-second ‘scan’. The behaviour for these children will be recorded along several different dimensions. Following this, a second group of 4 will be selected, observed, and information recorded. On occasion, you may be in a situation when you can only include 1 or 2 children in a scan, but try to stick to 4 as often as possible.

Try to sequence your scans at about one per minute. Select the children to be scanned before you actually begin your scan. Observe for ten seconds (using a second hand on your watch or counting “1 elephant, 2 elephant, 3 elephant...... 10 elephant”). Then write down the info. If it takes longer than a minute to write it down, that’s O.K. as the top priority is accuracy. If you get settled into a situation where recording is very simple and fast because they are doing the same thing for a long period of time, don’t speed up your scans. Just try to stick as closely as you can to about 1 per min.

If a very complex situation arises and it takes quite a bit of time to sort out how you are going to record it, be sure to mark the time down (see TIME section for further instructions) when you resume scanning. By doing this, we will be able to account for the excess time in recording, and not mistake the ‘activity time” as being particularly short because there were only a few scans during this time.

Although you don’t have to keep close track of which children have been observed, try to survey all the children equally by dividing the classroom (or large groups) into segments to be repeatedly focused on in succession. In cases where some kids are working independently while others are in a large group with the teacher, try to sample back and forth between the two general working contexts to get a representative sample of the activities occurring.

**Please try to be impartial** when starting your ‘scan times’. That is, just start your next scan a minute after the last scan (or as close as possible to that time). DO NOT wait until something “important” happens, such as the occurrence of a particularly good piece of teaching. If the scans are selective, then we will not get a representative ‘picture’ of the entire programme. As explained below, the observation strategy used here is ‘time-based’, not ‘event-based’.

Essentially, each observation scan is to act as a photograph. By collecting a series of observational photographs of all the pupils in quick succession, a composite picture can be developed. Each composite picture will reflect the emphasis of the teaching programme as revealed by the children’s behaviour. The greater the number of children engaged in an activity, and the longer the activity persists, the greater the chance that it will be observed and recorded during a 10-second scan. Consequently, the activity will be predominant in the summarized composite picture (Usher & Evans, 1977).

We are only interested in details of the Language Arts Programmes and the Observation sheets reflect that fact by having detailed categories for various aspects of Language Arts, and just a heading titled OTHER, for any other type of activity. However, even when children are involved in “Other” activities, keep scanning, as we are looking at proportion of times being spent on various areas and calibre of student involvement.

* The pupil activity scan is a procedure developed by Usher and Evans (1977). Some of the descriptions used here are taken directly from the authors’ 1977 Summary Report on the Early Childhood Study Evaluation Project for the Board of Education for the Borough of Etobicoke.
You may also need to use the OTHER category to write in something that has occurred which is relevant to Language Arts but that you feel cannot be adequately accounted for in the categories provided. Just be sure to write it in on each page for as long as it is needed so that any scans marked in this section will not be confused with the “regular” use of the OTHER category.

**NOTE:** Refer to the sample observation and cover sheets to see examples of the following instructions.

**COVER SHEET**

**Use pen for all observations.**

Before you start, fill in the: Observer's Name; Date; Time; School; Teacher; Grade; # of Students Present, and Classroom Layout on the cover sheet.

The ‘Time’ portion has 2 sections; ‘Official’ refers to the time that the school morning or afternoon “officially” starts, and ‘Actual’ refers to the time you actually started and ended your observation session (obviously the ‘Actual’ section which is the most important will be filled out after you have finished observing.)

The ‘Classroom Layout’ refers only to whether the class is housed in a single room (Closed), or in a large area shared by other classes (Open). You just need to check one off.

The ‘Notes’ section on the cover sheet is for clarifying anything you think is important that may not be covered by the observation record. Feel free to use this frequently, as any extra information is usually very helpful when looking at these later. If you need to explain anything on the data sheets, just write anywhere you can find a place.

**DATA (Observation) SHEETS:**

**Overview**

*Columns*

The data sheets are set up to record each scan along a number of dimensions. There is a set of 3 columns for each scan, representing the ‘status’ of the kids (i.e., supervised, nonsupervised, and engagement level) in your scan. Depending upon the information seen, you will mark some aspects of your scan under column 1, 2, or 3, or some combination of the 3 (this will be explained in more detail later under ‘SUPERVISION’).

*Rows*

There is a large number of rows which represent: 1) the time (of observation sessions), 2) the class grouping (how children are grouped during an activity or lesson), 3) supervision and engagement status of the kids scanned (and hence, these are cross referenced with those categories in the columns), and 4) programme components associated with Language Arts programmes.

After TIME and GROUP, the individual row descriptions are grouped under:

**13 MAIN CATEGORY HEADINGS**

1. SUPERVISION
2. ENGAGEMENT
3. ORAL VOCAB/LANG
4. AUDITORY PHONEMIC AWARENESS
5. PRINT-RELATED AND ORAL PHONICS
6. WRITING (Letter)
The following is a procedure checklist, -- things to remember for each new page and each new scan. After your read the rest of the instructions and are familiar with the categories, the following steps will make a lot more sense. BE SURE to memorize these steps so that you know which areas must be filled in for every page and every scan.

PROCEDURE CHECKLIST

A) Steps For Each New Page: (See sample data sheet for example)

1. Please number and initial each page as you go along. (You may find it easier to pre-number, and pre-initial each page before you start.)

2. In the TIME row, put the time of the first scan over the first observation column set, each time you start a new sheet. You don’t have to record the time again until the next new page (unless activities or group formation changes occur - described later).

3. In the GROUP row, put the group type (described under GROUP) over the first observation column set, each time you start a new sheet. (or when group formation changes).

B) Steps For Each New Scan: (See sample data sheet)

1. Fill in TIME and GROUP TYPE row only if warranted.

2. Under the appropriate column(s) mark your observation in the SUPERVISION category next to the appropriate row(s).

3. Under the appropriate column(s) (which will be the same as those marked in step #2), mark your observation in the ENGAGEMENT category next to the appropriate row(s).

4. Under the appropriate column(s) (which will be the same as those marked in step #2), mark your observation in any one (or more) of the other appropriate ‘main categories’ (e.g., ORAL VOCAB/LANG, WRITING, OR OTHER), next to the appropriate row(s).

5. If the activity is related to the Phonics Handbook Programme, then also mark your observation in the “PH” row.

6. Fill in MATERIALS row if warranted.
SUMMARY OF PAGE AND SCAN STEPS

A) New Page: On every page, number and initial it, fill in TIME and GROUP TYPE rows.

B) New Scan: In every scan, the SUPERVISION category, the ENGAGEMENT category, and at least ONE OTHER CATEGORY will be marked.

* Note: As an activity can be made up of many different components (e.g. sounding out words, rhyming, reading) at the same time, you may often mark more than one category (representing the various components) in a single scan.

EXPLANATIONS of TIME and GROUP TYPE

* TIME:
Please fill in the time in the TIME row (over the 1st scan) when you start your observation session for the day, and over the first scan on each new page. Fill in the time (over the most recent scan) whenever you stop the observation process (not an individual scan, but the sequence of scanning and writing) due to interruptions or the end of the session. BE SURE to mark the time when you resume the observation process.

Your observations may not begin right away as children will be getting their snow suits off, or may go to an assembly right away (however, we will try to avoid any special event day). Just start observing when everything is settled and write a brief note as to what went on up until the first observation. Mark the time of the first scan.

If kids are settled once ‘announcements’ start, you can start your scans then (and the main category would be marked under OTHER). You may also find that in ‘open’ classrooms, that all the sharing classes are grouped together during announcement time. Just divide up your scan amongst the whole group as it still will give us an idea of how engaged the general group is during this activity. And because the times will be marked for the beginning and end of announcements, we will get a sense of how much of the morning (or late afternoon) is allotted to this.

Sometimes teachers make use of even this time (announcements) to teach something. If you catch something ‘worthwhile’ during a scan at this time, mark your observation under the appropriate category heading, next to the appropriate row(s).

* GROUP TYPE:
This row describes the class configurations during various activities. It also is the place to mark when there is a change (transition) in activities or group formation, or when observations are stopped during some interruption (e.g., recess, or a lesson stopping due to disruptive behaviour).

F = Full Class

‘F’ is used to describe the situation where the full class is grouped and involved in an activity. The children may even be sitting at their desks, but the defining criteria is that they are involved in some activity as one large single unit. This would occur in ‘circle’ or any time the teacher is teaching or guiding a lesson or activity with/to the whole class.

Use ‘F’ even if one or two children are elsewhere in class (e.g. an autistic child with an aide), however, make note of this type of occurrence.
O = Other

This includes any other class configuration ranging from class split in half to all children working individually.

T = Transitions

The ‘T’ is used in the GROUP TYPE row to indicate when the class is in transition. This is described further in the next section.

D = Disruption

Use this code in the GROUP TYPE row when the observations are stopped due to a behavioural disruption. In other cases of stopped observation, just write in the reason (e.g., recess, fire-drill).

*Remember... mark in the time when observations stop and the time when they resume. (You will find you will be mostly using “F” and “O” with an occasional “T”.)

