DOCUMENT RESUME

ED 049 910 24 RE 003 626

AUTHOR	Johnson, Dale D.
TITLE	Factors Related to the Pronunciation of Vowel Clusters. Part I (of 3 Parts).
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO	TR-149-1
BUREAU NO	BR-5-0216
PUB DATE	Sep 70
CONTRACT	OEC-5-10-154
NOTE	85p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Child Language, *Elementary Grades, Language Research, *Pronunciation, Reading Ability, Reading Research, *Vowels, *Word Recognition

ABSTRACT

Children's pronunciations of vowel clusters in synthetic words were analyzed in relation to common English words containing the same vowel clusters. Subjects were 436 elementary-school students of both high and low reading levels from a suburban, an urban, and a rural community. Independent variables were grade level, sex, reading level, community type, response type (principle and secondary), and vowel cluster (ai, au, ay, ea, ie, oa, oo, ou, and ow). Dependent variables were difference scores between subsets' principle and secondary pronunciations of vowel clusters and the principle and secondary pronunciation proportions of vowel clusters on a 1963 modification of the Thorndike Frequency Count (type corpus) and on the 1967 Kucera and Francis computational analysis of modern American English (token corpus). Statement and discussion of the problem, development of the test instrument, and study procedures are reported in Part 1. Results and conclusions of the study are given in Part 2, RE 003 627, and appendixes are found in Part 3, RE 003 628. Tables are included. (VJ)

TECHNICAL REPORT

U.S. Office of Education Center No. C-03 Contract OE 5-10-154

Technical Report No. 149 Part I (of 3 Parts)

FACTORS RELATED TO THE PRONOUNCIATION OF VOWEL CLUSTERS

Report from the Project on Basic Pre-Reading Skills: Identification and Improvement

Richard L. Venezky, Principal Investigator

By Dale D. Johnson

Thomas Barrett, Professor of Curriculum & Instruction, Elementary Education
Chairman of the Examining Committee

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

September, 1970

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

NATIONAL EVALUATION COMMITTEE

Samuel Brownell
Professor of Urban Education
Graduate School
Yale University

Henry Chauncey
President
Educational Testing Service

Elizabeth Koontz
Wage and Labor Standards Administration, U.S. Department of Labor, Washington

Patrick Suppes
Professor
Department of Mathematics
Stanford University

Launor F. Carter
Senior Vice President on Technology and Development
System Development Corporation

Martin Deutsch
Director, Institute for Developmental Studies
New York Medical College

Roderick McPhee
President
Punahou School, Honolulu

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

Francis S. Chase
Professor
Department of Education
University of Chicago

Jack Edling
Director, Teaching Research Division
Oregon State System of Higher Education

G. Wesley Sowards
Director, Elementary Education
Florida State University

RESEARCH AND DEVELOPMENT CENTER POLICY REVIEW BOARD

Leonard Berkowitz
Chairman
Department of Psychology

Russell J. Hosler
Professor, Curriculum and Instruction

Stephen C. Kleene
Dean, College of Letters and Science

B. Robert Tabachnick
Chairman, Department of Curriculum and Instruction

Archie A. Buchmiller
Deputy State Superintendent
Department of Public Instruction

Clauston Jenkins
Assistant Director
Coordinating Committee for Higher Education

Donald J. McCarty
Dean
School of Education

Henry C. Weinlick
Executive Secretary
Wisconsin Education Association

Robert E. Grinder
Chairman
Department of Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Ira Sharkansky
Associate Professor of Political Science

M. Crawford Young
Associate Dean
The Graduate School

EXECUTIVE COMMITTEE

Edgar F. Borgatta
Brittingham Professor of Sociology

Robert E. Davidson
Assistant Professor, Educational Psychology

Russell J. Hosler
Professor of Curriculum and Instruction and of Business

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Anne E. Buchanan
Project Specialist
R & D Center

Frank H. Farley
Associate Professor, Educational Psychology

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

Robert G. Petzold
Associate Dean of the School of Education
Professor of Curriculum and Instruction and of Music

Robin S. Chapman
Research Associate
R & D Center

FACULTY OF PRINCIPAL INVESTIGATORS

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor of Educational Psychology