FURTHER NOTES REGARDING ‘TIME’ and ‘GROUP TYPE’

1. After you mark your starting time, mark the type of group formation in the GROUP TYPE row.

2. Any time there is a major group formation change and/or an activity change (e.g., going from ‘circle’ and breaking into small groups for ‘activity time’), mark the time the change started taking place (over the most recent scan). At this point, draw a vertical line in the GROUP TYPE row, to indicate the end of that group formation or activity.

   a) If the entire class is in transition, wait until they get settled into a new activity before resuming observations. When you do resume, BE SURE to mark the time (over the new scan). Also, in the GROUP TYPE row write a “T” (for transition) between the two times, draw a vertical line under the time you resumed, and mark the new group formation type (or the same if it is the same). Skip one column set (i.e., the set of 3 columns which make up one observation scan) and resume marking your observations in the next set of columns under the new TIME and GROUP TYPE.

Example:

<table>
<thead>
<tr>
<th>TIME</th>
<th>1:40</th>
<th>1:42</th>
<th>1:43</th>
<th>1:44</th>
<th>1:48</th>
<th>1:49</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP TYPE</td>
<td>F</td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td>OTHER</td>
</tr>
<tr>
<td>SUPERVISION</td>
<td>S</td>
<td>EN</td>
<td>NS</td>
<td>S</td>
<td>EN</td>
<td>NS</td>
</tr>
</tbody>
</table>

HOW TO MARK A SCAN

You need to become familiar with the categories for a full understanding of where to mark each scan, however, at this point I will explain simply “what” you mark down.
For each scan, you will place a number (representing the number of children in your scan) in the appropriate categories. If you scan a group of 4 children, you would place a 4 in the appropriate categories if all 4 children were doing the same thing and were involved to the same degree; OR, you may have to split up the number (e.g., 1 & 3) in each of the appropriate categories to best describe differing behaviours. This will become more clear with category explanations...BUT I will give a brief example here.

**EXAMPLE:** (See Scan #1 on Sample Sheet page 228)  
If your scan caught 4 children, listening to a teacher reading a story, but 3 were paying full attention and one was looking around, attending off and on, you would mark your scan as follows:

1. put a '4' under the 'S' column (indicating all 4 are involved in a supervised activity), next to the "T led" (teacher led) row category.  

2. put a '3' next to "fully engaged: passive" category under the ENGAGEMENT heading and a '1' next to "somewhat engaged: passive" row category.  

3. under the READING(Concepts of Print)heading, you would put a '4' next to the "T/A/V/ST reads to ch" (teacher/aide/volunteer reads to children) row category.  


**IMPORTANT NOTE:** The occasion will sometimes arise where your scan catches children attentive and waiting....BUT nothing is actually happening during those 10 seconds. For example (Scan #2 p. 228), a child may be in the middle of explaining something for show and tell, but during your scan she is searching for the right words or fiddling with the object of interest (e.g., toy). The 4 children in your scan who are watching her, may be "fully engaged: passive" because they are watching what she is doing, however nothing is happening that is relevant to any of the categories which describe "Language Arts". **IN THIS CASE,** put your '4' beside the "show and tell" category (CH disc stry/pic/Sh&tell) **BUT put a circle around the '4' to indicate a sort of "in limbo" anticipation.**

This also often occurs when the children are playing word games independently. We don't want to mark the OTHER category because that misrepresents what is happening, because during this scan the children are passively engaged in a language arts activity...and so we want to indicate this by marking the scan in the appropriate Language Arts category....BUT given the fact that these observations are trying to represent what "components" (e.g., vocab development, letter/sound correspondence etc.) the children are actually receiving during your scans we must have a way of indicating that during that particular scan, they "received" nothing because nothing was happening... so we circle it.

**Note,** that when children are not engaged in an activity, they also "receive" nothing, but this is indicated by the ENGAGEMENT category, and is due to their own lack of attentiveness, NOT due to the state of the activity in which they are (or should be) involved.

**If you feel confused** as to where, or how to mark a scan, try to remember the purpose for conducting these observations. We are trying to see how many of these various literacy programme "components" the children in each classroom are "receiving". We want to know what proportion of the school day is spent on various activities, and how effective they are (that is, how much real teaching or learning is occurring). Accurate time sampling scans should give us this information. For example, a good word game activity could be occurring for 30 minutes, but if your scan catches predominantly "in limbo" periods (due to dice rolling, disorganization, etc.), this...
tells us that these kids are not getting 30 min. of quality learning, and is representative of what is really occurring.

EXPLANATION OF MAIN CATEGORY HEADINGS

The following is a detailed explanation of the column and row categories.

#1. SUPERVISION: This heading is cross-referenced and refers to:

a) Column Headings
b) a MAIN CATEGORY HEADING

SUPERVISION:

a) Column Headings:
The row across the top of the columns is divided into sets of 3 columns per scan, marked S, EN, NS.

"S" means supervised. The number under this column represents the number of students in your scan who are participating in some kind of supervised activity. When marking further categories for these students (to indicate their involvement and the activity they are participating in), you will continue to use the "S" column for that scan (for the supervised students of your scan).

"NS" means non-supervised. The number under this column represents the number of students in your scan who are working without supervision. They could be working alone, in pairs, or in groups, but there is no direct supervision by a teacher, aide volunteer, or student tutor (described below).

"EN" means engagement and refers to the amount of student involvement in the activity (and when used, is marked with an "a, b, c, d, or e"). This column will not be used very often as the MAIN CATEGORY HEADING marked ENGAGEMENT is used to describe pupil involvement in every scan. However, there will be occasions when we will need clarification further down the data sheet, and this will be the time to use this column. This will be described in further detail when we get to the ENGAGEMENT heading.

SUPERVISION:

b) MAIN CATEGORY HEADING:

As a main category heading this is divided into 4 categories:

1. **“T led”** - Teacher led or teacher supervised
2. **“A/V led”** - Aide or Volunteer led or supervised
3. **“Student Tutor (ST) led”** - a student tutor (ST) is an older child from a higher grade who may be helping the kids. This often occurs as a reading activity (Reading Buddies) where the ST may read to the class or, more likely, to a child paired with her or him.
4. **“Non-supervised”** - children are working without supervision either alone, in pairs, or in groups,
EXAMPLE: (Scan #3 p. 228)
In your scan of 4 children, 3 of the children are in a group activity with the teacher, and one is working at her desk. You would mark the supervision part of the scan as follows: All kids fully engaged, 3 are passive, 1 is active.

1. Put a 3 under the "S" column next to the "T led" row category.

2. Put a 1 under the "NS" column next to the "Non-supervised" row category.* If you happen to stick your 1 beside any other of the row categories (e.g., T led) but it is still in the "NS" column, don't worry, we will still consider it non-supervised...it is the column heading which designates whether the child is supervised or not, the row category just pertains to the type of supervision for those who are supervised.

3. Put a 3 under “S” next to “fully engaged:passive” - and 1 under “NS” next to “fully engaged:active”.

This is a partial example, I have not mentioned what the activity is, so the sample sheet (p.228) shows a 'partial scan'.

NOTE: It may happen that during a scan, the teacher comes up to a non-supervised group and starts supervising. This is a judgment call on your part. Mark it as what seemed to be the status for the biggest part of the scan (e.g., if the teacher just walked in at the last couple of seconds, it would still be marked "NS").

NOTE: As mentioned above, T led means Teacher led or teacher supervised. This usually means that the teacher is leading the lesson or discussion etc. However, it can also mean that the teacher (or aide or volunteer) is supervising a situation where a child is leading the discussion, such as in 'show and tell'. There are a couple of areas on the obs. sheet where we want to make this distinction of a 'dual' situation. One is in the category of READING (described later), and one falls under the ORAL VOCAB/LANG category described below.

EXAMPLE: (Scan #4 p. 228)
A child is leading a full group 'show and tell' showing a puppet with the teacher supervising what is going on. Your scan would be marked as follows:

1. - all your observations would be under the "S" column (indicating supervised)

2. - in the Supervision category you would mark your observation next to T led

3. - fill in the appropriate spaces in the 'Engagement' category (e.g., 2 fully passive, 2 somewhat passive)

4. - in the Oral Vocab/Lang category, mark your obs. next to CH:disc.shtrypic/sh&tell

5. - check off the high quality description only if high quality words occur (see the description of what is considered high quality in the ORAL VOCAB/LANG category description. (Let say child describes puppet as marionette).

6. - Put object (and high quality word) in Materials category
**Should the next scan catch the teacher (or aide or volunteer) cutting in to discuss or elaborate upon something, even though the child is still in front of the group, you would fill in step # 4 next to T: disc. stry/pict/Sh&tell.

Remember, try to most accurately describe each individual scan...what best represents what is happening just at that 10 second interval.

There may be other times when children appear to lead a group because they stand in front and do something; - such as spell a word, or ask someone else to spell a word, however we will not distinguish this as different from teacher-led (or A/V led) as this is usually very closely supervised and greatly under the teacher's (or A/V's) guidance.

---

#2 ENGAGEMENT

This category heading describes how involved in the activity the children of each scan are.

"fully engaged: active" - the child is actively involved, answering questions, putting up hand in response to teachers question, reading out loud, engaged (and on task) in seatwork etc. If reading silently in an independent situation, then she/he is actively involved, but if the teacher (or fellow student) is reading to the class and the child is attending to this, or following along silently reading her/his own books (or a big one held up by the teacher), the child is considered to be passively involved.

"fully engaged: passive" - the child is attending to the teacher, or actions of others involved in the learning activity.

"somewhat engaged: active" - the child shows sporadic active involvement (e.g., puts up hand to answer question, but is looking out the window; is on and off task in a seatwork situation during the scan).

"somewhat engaged: passive" - the same as above, only the sporadic involvement is passive (e.g., attending to teacher for a few seconds, then talking to neighbour or looking out window, then looking back at teacher).

"not at all" - no involvement in learning activity (e.g., talking to others, wandering around the room, etc.)

"Prepare/Finish Off" - This category will take the place of any of the 5 above if you catch a student(s) in the process of preparing to start an activity (e.g., getting papers, pencils, crayons etc.) or in the processing of finishing off an activity (e.g., putting things away, tidying up). For those children you have marked in this category, no other category need be marked (except for SUPERVISION). (So if in your scan of 4 children, 1 was preparing or finishing off, only 3 children would be put in any further categories appropriate to their activity.)

This is similar to TRANSITION, but it is used for individual children...that is, in cases where 1 or 2 children in your scan of 4 are doing this while the rest of the class (or most of the rest of the class) is still working. **If the whole class is preparing or finishing off, this would be noted as "T" (transition) in the GROUP TYPE row, and the TIMES would be marked accordingly.**

NOTE: On the observation sheet, the first 5 ENGAGEMENT categories are labelled "a, b, c, d, and e". This is so you can use these letters to put into the "EN" column when needed. This will
sometimes need to be done when the 4 children in your scan are differentially engaged in different activities. We will need to know how involved the kids are in their particular activity, and so you would mark the appropriate letter (a-e) which represents the quality of engagement in the "EN" column next to the kids in the appropriate category row. The following example should help make this clear:

EXAMPLE: (Scan #5 p. 228)  
4 children are at a table. One is reading a library book and 3 are writing and/or drawing an accompanying picture in their journals. At the time of your scan, the teacher is helping the child who is reading the book sound out words and this child is fully engaged both reading and sounding out the words, interacting with the teacher. 2 of the children are colouring their pictures in their journal. One of these children is fully on task, the other is somewhat on task. The 4th child is fully engaged in writing in her/his journal.

Your scan would be marked as follows:

In the SUPERVISION category:

1. Put a 1 next to "T led" under the "S" column.
2. Put a 3 next to "Non-supervised" under the "NS" column.

In the ENGAGEMENT category:

1. Put a 1 next to "fully engaged: active" under the "S" column (representing the child with the teacher).
2. Put a 2 next to "fully engaged: active" under the "NS" column (representing the 1 child who is fully on task colouring the picture, and representing the 1 child who is fully on task writing in her journal).
3. Put a 1 next to "somewhat engaged: active" under the "NS" column (representing the 1 child who is somewhat on task colouring the picture).

In the PRINT-RELATED AND ORAL PHONICS category:

1. Put a 1 next to "word analysis (wp)" under the "S" column (representing the child with the teacher).

As this category does not pertain to the other children's activities, nothing else is marked here.

In the WRITING (Real) category:

1. Put a 1 next to "self-generated" under the "NS" column (representing the child who is fully on task writing in her journal).
2. You would also put an "a" under the "EN" column next to this child (1), to represent that the child who is writing in her journal is one of the children who have been marked as "fully engaged: active" under the "NS" column.
In the READING (Real) category:

1. Put a 1 next to "oral (Real)" under the "S" column (representing again, the child with the teacher).

In the OTHER category:

1. Put a 2 next to the blank space in the OTHER category, under the "NS" column (representing the 2 children who are colouring).
2. You would also put an "a" and a "c" under the "EN" column next to the "2", to represent the two different involvement levels of these 2 children.

(We cannot distinguish the differential involvement of the scanned children according to activity in these cases, just by looking at the ENGAGEMENT category markings alone. So we have to use the "a,b,c,d, and e" codes under the "EN" column).

#3. ORAL VOCAB/LANG

This includes all oral language development which usually occurs in "show and tell" or story discussion, etc. Be sure to mark the category, as well as marking the "high quality" if warranted. If you notice "high quality" language occurring during scans of other situations (e.g., during some phonics sounding out of a sophisticated word) mark "high quality" (see description below) even though the rest of your observations are in the PHONICS category.

Because of the above mentioned case where a teacher (or aide or volunteer) is supervising a situation but a child may be leading the discussion, the following categories are prefaced with "T" or "CH" to account for the two possible situations.

"T" -- means the teacher (or aide or volunteer) is actually leading the discussion (as well as supervising).
"CH" -- means the child is leading the discussion (but the teacher would still be supervising).

T: date/weather/attend  - This category includes "tomorrow is..., yesterday was... etc." Teacher (aide or volunteer) leads discussions of the date or weather or attendance. This category describes very routine stuff. (e.g., "Today is Tuesday and it is windy.") If she elaborates into a "real" discussion about hurricanes etc., then mark it under "T: other". If not routine, don't mark here, mark only in appropriate category.

Remember...
even if a child is in front and has been the leader, if your scan catches the teacher cutting in and elaborating at that moment, then mark the category that is prefaced by "T".

T: disc, stry/pict/Sh&tell  - Teacher (aide or volunteer) leads discussions on a story that she has just been reading, or about story pictures, or 'show and tell'. For example, the teacher may have just been reading a story, but by the time you started your scan she has stopped
reading and is talking to the class about it - asking questions etc. So you mark just this category. But, if your scan caught both reading and talking, mark this category, and also mark your scan in “READING (Concepts of Print)” next to the “T/A/V reads to ch” category.

T: other
- This includes T, aide, or volunteer.
- Note** This category is used whenever teachers are explaining things or giving instructions.
- This category is for any other topics (such as making up a story) that the teacher may be discussing with the children, or a video the children are watching. As mentioned above, if routine discussions of date, weather, or attendance gets elaborated into wide-ranging discussions of weather in general, or why “Billy” is away, or “why Feb 14 is Valentine’s day and how it got started,” then these will be marked under this (T: other) category.

**high quality**
- If the language (even one good word) in your scan is of high calibre, mark your number (representing the # of kids in your scan exposed to this language) beside this category. This refers mainly to vocabulary and will be marked if you catch the use of a word (or words) that is above the child’s (children’s) usual talking vocabulary. (If you are not sure if the word should be considered “high quality”, mark it anyway, and write the word in the “MATERIALS” section.)

Example: (Scan #6 p.228)
Your scan catches the teacher reading to the class, and stopping to ask the children what might happen next. The last few words that she read were “and the wind blew relentlessly against the brittle cabin door”. You would mark your scan as follows:

Steps:
1. All your obs. will be under the “S” column.
2. SUPERVISION = “T led”
3. ENGAGEMENT = whatever is appropriate according to the 4 kids you were watching. (e.g. 3 fully passive, 1 not at all)
4. ORAL VOCAB/LANG = “T: disc. story/pict/Sh & Tell”
5. = “high quality” would also be marked with a 4.
6. READING (Concepts of Print) = “T/A/V/ST reads to ch”
7. MATERIALS = put library book (to indicate it’s not a simple basal reader) and write the words “relentlessly” and “brittle” (if you have time).

CH: date/weather/attend - This is exactly the same as “T: date/weather/attend”. But it is a child who is leading the routine discussion (but the teacher, aide, or volunteer would still supervise).

CH: disc stry/pic/Sh&tell - same distinction as above.

CH: other - same distinction as above
# 4 AUDITORY PHONEMIC AWARENESS

**Overview**

This main heading refers to phonemic awareness (or auditory training) exercises that are not print related or letter name related. The sounds, not the letters, are emphasized. This means, children’s attention is not being specifically directed to printed letters or printed words (or letter names) when participating in some type of “word or sound analysis” activity.

**For example:** a teacher asking for words beginning with shhh (the sound of “sh”) your scan would go under this heading (next to “detect sounds in wds”).

If she asks for words beginning with “s” “h” (the letter names) you would not mark under this heading. It would go under the PRINT-RELATED and ORAL PHONICS main heading (to be described later) because letter names are emphasized in this case, not the sounds...or if letter names and sounds are emphasized, the scan would also be marked in Phonics.

The AUDITORY PHONEMIC AWARENESS heading is to be used only for oral/aural sounds; not letter names, or printed letters (either alone or associated with sounds).

**#Note:** There is a time when you would use this heading when print may be indirectly involved. Consider the following scenario:

The teacher has just introduced the phoneme “ch”. She has told a story about a choo-choo train and has been stressing the sound of the ch phoneme = (sound ch), not the letter names of the ch phoneme = (“c” “h”). The ch phoneme is printed on the chart beside her, and although she has pointed out that “c” “h” say ch (sound ch), no more print, or writing of words have been involved. She is asking for words that start with the ch sound.

Possible scans:

#1. If your scan just caught her pointing out that “c” “h” (letter names) says ch (sound ch), then you would mark it under PRINT-RELATED and ORAL PHONICS, next to the “letter - sound corresponds (o/wp)” category.

#2. If your scan caught only the asking for words starting with ch (sound ch), then you would mark it under AUDITORY PHONEMIC AWARENESS, next to the “detect sounds in wds”.

#3. If your scan caught both events, mark both categories.

**Remember**, often more than 1 category will be marked to best describe what is happening in a scan, either because 1 scan caught 2 events, or, 1 event has more than just 1 component, and therefore will need 2 or more categories to account for it.

**Category Descriptions**

“Blending” - includes any oral attempt at putting together (blending) the single phonemes into a word.

- includes detecting a word task, -- the teacher may say "I'm going to say some sounds and I want you to tell me what word you hear"...and she sounds out "p-o-t".
Remember - THIS is without print. If the teacher is pointing to letters and asking the children to blend them into words, this would be then marked under the PRINT - RELATED and ORAL PHONICS heading, next to the "word analysis (wp)" category.