James Moser
Assistant Professor of Mathematics Education; Visiting Scholar

Richard L. Venezky
Assistant Professor of English and of Computer Sciences

Ted Czajkowski
Assistant Professor of Curriculum and Instruction

Lester S. Golub
Lecturer in Curriculum and Instruction and in English

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Alan Voelker
Assistant Professor of Curriculum and Instruction

Robert E. Davidson
Assistant Professor of Educational Psychology

John G. Harvey
Associate Professor of Mathematics and of Curriculum and Instruction

Milton O. Pella
Professor of Curriculum and Instruction (Science)

Larry Wilder
Assistant Professor of Curriculum and Instruction

Gary A. Davis
Associate Professor of Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Thomas A. Romberg
Associate Director, R & D Center
Professor of Mathematics and of Curriculum and Instruction

Peter Wolff
Assistant Professor of Educational Psychology

M. Vere DeVault
Professor of Curriculum and Instruction (Mathematics)

Donald Lange
Assistant Professor of Curriculum and Instruction

B. Robert Tabachnick
Chairman, Department of Curriculum and Instruction

MANAGEMENT COUNCIL

Herbert J. Klausmeier
Director, R & D Center
V.A.C. Henmon Professor of Educational Psychology

Thomas A. Romberg
Associate Director

James Walter
Director
Dissemination Program

Dan G. Woolpert
Director
Operations and Business

Mary R. Quilling
Director
Technical Development Program

* COMMITTEE CHAIRMAN

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from the Basic Pre-Reading Skills: Identification and Improvement Project in Program 1. General objectives of the Program are to generate new knowledge about concept learning and cognitive skills, to synthesize existing knowledge, and to develop educational materials suggested by the prior activities. Contributing to these Program objectives, this project's basic goal is to determine the processes by which children aged four to seven learn to read and to identify the specific reasons why many children fail to acquire this ability. Later studies will be conducted to find experimental techniques and tests for optimizing the acquisition of skills needed for learning to read.

ACKNOWLEDGMENTS

Like most other doctoral students, I owe more debts than I can ever repay, or even recall. Certain people, however, have been particularly helpful in the development of this thesis and with my graduate program in general.

I am very grateful to Dr. Thomas Barrett, my major professor, whose advice and support throughout my program and with this study have been invaluable. Sincere appreciation is expressed to Dr. Richard Venezky under whose guidance the idea for this study was developed and whose assistance and suggestions were very helpful. Both of these men gave me more of their time than I had any right to expect.

Gratitude is expressed to Dr. Robert Calfee, Dr. Robin Chapman, Dr. Ken Dulin, Dr. Albert Yee, and Dr. Richard Smith whose suggestions and advice were valuable in the design and preparation of this investigation. Thanks are also extended to Dr. Charles Scott, my minor professor, whose suggestions throughout my program were always useful.

I am indebted to my good friend and colleague, Dr. Robert Heidman, who helped me in so many ways during my three years in Madison.

I also wish to thank the students, teachers and administrators in the participating schools in Waterloo, Seneca, Cedarburg and Racine, Wisconsin, whose cooperation and support made this study possible.

Gratefulness is expressed to Mrs. Marilyn Mantei, Computer Programmer, and to the University of Wisconsin Computer Center for valuable assistance. The author is indebted to the Wisconsin Research and Development Center for Cognitive Learning for providing funding and special assistance throughout this study.

Special thanks are extended to Dr. Joe Strain, Dr. Naim Gupta, and Dr. Edward Merryman, all of Ball State University, for their help and encouragement.

I am grateful to Mrs. Sharon Volovsek, Mrs. Shirley McCormick, Mrs. Kathleen Hartung, Miss Cheryl Lee, and Mrs. Betty Merryman for clerical and typing assistance in the preparation of classroom materials and the several drafts of this manuscript.

Finally, my deepest gratitude is extended to my wife, Sandy, and to my children, Lisa, Julie and Kirk, whose patience, encouragement and understanding led to the successful completion of this program.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Statement of the Problem

The purpose of this study was to determine the pronunciations children give to synthetic words containing vowel cluster spellings, and to analyze the observed pronunciations in relation to common English words containing the same vowel clusters. The pronunciations associated with vowel cluster spellings are among the most unpredictable letter-sound correspondences in English. If learning to read includes learning to translate from spelling to sound, then vowel clusters should pose a particularly difficult problem for children. Determining the manner in which children solve this problem--i.e., the factors related to children's pronunciations of vowel clusters in unfamiliar words--could shed more general light on this complex decoding act.