"segmenting words/sent" - includes any oral or (mental) breaking up of words (or sentences) into their parts, -- either syllables (va-ca-tion, De-cem-ber, etc.), onset and rimes (br-at, sp-at, r-at, str-ong, wr-ong, l-ong, etc) or individual phonemes (sh-ee-p, f-i-sh, f-i-s-t, t-r-ai-n) or sentences (How many words in the sentence: "I went to ..")
- includes 'finish the word' task.

For example:
Teacher says the whole word, and then breaks it down getting the children to finish the word: split

teacher says "s"...class says "plit"
"sp" "lit"
"spl!" "lt"
"split" "t"

OR - Hearing the word after the initial sound is gone (onset and rimes)
-teacher says "pink...now take away the first sound"....class says.."ink"
  mice...ice
  bus...us

"detect sounds in wds" - includes any oral emphasis on initial, medial, final sounds in words (e.g., "Who has "SSSS" sound in their name? Where is it -- at the beginning, middle or end? Can you give me a word with 'shhhhh' in it?", "What sound is at the end of 'sit'?" etc.)

* If the teacher asked for all the sounds in a word (e.g., all the sounds in 'sick'), this would be marked under "segmenting wds/sent".

"Explicit rhyme practise" - remember - no print (this is auditory)
- this would include explicitly pointing out that words rhyme, or asking what rhymes with ..., etc.
- that words that have the same ending sound rhyme.
This can occur in the context of the teacher reading a story (or poem), but it must be explicitly said (but not show print). For example, your scan must catch the teacher pointing out that "hill and pill rhyme" - when reading a poem etc., but she wouldn't show the words. If she is just reading poems (without explaining anything) it would be marked in the next category. (If she refers to the spelling/print this would fall under PRINT-RELATED PHONICS Headings.) (See later)

"rhymes/poems/songs" - this includes listening to poems, sing songs, reciting memorized poems (rhymes) - not print related. If the kids are following along repeating some rhyme that is printed on a chart (or book, or board) while the teacher is pointing, this would not be marked in this category. It would fall under: READING (Concepts of Print), next to the "Read:chant/rep/mem/" category.

"tongue twisters" - e.g., "peter piper picked a peck of pickled peppers"
# 5 PRINT - RELATED AND ORAL PHONICS:  orally = (o); with print.(wp)

Overview

This heading includes categories depicting:

**Oral with print:** discussions of letters(letter names), or letter/sound correspondences either with or without print. (Even though letter names are not really considered as phonics, I am putting them here, because they fit better than in the AUDITORY PHONEMIC AWARENESS heading which is meant to cover strictly phonemic sounds without print).

- phonics handbook actions - sound - letter correspondences with or without print (see details below)

Just **oral:**

- spelling orally (not print-related) but uses letter names.

Just **with print**

- any kind of print-related word analysis. Many of the things that have been mentioned in the AUDITORY PHONEMIC AWARENESS section will be marked here when Print-related (involving print) e.g., segmenting, blending, rhyming, sound detection etc., or involving the aural names of letters.

**Category Descriptions:**

* Note:  
  (o) = orally only
  (wp) = with print only
  (o/wp) = both orally and/or with print

**"letter-name learn (o/wp)"**

- this category includes all letter name learning (not words)
  (e.g. singing the alphabet, pointing to letters and saying their names)
  - "What letter is this?"
  - "This letter is a."

  * This involves only learning individual letters (not sounds). If spelling is involved, it would be marked either under "Oral Spelling (o)" or "Word Analysis (wp)", or "Spelling Dictation (wp)".

**"Oral Spelling (o)"**

- this involves only oral spelling (no print) (e.g.) - "How do you spell cat?" (and children respond orally), or "What letter does dog start with?"
or "What do you add to 'wind' to make 'windy'?"

* (As you can see "oral spelling" can involve some simple oral word analysis.)

**"Practise ‘Spelling’ (wp)"**

- with print
  - this includes children “spelling” (reading the letter names) while looking at a word, or a teacher writing a word for the children to see.

  * This does not involve any real word analysis, just straight visual spellings.

**Example:**

- repeating the **spelling** of words as the teacher is pointing to them on the chart or board.
or
- when a child reads back the **spelling** of the words on a winning card when playing Word Bingo.
or
- a teacher saying the letter names as she is writing a word on the board (or after she has orally spelled it)
or
- a teacher writing a word on the board after a child has just said the word. etc.
"letter-sound corre (o/wp)"  - This can occur orally or include print. This category is basically for recording obs. of direct letter/sound teaching. It includes such things as 'S' says sssss and 'a' says a, or showing a single letter and asking what sound it makes, or asking what letter says 'sss'.

For the Phonics Handbook programme, this will often occur each time a new letter is introduced, along with its action. If your scan catches a letter-sound correspondence and the action too, still score it in this category and also the "JPAction" category, and also mark it in the 'PH' (phonics handbook) category, and write "action" in MATERIALS.

BUT if your scan catches only the action, or only the action- letter correspondence, or only the action-sound correspondence, then it would be marked just in the JPAction.

Remember, this category is for direct single letter-sound connections (oral or print)

* NOTE: If letter-sounds are being used to decode a word, it would be marked next to only the "word analysis (wp)" category.

So: - this is similar to "letter-name learn (o/wp)". BUT instead of learning individual letter names, in this category the children are learning letter sounds.

"JPActions(o/wp)"

- This category is for marking scans every time you see people doing the (Jolly Phonics) action: action alone, action connected to the letter name (verbal or print), action connected to letter sound, or action, sound, and letter name altogether
- Because the scans are only 10 seconds long you may often catch the children enthusiastically doing the action with its sound for the full ten seconds. In this case, your scan is not catching direct "letter-sounds" - so we can't mark it as such. However, we do not really want to mark it under OTHER heading, because it is still a phonics-related activity - so we mark it here.

* Remember
- If during your scan you catch action with both the sound and letter, you would mark it under the "letter-sound corresp (o/wp)", and here. (You would also mark P.H. category and write 'action' next to MATERIALS).

Example:

Your scan catches the teacher showing the letter "S" and the children are responding with the "S" action (waving their arms in the form of a snaky "S") and they are saying "SSSSS".
- You would mark your scan as follows:
(Shortened version of obs. sheet):

<table>
<thead>
<tr>
<th>Time:</th>
<th>10:41</th>
<th>10:42...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. SUPERVISION</strong></td>
<td><strong>S</strong></td>
<td><strong>EN</strong></td>
</tr>
<tr>
<td>T led</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2. ENGAGEMENT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fully engaged:active</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>6. PHONICS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Spell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prac Spell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/Sound Corresp</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>JPActions</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Word Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>13. MATERIALS/ACTIVITY</strong></td>
<td>JP flash card</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Partial version of observation sheet (with only pertinent categories showing)

**Key:**
- **S** = supervised, **NS** = nonsupervised,
- **EN** = engagement (used only if children in the scan were differently engaged),
- **Tled** = Teacher led activity,
- **4** = represents the 4 children in that particular scan.

The "JPAction" category could also include the teacher (or a student) making the actions in quick sequence and the children calling out the sounds, blending them together to determine the word. Because the action represents **both the letter and the sound**, it is considered to be a Phonics activity. You would also mark "Word Analysis" (see below), because children are analyzing (blending words, -- or segmenting for the child who is making the actions).

**Example:**
- a teacher-supervised activity, your scan of 4 children caught **one child “miming”** (acting out the sounds for a word), and **3 children closely watching**, trying to guess the word. **You would mark the scan as follows:**

  Steps - all markings would be in the “**S**” column
  1. **SUPERVISION** - mark **4** next to “T led” category.
  2. **ENGAGEMENT** - mark **1** next to “fully engaged/active”.
     - mark **3** next to “fully engaged: passive”.
  3. **PHONICS**
     - mark **4** next to “JPAction” subcategory
     - mark **4** next to “**Word analysis**” subcategory.
  4. **PHONICS HANDBOOK** - Mark 4 in this category - this is to be described later.
  5. **MATERIALS CATEGORY** - Also mark “miming actions”

*Note:* (Scan #7 p.228)


"word analysis (wp)" * It must include print *
- this category would represent the bulk of instruction for learning to read words and spell words.
- includes sounding out words (e.g. teacher pointing to each letter in ‘bat’ and saying along with the class “b-a-t” ... bbbaaattt). As you can see, this incorporates segmenting and blending with reference to printed letters and words.
- includes any kind of printed word/sound manipulations, such as “making new words”.
(eg.) “start with pin” - “now add a letter and make spin”
- “now change a letter to make spit”
- “now change the ‘Y’ sound to an ‘o’ sound; - what have you got?”
- includes pointing out and generating printed rhyming words. (e.g. cat, hat, bat, sat).
- includes generating single words (working out words), with teacher or as seat work. (e.g. worksheets asking to fill in the blank; or making a list of words that go with winter; or writing words that start with the “j” sound; etc.)

* This category is for analyzing how words are spelled using print referents.
* If there is straight spelling dictation, that would come under the “Spelling dictation(wp)” category. - If oral -- use “oral spelling(o)”.

"read/find (sight) words"-with print
- this category includes “sight” word activities. “Sight” words are words that the children know, without sounding out. In kindergarten, a child’s sight word vocabulary is very limited if it exists at all.
Scans marked in this category would catch activities like:
- reading isolated words on the word wall (or a teacher may say “find” the word “boy”, and a child goes and gets the word, shouting out the word as a teacher flashes a word card; etc.)