The study dealt with the following independent variables: grade level (second, fourth, and sixth), sex, reading level (high and low), community type (suburban, urban, and rural), vowel cluster (a subset of nine--ai, au, ay, ea, ie, oa, oo, ou, and ow), and response type (principal and secondary). The dependent variables were difference scores between the subjects' principal and secondary pronunciations of vowel clusters and the principal and secondary pronunciation proportions of vowel clusters on two corpora--a 1963 modification of the 20,000 most common words on the Thorndike Frequency count (Type Corpus) and the most frequent 1,000 words on the 1967 Kucera and Francis computational analysis of present-day American English (Token Corpus).

Procedures

Two pilot studies were conducted to refine and modify the testing instrument, a 100 item multiple choice test. The instrument included 90 synthetic words containing vowel clusters, (ten synthetic words for each of the nine selected vowel clusters) and ten check items. Four real word distractors contained the major pronunciations for the vowel cluster on the type and token corpora.

The sample consisted of 436 elementary pupils from a suburban an urban, and a rural community, all in Wisconsin. Second, fourth and sixth grade boys and girls of both high and low reading levels were included. Each subject responded to two 50 item halves of the instrument on two consecutive days.

To test twelve hypotheses and answer three questions two analyses were performed. In each analysis the design was a 3 x 2 x 2 x 3 x 8 (or 7) x 2 analysis of variance, in which the main effects were grade, sex, reading level, community type, vowel cluster (eight on the type analysis and seven on the token analysis) and response type.

Results

1. Grade level was significantly related to vowel cluster pronunciation. There was an upward progression from second to sixth grade in the proportion of principal vowel cluster pronunciations given in both analyses.

2. There were no significant sex differences in either analysis.

3. Subjects of high reading level consistently gave more principal pronunciations to vowel clusters in both analyses than the poorer readers.

4. Suburban subjects tended to give the principal pronunciations of vowel clusters more consistently than urban and rural subjects.

5. Subjects' pronunciations were more closely related to word types than to word tokens, particularly to the principal pronunciations in the type corpus.

6. Word configuration seemed related to vowel cluster pronunciations.

Conclusions

1. As children progress through the grades, their vowel cluster pronunciations more closely parallel the correspondences occurring in common English words.

2. Being a preference inventory, not a test of "correctness", sex differences were not significant.

3. Better readers are less deviant from correspondences in common words in their pronunciation of vowel clusters than are poorer readers.

4. Suburban children tend to more closely approximate the vowel cluster correspondence frequencies in common words than urban and rural pupils.

5. Principal pronunciations of vowel clusters in word types relate more closely to children's pronunciations than do the correspondences in word tokens.

6. Contextual environment and word postion seem to influence vowel cluster pronunciations by children.

INTRODUCTION

It has been said that the act of learning to read is perhaps the greatest intellectual feat of anyone's lifetime, and teaching people to read has been the concern of educators since the development of the first alphabet. In spite of this long history of teaching reading, there is yet no universally accepted definition of "reading."

The history of reading instruction in this country, perhaps more than any other educational endeavor, has been characterized by a variety of methodologies and by missionary zeal. More research has been done on reading than any other school subject. Reading materials appear, flourish, and fade with amazing rapidity. Phrases come and go. Ten years ago "decoding" was uttered only by the unenlightened--today its popularity is immense.

In the past decade the field of reading has felt the influence of linguistic science, and the impact has been profound. Linguists study language, and their discoveries and theories have offered insights to those who would help children learn to read. Linguists have isolated the features of language, and have described their functions and relationships. Studies have been done of phonology, morphology, and syntax, of pitch, juncture and stress, and of competence and performance and of countless other related areas. Some studies have dealt with the relationship between orthography and phonology.