* Note - This is not reading sentences, just words.
** Note - If you catch someone sounding out the word, it is not marked in this category. (It would be under “word analysis”).

"spelling dictation (wp)" -with print
- this is straight dictation where the teacher says a list of words (or s/he may emphasize the phonemes -- ‘d-o-g’), and the children write the words.

Example Scenario: (Scan #8 p.228) -for different spelling categories
Your scan catches the teacher asking the children how to spell “hat”. A child responds with “h” “a” “t”. The teacher then writes it on the board. (The 4 children in your scan were participating by watching).

You would mark your scan under both “Oral Spelling(o)” (to show it was initially an oral activity) and under “Practise Spelling (wp)” (to show that children were getting to practise spelling by being exposed to the printed word. If your next scan catches the teacher “manipulating” the word (e.g. changing ‘h’ to ‘c’ or ‘t’ to ‘n’ and add ‘d’) to make new words - you would mark it next to “word analysis (wp)”
# 6 WRITING (LETTER FORMATION PRACTISE)

"specific description/prac" - teacher describes how to make letters. This includes children practising making letters in the air, tracing letters or printing letters (or words) during specific instruction.

"copy: letters/wds/sent" - this includes copying (or tracing) letters, words, and or sentences from charts, wall, or a work sheet; after or without specific instruction.

"printing own name/letters" - whenever a scan catches a child writing her or his name, or writing random letters (without specific letter formation instruction).

# 7 WRITING (Real) - NOT single words. There must be at least one sentence.

"guided" - This is very strictly directed by the teacher. It would only be used mainly when the teacher virtually tells the children exactly what to write. For example: In kindergarten the teacher may tell the children to draw a picture and write under it the sentence "This is me playing". But they would not copy this sentence. It would not be pre-written for them (although the teacher may help them to spell it). (If it were pre-written, it would come under "copy: letters/wds/sent")

"self-generated" - anything that could vaguely be taught of as a composition, from a single sentence to a story or journal writing. The teacher may provide an idea, or even a few words, but the writing is mainly generated by the student (even though the teacher may help with the spelling).

# 8 GRAMMAR/SYNTAX

- This is a single heading/category. It will often be marked along with other categories. It is to be used whenever your scan catches some grammar explanation (e.g., describing past tense) punctuation explanation (e.g. “at the end of a sentence we put a ‘full stop’ (period)”), or sentence structure (e.g., “Does that sentence make sense?”)

For example: (Scan #9 p. 228)

A number of children each have a word card (with a sight word they should know) and they are to arrange themselves so that they (and their word cards) are in the correct order to make a sentence. Your scan catches the teacher and children reading the sentence, and then the teacher explains that it doesn’t make sense. (the 4 children are reading the sentence and listening to teacher explain why it doesn’t make sense).

You would mark your scan next to the GRAMMAR / SYNTAX category, and next to “oral: (real)” category under the READING (Real) Heading (described later).
# 9 READING (Concepts of Print)

This heading reflects situations where the children are not “really” reading but are gaining something from someone else reading and/or learning about reading and print. They may "pretend read” or pseudoread, and may occasionally know one or two words of what they are “reading”, but it is mostly chanting or repeating, or “reading” memorized stories or patterns.

Category Descriptions

"T/A/V/ST reads to ch" - teacher, aide, older student (ST), (* only if really reading) or volunteer reads to a child, or to the children. (So the children are passively listening.)

"Read (chant/rep/mem)" - a child or children are "pretend reading" to (or along with) the teacher, A/V/ST themselves, or other children. This includes:

- repeating words or sentences the teacher points to but are mainly beyond their level.

- orally "reading" (memorized) material to a child or children, T/A/V/ST.

- following a 'big book' chanting story (pattern book) (e.g. "Brown bear, Brown bear, what do you see?" "I see an elephant looking at me." "Brown bear, Brown bear, what do you see?" "I see a monkey looking at me.", etc.)

- “reading” or perusing a book to themselves.

"Ch. being “read to” by ch/ST*" - child or children listening to “pretend reading”.

- This category was included to cover the situation where your scan catches children being the recipient of “fake” (pretend) reading. This will occur when a child (or children) listen to another child “read” her/his book (e.g., a pattern book), or where a child is listening to an older student (ST), who may not really be a reader, and “pretend read” to the kids. This may be a judgment call on your part. If these ST (reading buddies) appear to be really reading, score your scan under “T/A/V/ST/reads to ch”. (This means the children in your scan would be listening to real reading.)

"T reads wrds/sent on chart" - teacher (aide or volunteer) reads (or writes) words or sentences that are on a chart or board etc. These words are often beyond the children’s reading ability-- particularly in kindergarten. This often occurs during discussions (vocab/lang development time).

For example: "Things I like about winter -- snowman, snowsuits, tobogganing" etc.

Scans:

#1. If your scan catches the teacher reading this sort of thing to the class, your scan would be marked in this category.

#2. If the children are “reading” repeating along with the teacher, it would be marked under “Read (chant/rep/mem)”.  

#3. (Scan #10, p.229)  

If you really feel, as the children are “reading”, that there are some words they really know, then mark your scan under 2 categories of “Read (chant/rep/mem)”, and “read/find (sight) words” under the PRINT-RELATED AND ORAL PHONICS heading. This would show that although children are mainly “pretend reading” they are getting some practice in sight words.
Sometimes the teacher underlines certain words that the kids should know, and this will help you determine if the read/find (sight) word category should be marked along with the "Read (chant/rep/mem)" one.

(e.g.) - "Things I like about winter."

For the most part though, the words are beyond the children’s reading ability if this category is marked. When these words are too difficult, the children are not really learning the spelling, however, they are learning concepts about print (e.g., what words are, spoken words can be translated into letters and written down, there are spaces between words, long words have many letters and short words don't, etc.). And they are also developing vocab/language, some of which is high quality which need to be marked accordingly, if you hear words beyond their speaking vocabulary.

- If they can really read the sentences, only 'oral (real)' would be marked.
- If they were really reading words (lists), only 'read/find (sight) words' would be marked.

# 10 READING (Real)

This category includes any real reading in a book (or sentences) at their own level that children appear to be doing (either oral or silent).

"oral (real)" - the child (or children) is orally (really) reading to or along with the teacher, aide or volunteer, student tutor, other children, or self. We can tell whether she is with a T ( or aide, volunteer) by seeing if it is marked in the supervised (S) or non-supervised (NS) column.

"silent (real)" - the child (or children) is silently (really) reading; -- either following along in their own books as others are reading, or independently. Again - for silent reading, this will be a judgment call.

"Using Context/Pict" - a child uses the context of the story and/or the pictures to help figure out a word or words.

- This is more of a descriptive category and can be used along with a number of other categories.

For example: (Scan #11 p.229)

- A child is "reading" a pattern book to the class that she has pretty much memorized (e.g. "Brown Bear, Brown Bear, What do you see?" "I see a giraffe looking at me"), and looks at the picture to "read" giraffe. You know if you showed the word giraffe in isolation, she wouldn’t be able to read it.

2 perspectives:
This scan would be marked a “1” next to “Read (chant/rep/mem)” and next to “Using Context/Pict”.
The 3 “listeners” in your scan are marked next to “ch being “read to” by ch/ST”.

225
You may see a child do this in real reading as well, and so mark the appropriate category. - “The man was driving the ... truck”. (The child would probably use both the context - (driving) and the picture to help figure out the word.)

* Note: (Scan #12 p.229)
This category would often be used when kids are “reading” a sentence about the date or weather using days of the week, or months of the year cards when doing routine Date/weather/sh & tell activities. It is the context of the activity which help them “read” Tuesday (for example). For this scan you would use “Read”(chant/rep/mem)” and “Using Context/Pictures” categories.

# 11 Phonics Handbook (Jolly Phonics Materials)

This is also a descriptive category which will be used along with other categories. Whenever any of the activities include part of the Jolly Phonics programme stuff, either using the materials from the Phonics Handbook and/or using the ideas, then mark your observation here as well as next to the appropriate activity category.

Please familiarize yourself again with the Jolly Phonics (in the Phonics Handbook) descriptions.

Activities that are obviously part of the Jolly Phonics Programme (in the Phonics Handbook):
1. Auditory Blending and Segmenting task (described earlier) e.g., s - pit
   sp - it
   spi - t
2. Introducing the letters and sounds by way of a story
3. Using the actions with the letters and sounds

P.H. MATERIALS
1. Sound Sheets (sheets with pictures and sound actions and letters to copy and trace)
2. Sound Books (Booklets with just the letters pasted in them that the kids use to practise the sounds)
3. Flash Cards with the pre-printed P.H. letters on them
4. Reproducible Games and Worksheets (see your Phonics Handbook for examples of these P.H. materials)
5. Word Boxes (or Bags) (see your Phonics Handbooks)
6. Jigletts (these are extra P.H. puzzles some teachers have that are pictures of something and a word printed on it - I think there are only 4 - a pig, a boat, a train and ?
7. Stencils (extras too - a stencil of a dog, or pig etc. with the stencil of the word underneath).

pig
There are other things that the Jolly Phonics programme says to do, like detecting sounds in words etc. (e.g. who has an a sound in their name). BUT other programmes do this too.

Mark the P.H. category when the activity (materials) are **uniquely the Jolly Phonics programme**.
- A lot of the Auditory Phonemic Awareness will be Jolly Phonics as they stress “auditory training”.
- But other auditory stuff - like rhyming and songs are common teaching practices, so do not mark these as P.H. But if the “actions” are incorporated into the activities, then **do** mark the P.H. category. These actions are unique to the Jolly Phonics programme.

* If in doubt about an activity - don’t mark P.H.
- Although this may sound difficult to determine, it is usually pretty clear when the P.H. components are being used, either alone, or as part of something else.

### # 12 OTHER

There are 2 blank rows under OTHER. Mark your scans here (in one of these) for any activity other than those described above (e.g. math). BE SURE to continue the scanning process through all activities, no matter how unrelated they are to “language arts”. You may find that even in unrelated activities, opportunities for phonemic awareness, or language development will arise, so BE SURE to mark the other appropriate categories as well.

Also, if there is a 'language arts' type of activity that is not on the list described above, fill it in under the OTHER section... BUT BE SURE that it is consistently used on all the pages, or if not... be sure it is clearly marked (write as much info as needed to explain) what you have intended.

* If you can, try to describe the activity taking place if it is not evident from the categories. For example: “T/A/V/ST reads to ch” is self explanatory.
  But children making a sentence out of individual word cards may not be so clear by just reading the scored observation sheet. (This is just for extra information, if possible).

Example for OTHER: (Scan #13 p.229)
4 kids colouring their “sound sheets” by themselves.

### # 13 MATERIALS

Please give a fairly clear account of what types of materials are used (e.g., 'big book', phonics work sheets, flash cards, blackboard or charts, readers, plastic/wooden letters, etc.). It is not necessary to record the same thing in every scan. Just put it down once, and write in the new materials under the most recent scan after materials have changed.
<table>
<thead>
<tr>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP TYPE</td>
</tr>
<tr>
<td>1. SUPERVISION</td>
</tr>
<tr>
<td>T. told</td>
</tr>
<tr>
<td>2. ENGAGEMENT</td>
</tr>
<tr>
<td>a. fully engaged: active</td>
</tr>
<tr>
<td>b. fully engaged: passive</td>
</tr>
<tr>
<td>c. somewhat engaged: active</td>
</tr>
<tr>
<td>d. somewhat engaged: passive</td>
</tr>
<tr>
<td>Prepare/Finish: Off</td>
</tr>
<tr>
<td>3. ORAL VOCABULARY</td>
</tr>
<tr>
<td>T: disc. stry; pict; Sh &amp; tell</td>
</tr>
<tr>
<td>T: other</td>
</tr>
<tr>
<td>CH: date/weather/attend</td>
</tr>
<tr>
<td>CH: disc stry; pict; Sh &amp; tell</td>
</tr>
<tr>
<td>CH: other</td>
</tr>
<tr>
<td>4. AUDITORY PHONEMIC AWARENESS</td>
</tr>
<tr>
<td>blending</td>
</tr>
<tr>
<td>segmenting words / sounds</td>
</tr>
<tr>
<td>detect sounds in words</td>
</tr>
<tr>
<td>explicit rhyme practice</td>
</tr>
<tr>
<td>thymes/poems/song</td>
</tr>
<tr>
<td>tongue twisters</td>
</tr>
<tr>
<td>5. PRINT-RELATED AND ORAL PHONICS: orally = (o); with print = (wp)</td>
</tr>
<tr>
<td>letter-name learn (o/wp)</td>
</tr>
<tr>
<td>Oral Spelling (o)</td>
</tr>
<tr>
<td>practice spelling (wp)</td>
</tr>
<tr>
<td>letter-sound correct (o/wp)</td>
</tr>
<tr>
<td>6. WRITING (Letter Formation Practice)</td>
</tr>
<tr>
<td>specific description/prac</td>
</tr>
<tr>
<td>copy: letters/words/sent.</td>
</tr>
<tr>
<td>printing own name/other</td>
</tr>
<tr>
<td>7. WRITING (Real)</td>
</tr>
<tr>
<td>guided</td>
</tr>
<tr>
<td>self-generated</td>
</tr>
<tr>
<td>8. GRAMMAR/SYNTAX</td>
</tr>
<tr>
<td>9. READING (Concepts of print)</td>
</tr>
<tr>
<td>T/A/V/I reads to ch</td>
</tr>
<tr>
<td>&quot;Read&quot; (chart/rep/men)</td>
</tr>
<tr>
<td>Ch. being &quot;read&quot; to/with</td>
</tr>
<tr>
<td>T. reads words/contacts</td>
</tr>
<tr>
<td>10. READING (Real)</td>
</tr>
<tr>
<td>oral (read)</td>
</tr>
<tr>
<td>silent (read)</td>
</tr>
<tr>
<td>11. PHONICS HANDBOOK</td>
</tr>
<tr>
<td>12. OTHER</td>
</tr>
<tr>
<td>colouring</td>
</tr>
<tr>
<td>spelling dictation (w)</td>
</tr>
<tr>
<td>13. MATERIALS</td>
</tr>
<tr>
<td>library book</td>
</tr>
<tr>
<td>toy</td>
</tr>
<tr>
<td>228</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>7.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>8.</td>
</tr>
<tr>
<td>9.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
|      | Ch brings read "in/over"
T reads words/extracted |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
| 10.  | READING (real) |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | oral/real |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | silent (real) |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | using context/pict |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
| 11.  | PHONICS HANDBOOK |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
| 12.  | OTHER |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | Colouring |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | Spelling dictation (on) |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
| 13.  | MATERIALS |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | children's story to make sentence |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | Word on chart |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | Pattern Book |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |
|      | Days of the week |     |     |     |     |      |       |       |       |       |       |       |       |       |       |       |

---

229 Sound Sheets
Appendix D

Classroom Observations:

Reliability

Data Calculation Procedure: Percentage of School Day
Observations: Inter-rater Reliability

Reliability of the observation measure was established through inter-rater observations. Five of the 60 observations (8%) were conducted with two researchers observing the same classes (the experimenter was always one of the inter-raters). The reliability for each of the literacy component totals was calculated using the percentage method, a traditional method used for calculating the reliability of duration data (Hartmann, 1975; Herbert 1973) in which two observers record the total time that a "behavior" (or in this case an activity) occurs within the observation session. Then, by dividing the smaller total by the larger total time, the reliability coefficient is calculated. Since, in the current study, the number of raw observations recorded in a Literacy Component category is a representation of the time spent in that type of activity, it was felt these could be considered duration data, and that the percentage method was an appropriate method of reliability measurement.

It is acknowledged that the percentage method of calculating reliability may provide a less stringent estimate than does the "exact agreement" formula -- Agreement of behaviors /Agreement + Disagreement (Saudergrass & Lentz, 1986). However, since the two observers in the current study were not agreeing on specific instances of a behavior (activity) taking place by tracking the same individuals at the same time, the exact agreement method cannot be used. Although observers tracked all the same individuals, specific children were not necessarily being observed at the same time (per observer). Therefore, the percentage method used relies on total frequencies of behavior (activities) seen. It is believed that this presents a reasonably reliable picture of the overall distribution of activity time throughout the school day.

In some of these inter-rater protocols, a few Literacy Component activities occurred rarely resulting in minimal (or no) recorded observations. So, using Chi-Square tests of significance, the averages of the combined observations in each category were also compared (i.e., category averages of the five experimenter protocols compared with those of the five inter-rater protocols). Where no (or extremely low) instances of activities in a category were seen in the set of inter-rater protocols, reliabilities were not calculated.

Procedure

Percentage Method
In each set of inter-rater observations, the two observers differed slightly in the total number of observations made throughout the entire observation period, therefore, raw observations were converted to percentages for each observer before reliability was calculated. That is, the total
number of raw observations for each Literacy Component was first converted to a percentage of the overall total number of observations (Grand Total).

**Example:**
Observer #1 had a Grand Total of 521 observations and 143 of these raw observations were in the Vocabulary category. The percentage of all the observations that was composed of observations in the Vocabulary category was \((143/521) \times 100 = 27.447\%\).

Observer #2 had a Grand Total of 502 observations, and 112 of these raw observations were in the Vocabulary category. The percentage of all the observations that was composed of observations in the Vocabulary category was \((112/502) \times 100 = 22.31\%\).

The reliability for this category was then calculated by dividing the smaller total time(percentage) by the larger, resulting in a coefficient of .813.

This was done for each Literacy Component category for the 5 sets of inter-rater observations. The five reliability coefficients for each category were then averaged to present the mean reliability coefficient for each category. For some of the category averages, fewer than the full five sets of protocols were used because: i) in one (or more) of the protocols there were no observations in that category (and the complete agreement would inflate reliability means); or, ii) there were very few (i.e., 2-3) raw observations in one protocol, and 0 in the other (rendering computation of a coefficient impossible due to division by 0). As well so few observations would unduly lower the reliability estimate.

Table 1 presents the averages, standard deviations, and ranges of the percentage coefficients for each of the main Literacy Component categories and "Other" (i.e., nonliteracy-related activities). In addition, the average number of raw observations per category (for the 5 sets of inter-rater protocols) are shown, indicating the reason for lower reliabilities in some categories; that is, categories with few observations (i.e., where very few instance of a particular activity took place) will have lower reliabilities. This is a difficulty which often occurs with low-occurrence reliability ratings (Saudergrass & Lentz, 1986). With the current study's time sampling technique, if a behavior/activity occurred rarely, then the rare occurrence could easily be missed.
Table 1.