Language is a system of oral codes through which humans communicate, and writing is a graphic representation of these oral and audible symbols. Children learn to listen, speak, read and write--usually in that order, and most develop considerable aural/oral facility before any formal reading instruction begins. They have learned to discriminate and articulate most of the phonemes of their native language, and to comprehend and generate meaningful sentences before they enter school.

Of all the skills of language and thought, perhaps the only one unique to beginning reading is the ability to translate what is written to oral language already possessed. Symbols represent sounds; unfortunately, some symbols represent many sounds, and some sounds are represented by many symbols in English. Recent studies (Venezky, 1966; Hanna _et al._, 1966) aided by computer technology have tabulated correspondences between spelling and sound and sound and spelling in common English words. However, little research has been done to determine whether or not these relationships are actually used by competent readers--and if they are, how children acquire them.

If initial reading includes the translation from spelling to sound, and if these symbol-sound correspondences are not one to one, research is needed regarding children's pronunciation behavior. When a child encounters an unfamiliar written word, what factors influence his choice of pronunciation? The present study is concerned with this question.

Chapter I

STATEMENT, BACKGROUND AND RATIONALE OF THE PROBLEM

Statement of the Problem

This study was designed to investigate elementary pupils' pronunciations of selected vowel clusters for which predictable letter-sound correspondences rarely exist, and to compare the observed pronunciations to pronunciation frequencies of vowel clusters in common English words.

In more specific terms, the investigation dealt with the following questions:

1. How well do children's pronunciations of vowel clusters in synthetic words approximate the actual pronunciation frequencies of the same vowel clusters?
2. What differences are there in the vowel cluster pronunciation frequencies of good and poor readers?
3. Do boys and girls differ in their pronunciations of vowel clusters?
4. What differences are there in the vowel cluster pronunciations of second, fourth, and sixth grade subjects?
5. Do children of different community types differ in their pronunciations of vowel clusters?
6. Will children's pronunciations of vowel clusters be more closely related to the letter-sound correspondences on a type corpus or a token corpus?
7. Will word position or consonant environment affect the pronunciation of vowel clusters in synthetic words?

Background

Educators who are concerned with the improvement of reading instruction are rarely satisfied with the status quo. For years the field of reading has been characterized by the ubiquitous debate over beginning reading methodology. Proponents of such methods as language experience, whole-word, individualized instruction, synthetic phonics, linguistic, and modified alphabets have been prolific in their research, writing and speaking.

The publication of Chall's survey (1967) generated increased polemics, for her conclusions tended to dispel "conventional wisdom". Admitting that no single approach to beginning reading was all good or all bad, she nevertheless concluded that ". . . the first step in learning to read one's native language is essentially learning a printed code for the speech we possess" (p. 83). A growing number of reading specialists (Clymer, 1968; Goodman, 1964; Burns, 1965; Betts, 1964; Lamb, 1968; and others) and linguists (Fries, 1963; Bloomfield, 1961; Hall, 1961; Venezky, 1966; Weber, 1968; Weir, 1964; and others) are generally in accord with this view.

This study was not designed to compare methodologies in either the "meaning emphasis" or "decoding emphasis" philosophies, but was intended to examine in detail one aspect of the "code." American English uses more than 40 phonemes, depending on regional dialect, represented in a variety of ways by the 26 letters of the Roman alphabet. The net result is several hundred letter-sound correspondences (Bronstein, 1960). Whatever method of reading instruction is used with the beginning reader, the child must somehow develop the ability to translate the written form of English into its oral counterpart. He must sooner or later be taught--or discover for himself--the code.

Only a knowledge of the code--the relationship between some two dozen letters and 40 or more sounds--will permit readers to increase their reading fluency and vocabulary. Without this understanding and its inherent transfer, each word would have to be memorized. In fact, studies show that good "whole-word" readers have discovered and use letter-sound correspondences (Bishop, 1964).

Language has been dissected in various ways by linguists and educators. Bloomfieldian linguists refer to four levels of language: the phonemic, morphemic, syntactic, and semantic (Hockett, 1958). The transformational-generative linguists of the Chomsky school speak of competence and performance--the deep structure and surface structure, and the syntactic, lexical, semantic, and phonetic components (Chomsky, 1957), while the importance of the suprasegmental phonemes of stress, intonation, and juncture, is advanced by Trager and Smith (1957). Regardless of one's notions of the structure and component parts of the language, the graphic representation of language, the orthography, is the barrier which must be crossed in learning to read. The orthography--and its relation to sound--is the most important feature of language for anyone learning to read.