**Average Reliability Coefficients for Literacy Component Categories**
**(and Raw Observations per Category)**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Raw Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Phonics</td>
<td>4</td>
</tr>
<tr>
<td>APA</td>
<td>5</td>
</tr>
<tr>
<td>StWd</td>
<td>3</td>
</tr>
<tr>
<td>Grammar*</td>
<td></td>
</tr>
<tr>
<td>RIrd*</td>
<td></td>
</tr>
<tr>
<td>ConPrint</td>
<td>5</td>
</tr>
<tr>
<td>RIWrite*</td>
<td></td>
</tr>
<tr>
<td>LForm</td>
<td>3</td>
</tr>
<tr>
<td>LtNmLn</td>
<td>3</td>
</tr>
<tr>
<td>Vocab</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
</tr>
</tbody>
</table>

**Note:** *coefficients not computed — too few or no instances observed on inter-rater protocols*

**Key:**
- N = Number of inter-rater sets included in calculations
- APA = Auditory Phonemic Awareness, StWd = Sight Word, RIrd = Real Reading,
- ConPrint = Concepts of Print, RIWrite = Real Writing, LForm = Letter Formation,
- LtNmLn = Letter Name Learning, Vocab = Vocabulary/Language Development,
- Other = Other nonliteracy activities.

**Chi-Square Calculations**

The number of raw observations in each Literacy Component category was tallied and averaged for each set of protocols — five experimenter protocols (Experimenter Set 1), and, five inter-rater protocols (Inter-rater Set 2). The remainder of the total overall observations (all but the category of interest) was tallied and averaged for each set. A contingency table was constructed for each of the Literacy Component categories with Set (Set 1 vs. Set 2) as one factor and Category Observations (average number of observations in category vs. average number of observations not in category) as the second factor. Chi-Square tests of significance were then performed on each contingency table representing each Literacy Component category. The results in Table 2 demonstrate that there were no significant differences between Set 1 (5 experimenter protocols) and Set 2 (5 inter-rater...
protocols) in any of the Literacy Component categories tested. Three Literacy Components could not be tested (Grammar, Real Reading, and Real Writing) due to the low number, or lack, of observations in these categories for these sets of protocols.

Table 2.
Total All (Raw) Observations for Each Literacy Component Category by Experimenter Protocols vs. Inter-rater Protocols

<table>
<thead>
<tr>
<th>Set 1 (Experimenter)</th>
<th>Set 2 (Inter-rater)</th>
<th>$\chi^2$</th>
<th>$p &lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD ( $\bar{X}$ TObs)</td>
<td>M</td>
<td>SD ( $\bar{X}$ TObs)</td>
</tr>
<tr>
<td>Phonics</td>
<td>18.2</td>
<td>13.5 (463)</td>
<td>21.8</td>
</tr>
<tr>
<td>APA</td>
<td>52.4</td>
<td>53.5 (463)</td>
<td>51.4</td>
</tr>
<tr>
<td>StWd</td>
<td>4.8</td>
<td>4.4 (463)</td>
<td>2.8</td>
</tr>
<tr>
<td>ConPrint</td>
<td>43.2</td>
<td>19.2 (463)</td>
<td>41.2</td>
</tr>
<tr>
<td>LForm</td>
<td>8.8</td>
<td>8.6 (463)</td>
<td>5.8</td>
</tr>
<tr>
<td>LtNmLn</td>
<td>13.6</td>
<td>23.4 (463)</td>
<td>16.0</td>
</tr>
<tr>
<td>Vocab</td>
<td>139.8</td>
<td>37.3 (463)</td>
<td>135.6</td>
</tr>
<tr>
<td>Other</td>
<td>207.0</td>
<td>36.4 (463)</td>
<td>207.4</td>
</tr>
</tbody>
</table>

$\chi^2(1, N = 920)$

Note: *2 cells with expected counts less than 5.0

Key: ( $\bar{X}$ TObs ) = average of total number of observations per set

APA = Auditory Phonemic Awareness, StWd = Sight Word, ConPrint = Concepts of Print,
LForm = Letter Formation, LtNmLn = Letter Name Learning, Vocab = Vocabulary/Language Development
Other = Other nonliteracy activities.

1 Chi-Square tests were also performed on individual sets of inter-rater categories with the lowest reliabilities. The results also showed that there were no significant differences between the individual sets of these inter-rater categories.
Percentage of School Day Calculations for Observations

In order to determine how much of the school day was spent in various activities, percentages were calculated using scans and numbers of observed children in a scan.

Scan = 1 per minute
Observations = ranged from 3-5 per scan (but usually 4).

A. Steps for percentage calculations of activity time.
Using Scans:

1. Total Scans = the total number of minutes of observed time.
   EXAMPLE:
   tot scans = 111 = total number of minutes of observed time = total activity time.

2. Percentage of the complete school time that is spent in activity (total activity time) =
   (Total scans/total minutes of school) x 100.
   EXAMPLE:
   (111/150) x 100 = 74%

B. Steps for percentage calculations of non-literacy time.
Using Minutes:

1. Percentage of school time that was spent on transitions (e.g., 20 min trans.), or outdoor play etc.
   (Number of transition minutes/total minutes of school) x 100.
   EXAMPLE:
   (20/150) x 100 = 13.33%

C. Steps for percentage calculations of literacy activities.
Using Observations:

1. The number of minutes spent in each literacy activity must first be determined:
   i) the average number of observations (ANO) per minute = total observations / total scans.
   ii) total number of minutes (in a literacy category) = total observations in that literacy category / ANO
2. Finally, the percentage of the school day spent on various activities =
   (literacy category minutes/total minutes in school day) x 100.

EXAMPLE:
Minutes in full school day = 150
Total observations = 438
Total Scans = 111
Total Observations in Vocabulary Category = 64

Percentage of school day spent in Vocabulary activities:
   i) average number of obs per minute (ANO) = 438 / 111 = 3.9459459
   ii) minutes spent in Oral Vocab = 64 / 3.9459459 = 16.219178 minutes.
   iii) percentage of school day = (16.219178 / 150) x 100 = 10.81278 = 10.81%
APPENDIX E

Outcome Data Alternate Analyses:

MANOVA's Using Class Means
Alternate MANOVA's for Phase 2 Outcome Measures
Using Class Means

Table 1.
Set 1 (Standard Scores Measures)
Class Means of Jolly Phonics vs. Control

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 10)</th>
<th>Control (n = 10)</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>WaSS</td>
<td>99.33 ±8.97</td>
<td>86.34 ±4.85</td>
<td>16.21</td>
<td>.001</td>
<td>2.68</td>
</tr>
<tr>
<td>WrRdSS</td>
<td>106.48 ±4.90</td>
<td>100.55 ±4.09</td>
<td>8.62</td>
<td>.01</td>
<td>1.45</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>103.78 ±5.35</td>
<td>97.87 ±4.40</td>
<td>7.27</td>
<td>.015</td>
<td>1.34</td>
</tr>
</tbody>
</table>

MANOVA, F(3,16) = 5.05, p = .011.

Table 2.
Set 2 (Full Word Measures)
Class Means of Jolly Phonics vs. Control

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 10)</th>
<th>Control (n = 10)</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>WrRdWrds</td>
<td>2.51 ±1.12</td>
<td>1.28 ±0.42</td>
<td>10.79</td>
<td>.01</td>
<td>2.92</td>
</tr>
<tr>
<td>B&amp;R</td>
<td>5.81 ±4.14</td>
<td>2.86 ±1.26</td>
<td>4.64</td>
<td>.05</td>
<td>2.34</td>
</tr>
<tr>
<td>WrSpWrds</td>
<td>1.10 ±0.86</td>
<td>0.44 ±0.29</td>
<td>5.23</td>
<td>.05</td>
<td>2.28</td>
</tr>
<tr>
<td>WaRS</td>
<td>3.36 ±2.11</td>
<td>0.96 ±0.86</td>
<td>11.09</td>
<td>.01</td>
<td>2.79</td>
</tr>
<tr>
<td>NwRd</td>
<td>2.74 ±1.84</td>
<td>0.78 ±0.55</td>
<td>10.41</td>
<td>.01</td>
<td>3.56</td>
</tr>
</tbody>
</table>

MANOVA, F(5,14) = 5.54, p < .01.
Key: WrRdWrds = WRAT Reading Full Words, B&R = Burns & Roe Word Recognition, WrSpWrds = WRAT Spelling Full Words, WaRS = Woodcock Word Attack Raw Score, NwRd = Nonword Reading Task,
Table 3.