Mathews attributes the extraordinary greatness of the Greeks to the acceptance of this linguistic fact:

> Although various peoples had been writing for thousands of years before the Greeks, the latter outstripped all those who had preceded them in this field. The secret of their phenomenal advance was in the vividness of their conception of the nature of a word. They reasoned that words were sounds, or combinations of ascertainable sounds, and they held inexorably to the basic proposition that writing, properly executed, was a guide to sound. Their firm adherence to this view caused them to be dissatisfied with the failure of the Egyptians and the Semites to take full account of all the speech sounds, the vocalic ones in writing being slighted.
>
> Other peoples, such as the Babylonians and the Egyptians, had caught glimpses of the desirability of having signs represent sounds, not things, but they were never able to break with convention to the extent of setting aside picture writing in favor of letter writing. The fundamental defect of picture writing was that it was not based upon sounds at all. The Greeks saw this basic weakness and by avoiding it achieved everlasting distinction (Mathews, 1967, p. 7).

If one accepts the fact that language is oral and writing is a representation of speech, and further that the ability to read involves, either consciously or subconsciously, the translation from written symbols to sound, it must follow that accurate information about the symbol-sound relationships of English is needed. Until this decade little information of this nature had been accumulated scientifically. Spelling reformers had, perhaps, contributed the greatest quantity of literature on English orthography, but their arguments were based on the assumption that alphabets should be perfectly phonetic or phonemic, i.e., for each sound there should be a letter. Nevertheless, many interesting observations of the nature of speech came from reformers such as Hart.

Since the 16th Century, studies of symbol-sound relationship in English have been published. Abercrombie (1948) and Dobson (1959) survey most (if not all) of the earlier works. Perhaps the most complete analysis of spelling-to-sound correspondences during those early times was that of Douglas in 1740 (Holmberg, 1956). In fact, some of the more recent descriptions of letter-sound relationships are inferior to the work of Douglas.

Robert Hall published a monograph barely a decade ago, concerned with the relationship between letters and sounds (1961). His primary intent was to present solid linguistic information which might contribute to the demise of the then prevalent "lock-say" method of reading instruction. His work includes lists of English phonemes and their various graphemic representations. Hall's contention that many of the "irregularities" in English were intentionally devised some 500 years ago to keep reading and writing in the hands of the upper classes, runs contrary to language history. However, he feels the only way to teach reading effectively is to establish in the learner's mind a correlation between letters and sound (p. 60).

In 1961 Venezky (1963) developed a computer program to derive and tabulate spelling-to-sound correspondences in a corpus of 20,000 common English words. The computer analysis provided:

> A complete tabulation of the spelling-to-sound correspondences in a corpus, based upon the position of consonant and vowel clusters within the printed words. For any continuous string of vowels or consonants found in a printed word, the tabulations include all of the pronunciations found for that string, along with the totals and percentages for each pronunciation in each word-position (initial, medial and final), and complete word lists for each correspondence found . . ." (Venezky, 1967).

Weir (1964) advanced the hypothesis that if the writing system of English is viewed as a morphophonemic system, there is a much greater degree of regularity evident than if a letter-to-sound relationship were assumed. That study and later work by Weir and Venezky (1965) lent considerable support to this hypothesis. One of the results of their analyses was the description of language-dependent units on the graphemic level, called functional units, which are significant for the prediction of sound from spelling. Functional units are divided into two classes--relational units and markers. Relational units refer to a string of one or more graphemes which relate directly to a morphophonemic correspondent; for example, t —>/t/ and ph —>/f/ not /p/ + /h/. Markers are one or more graphemes whose primary function is to indicate the correspondences of relational units, or to preserve a pattern, For example, the e in rate marks a —>/e/. The major relational units include simple and compound consonants, single vowels, and vowel clusters.

Venezky's and Weir's work showed that many letter-to-sound correspondences are very predictable, while others are not. F is /f/ in all English words except of, for example, and c is /k/ before a, o and u and /s/ before e, i, and y (with certain exceptions, cello, social), but oo may be either /u/, /ʊ/ or /ə/ before d as in food, good and blood.*

Other examinations of letter-sound correspondences have been conducted by Oaks, Fry, Clymer, and Burmeister (Burmeister, 1968). The principal purposes of these studies were to test the usefulness of commonly taught phonics generalizations. These writers generally concluded that many of the phonics "rules" being taught were of little

*All phonemic symbols are from the International Phonetic Alphabet, devised by the International Phonetic Association. A listing of most phonemic symbols of American English is given in Appendix A.

value because of the numerous exceptions. These studies, though interpreted by the authors to show the irregularities of English, point out in a limited way what the computers have divulged more thoroughly--English orthography is not the highly irregular system many have thought it to be.

Hanna approached the problem of symbol-sound relationships from the other direction, that is, from sound to symbol (Hanna, et al., 1966). Since Hanna et al., were concerned with spelling rather than reading improvement, they tabulated the different spellings for a given sound, rather than the different sounds for a given spelling. They developed a 17,000 word corpus extracted from the Thorndike-Lorge, Teacher's Word Book of 30,000 words and Merriam-Webster, New Collegiate Dictionary. Their computer analysis of the corpus provided a complete analysis of sound to spelling correspondences in these English words.

To test the utility of certain phonic generalizations, Burmeister attempted to identify the most common sounds of each vowel pair through an analysis of the aformentioned Hanna study (Burmeister, 1968). However, her tables list only 26 of the many vowel clusters listed by Hanna; among the omitted clusters is io, which is the most common in English. Furthermore, some of her conclusions lack observable support. For example, her tables show ea —>/i/ 50.5% and ow —>/o/ 50%, yet she includes them in the category of vowel clusters which can be profitably taught with the "two vowels go a 'walking'" rule. Some of her terminology is confusing. "Ordinarily when two vowels appear together they should be viewed as a grapheme . . ." (p. 445), probably refers to the fact that contiguous vowels usually represent one phoneme.

In general, her investigation seemed less than rigorous.

While both linguists and educators have examined the relationship between symbol and sound in the English language, very little research has been conducted on the child's acquisition of the symbol-sound code. Descriptions of two such studies follow.

Biemiller and Levin conducted a study of the latency of oral response to words containing digraph spellings (sh, ai, ng) and "common clusters" (sl, cl). Their study was designed to examine the importance of auditory versus visual processing of stimulus words. They sampled 48 children drawn from the second, third and fourth grades, and presented them with words either preserving the intactness of the digraph (sh ed) or breaking the two letters (s hed).

> Results indicated that second and third graders took nearly one second longer to read words whose initial and final digraph were broken than they did to read words whose digraphs were presented intact. The effect did not occur for fourth grade children. There were no latency effects attributable to breaking medial, vowel digraph (Levin, et al., 1968, p. 178).

Of relevance to the current study is the conclusion that pronunciation of vowel clusters is not affected by division of the cluster. This suggests that children expect one phoneme rather than two when encountering vowel clusters.

Another investigation of children's acquisition of symbol-sound correspondences was undertaken by Calfee, Venezky and Chapman (1968) whose major concern,

> was to find the extent to which the reader used (regular) correspondences in pronouncing synthetic words, and how they pronounced synthetic words for which no such regular correspondence existed (p. IX).

They used 40 synthetic words, each typed in sans-sarif capital letters and mounted on a 35mm slide, which the subjects were to pronounce. A total of 245 students from the third, sixth, eleventh and twelfth grades, and college, participated in the study. Their general conclusion was that good readers consistently gave more appropriate responses to predictable letter-sound correspondence patterns than poor readers, though no group--even the oldest and the best readers--gave appropriate responses all the time.

Some of the synthetic words contained vowel clusters which have, as noted earlier, unpredictable letter-sound correspondences. Calfee, _et al_., found,

> With the exception of _ea_ and _ee_, overall agreement on a preferred pronunciation for a vowel digraph was not high; neither was there high agreement on specific items. Shifts in pronunciation of a given digraph as a result of context were observed, however, suggesting that choice of pronunciation may be contextually bound. Whatever the moderating mechanism, the spread of observed pronunciations for most digraph spellings suggests that it tends to be idiosyncratic (p. 167).