Set 3 (Phonemic Awareness, Skills, Coding, and Phonemic Analysis)
Class Means of Jolly Phonics vs. Control

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (n = 10)</th>
<th>Control (n = 10)</th>
<th>F</th>
<th>p &lt;</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Subset A:  Phonemic Awareness and Basic Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAAS</td>
<td>4.00</td>
<td>1.32</td>
<td>3.38</td>
<td>0.89</td>
<td>1.52</td>
</tr>
<tr>
<td>LtNm</td>
<td>21.53</td>
<td>1.99</td>
<td>20.25</td>
<td>1.75</td>
<td>2.31</td>
</tr>
<tr>
<td>WrtAlph</td>
<td>13.29</td>
<td>2.37</td>
<td>12.23</td>
<td>2.39</td>
<td>0.99</td>
</tr>
<tr>
<td>Subset B:  Alphabetic Coding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/SRecog</td>
<td>17.04</td>
<td>4.38</td>
<td>11.85</td>
<td>4.20</td>
<td>7.33</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>19.15</td>
<td>6.56</td>
<td>10.30</td>
<td>4.25</td>
<td>12.80</td>
</tr>
<tr>
<td>Subset C: Phonemic Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WrRdPh</td>
<td>15.59</td>
<td>4.14</td>
<td>9.37</td>
<td>3.45</td>
<td>13.33</td>
</tr>
<tr>
<td>B&amp;RPh</td>
<td>24.37</td>
<td>11.93</td>
<td>15.90</td>
<td>6.52</td>
<td>3.89</td>
</tr>
<tr>
<td>WrSpPh</td>
<td>27.17</td>
<td>10.84</td>
<td>15.96</td>
<td>5.74</td>
<td>8.36</td>
</tr>
<tr>
<td>WaPh</td>
<td>13.49</td>
<td>3.64</td>
<td>5.28</td>
<td>2.96</td>
<td>30.61</td>
</tr>
<tr>
<td>NwRdPh</td>
<td>21.37</td>
<td>5.89</td>
<td>8.86</td>
<td>3.21</td>
<td>34.81</td>
</tr>
<tr>
<td>NwSpPh</td>
<td>27.69</td>
<td>8.27</td>
<td>13.53</td>
<td>5.96</td>
<td>19.30</td>
</tr>
</tbody>
</table>

MANOVA, $F(11, 8) = 3.34$, $p < .05$.

*Note: approaching significance

Key:
Subset A:  TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, WrtAlph = Write Alphabet Task
Subset B:  L/SRecog = Letter-Sound Recognition Task, L/SRecall = Letter-Sound Recall Task
Subset C:  WrRdPh = WRAT Reading — Phonemic Analysis, B&RPh = Burns and Roe Word Recognition — Phonemic Analysis, WrSpPh = WRAT Spelling — Phonemic Analysis
APPENDIX F

At-Risk Subsample Alternate Analyses:

Phase 1 (Pretest) and Phase 2 (Outcome) Comparisons: Jolly Phonics vs. Control Group At-Risk Children (with ESL participants removed)
At-Risk (minus ESL) Subsample Comparisons

Table F1.

Phase 1 At-Risk Comparisons (Pretest Measures)
Jolly Phonics (nonESL) AR vs. Control (nonESL) AR Children

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (nonESL) AR (n=26)</th>
<th>Control (nonESL) AR (n=25)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1.80 1.80</td>
<td>1.88 1.27</td>
<td>0.11</td>
<td>ns</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LtNm</td>
<td>5.73 4.46</td>
<td>6.48 4.36</td>
<td>0.37</td>
<td>ns</td>
</tr>
<tr>
<td>L/SRecog</td>
<td>1.50 2.32</td>
<td>1.48 1.81</td>
<td>0.44</td>
<td>ns</td>
</tr>
<tr>
<td>L/Recall</td>
<td>1.39 1.42</td>
<td>1.28 1.43</td>
<td>0.90</td>
<td>ns</td>
</tr>
<tr>
<td>B&amp;R*</td>
<td>0.00 0.00</td>
<td>0.04 0.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MANOVA, F(4, 46) = 0.362, p = .872.

Key: TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, L/SRecog = Letter-Sound Recognition Task, L/Recall = Letter-Sound Recall Task, B&R = Burns and Roe Word Recognition

* Note: B&R was not included in analysis due to lack of sufficient variance in both AR groups.

Table F2.

Phase 2 At-Risk Comparisons (Outcome Measures)
Jolly Phonics (nonESL) AR vs. Control (nonESL) AR Children: Set 1 (Standard Score Measures plus Burns and Roe Word Recognition)

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (nonESL) AR (n=26)</th>
<th>Control (nonESL) AR (n=25)</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>90.07 23.21</td>
<td>86.16 16.69</td>
<td>.47</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WaSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WrRdSS</td>
<td>98.31 11.81</td>
<td>87.88 11.40</td>
<td>10.28</td>
<td>.01</td>
<td>.92</td>
</tr>
<tr>
<td>WrSpSS</td>
<td>94.12 15.88</td>
<td>85.04 13.17</td>
<td>4.92</td>
<td>.05</td>
<td>.69</td>
</tr>
<tr>
<td>B &amp; R*</td>
<td>1.81 2.65</td>
<td>0.20 0.41</td>
<td>8.97</td>
<td>.01</td>
<td>3.93</td>
</tr>
</tbody>
</table>

MANOVA, F(4, 68) = 3.57, p<.05.


* Note: B&R was included in Set 1 outcome analysis because it was the only Set 2 variable that could be successfully transformed.
As in the main At-Risk analyses (*with* ESL children), a MANOVA could not be performed on the Set 2 (Full Word Raw Score) measures (except for the Burns and Roe Word Recognition measure included in the Set 1 analysis) due to very low/no scores for the control children. However, Table F3 displays the descriptive statistics for the (nonESL) At-Risk subsample presented here (i.e., means and standard deviations) for the Full Word Raw Score measures.

**Table F3.**

**Phase 2 At-Risk Comparisons (Outcome Measures)**

*Jolly Phonics (nonESL) AR vs. Control (nonESL) AR Children: Set 2 (Full Word Raw Score Measures)*

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (nonESL) AR</th>
<th>Control (nonESL) AR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 26)</td>
<td>(n = 25)</td>
</tr>
<tr>
<td>WrRdWrds</td>
<td>0.81 1.33</td>
<td>0.08 0.28</td>
</tr>
<tr>
<td>WrSpWrds</td>
<td>0.46 0.86</td>
<td>0 0</td>
</tr>
<tr>
<td>WaRS</td>
<td>1.56 2.73</td>
<td>0 0</td>
</tr>
<tr>
<td>NwRd</td>
<td>0.96 2.43</td>
<td>0 0</td>
</tr>
</tbody>
</table>

Key: **WrRdWrds** = WRAT Reading Full Words, **WrSpWrds** = WRAT Spelling Full Words, **WaRS** = Woodcock Word Attack—Raw Score, **NwRd** = Nonword Reading Task
### Phase 2 At-Risk Comparisons (Outcome Measures)

**Jolly Phonics (nonESL) AR vs. Control (nonESL) AR Children:**

**Set 3** (Phonemic Awareness, Basic Skills, Coding, and Phoneme Analysis Measures)

<table>
<thead>
<tr>
<th></th>
<th>Jolly Phonics (nonESL) AR (n=26)</th>
<th>Control (nonESL) AR (n=25)</th>
<th>F</th>
<th>p&lt;</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subset A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(Phonemic Awareness &amp; Basic Literacy Skills)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAAS</td>
<td>3.19 1.83</td>
<td>2.80 1.38</td>
<td>0.74</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>LtNm</td>
<td>16.38 8.44</td>
<td>11.24 5.94</td>
<td>6.29</td>
<td>.05</td>
<td>.87</td>
</tr>
<tr>
<td>WrtAlph</td>
<td>7.35 4.21</td>
<td>6.56 2.90</td>
<td>0.60</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td><strong>Subset B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(Alphabetic Coding)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/SRecog</td>
<td>13.50 7.09</td>
<td>4.31 4.52</td>
<td>30.20</td>
<td>.001</td>
<td>2.03</td>
</tr>
<tr>
<td>L/SRecall</td>
<td>13.92 7.33</td>
<td>2.52 3.58</td>
<td>49.15</td>
<td>.001</td>
<td>3.18</td>
</tr>
<tr>
<td><strong>Subset C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(Phonemic Analysis)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WrRdPh</td>
<td>6.77 8.64</td>
<td>1.24 2.65</td>
<td>9.39</td>
<td>.01</td>
<td>2.09</td>
</tr>
<tr>
<td>B&amp;RPh</td>
<td>12.65 18.59</td>
<td>1.68 2.50</td>
<td>8.55</td>
<td>.01</td>
<td>4.39</td>
</tr>
<tr>
<td>WrSpPh</td>
<td>16.54 13.43</td>
<td>4.12 5.32</td>
<td>18.57</td>
<td>.001</td>
<td>2.34</td>
</tr>
<tr>
<td>WaPh</td>
<td>8.41 8.14</td>
<td>0.76 1.90</td>
<td>20.99</td>
<td>.001</td>
<td>4.03</td>
</tr>
<tr>
<td>NwRdPh</td>
<td>14.77 13.62</td>
<td>1.28 1.75</td>
<td>24.14</td>
<td>.001</td>
<td>7.71</td>
</tr>
<tr>
<td>NwSpPh</td>
<td>18.19 13.95</td>
<td>2.04 2.62</td>
<td>32.41</td>
<td>.001</td>
<td>6.16</td>
</tr>
</tbody>
</table>

**MANOVA, F(11, 39) = 5.97, p<.001.**

**Key:**

**Set 3 Measures:**

- **Subset A:** TAAS = Rosner Test of Auditory Analysis Skills, LtNm = Letter Name Task, WrtAlph = Write Alphabet Task
- **Subset B:** L/SRecog = Letter-Sound Recognition Task, L/SRecall = Letter-Sound Recall Task
- **Subset C:** WrRdPh = WRAT Reading -- Phonemic Analysis, B&RPh = Burns and Roe Word Recognition -- Phonemic Analysis, WrSpPh = WRAT Spelling -- Phonemic Analysis, WaPh = Woodcock Word Attack -- Phonemic Analysis, NwRdPh = Nonword Reading Task -- Phonemic Analysis, NwSpPh = Nonword Spelling Task -- Phonemic Analysis