No other research regarding children's pronunciations of vowel clusters has been undertaken, to the knowledge of this investigator.

While developing reading ability, children's generalizations result more often from example than from rule. Therefore, more information is needed about _how_ children generalize from language input data, and the first step is to know what the input is. Since reading involves the ability to translate written symbols to sound, and since the letter-sound correspondences of vowel clusters are generally not predictable, awareness of the frequencies of the several pronunciations of each vowel cluster spelling in common English words is required.

Nearly one-third of the most common 20,000 words, and one-fifth of the 1,000 most frequent words in English contain vowel cluster spellings. Unless each of these is to be learned as a sight word, which is not practicable, a listing of all vowel clusters in English, and the phonemes and phonemic strings they represent, is needed. This information is an essential basis for an investigation of children's generalization behavior with vowel cluster spellings.

What do young readers do when they encounter familiar vowel clusters in unfamiliar words? When they come upon an unfamiliar word with an ea cluster, do they attempt /i/ as in teach, or /ɛ/ as in dead, or /e/ as in great? Accurate and all-encompassing generalizations about compound vowel pronunciations cannot be taught, as they can with many other letter-sound correspondences--(t usually is /t/, c is usually /k/ before a, o, and u, etc.). Information is needed about how young readers pronounce unfamiliar words containing vowel clusters. Flexibility has long been a goal of reading instruction. Are children flexible readers? For example, when asked to pronounce unfamiliar words containing ow, will a child always give the /o/ pronunciation as in grow, or the /au/ sound as in now, or will he vary his pronunciations? Research is needed which will relate the pronunciation preferences of children to actual characteristics of the language.

Rationale for the Investigation

Vowel clusters are perhaps the most complex and unpredictable components of the letter-sound correspondence code. Vowel cluster spellings differ from single vowel spellings in several ways. They rarely appear before geminate consonant clusters; some, such as ai and au occur infrequently in word-final position. Others, oa, ie, etc., rarely begin a word in English.

Some vowel clusters have a major phonemic correspondent, and possibly several minor correspondents. The major correspondent of ai is /e/ as in bait, and it represents this sound in 85% of its occurrences. It represents /ə/, villain; /ai/, aisle, /ɛ/, again; /æ/, plaid; and others much less frequently. Other vowel clusters have two or more major correspondents, as well as minor correspondents. For example, ow is /o/ as in own 51% of the time and /au/ as in owl 48%. Its only minor correspondent is /a/ as in knowledge. On the other hand, all single vowel spellings have two major correspondents, (e.g., a —>/e/ or /æ/) plus several minor correspondences.

While single-vowel spellings can be traced to the earliest English writing, most vowel cluster spellings are much more recent, having been introduced during the late Middle English period. Consequently, vowel cluster correspondences underwent considerably fewer sound changes than did single-vowel spelling, though they did undergo some change (Mosse, 1952). For example, the Middle English diphthong /au/, spelled au or aw, developed in a complex manner (including French borrowings). With the Great Vowel Shift, Middle English /au/ became /ɔ/, though the au-aw spelling was retained. The vowel cluster oo first appeared in the 14th Century to represent /o:/, but did not become established until the 16th Century. /o:/ changed to /u:/ through the Great Vowel Shift, and to /u/ in Modern English. In some oo spelled words, /u:/ shortened to /ʊ/ as in book, and was later unrounded to /ə/ as in flood (Venezky, 1963).

The goals of the present study are to examine vowel cluster-sound relationships. Given that certain factors of pronunciation exist, (that is, in common English words many vowel clusters have six or more pronunciations), what factors influence a child's pronunciations of vowel-cluster words? Are good readers' pronunciations more often in the range of theoretical possibility than poor readers? What differences arise as children progress through the elementary grades? Is pronunciation related to community environment or sex? Does consonant environment affect pronunciation choice?

The present study seeks to answer questions about the relationships between the pronunciations of vowel clusters by a representative sample of elementary school children, and vowel cluster--phoneme correspondences in a large corpus of common English words. Such information should provide a source for the modification of beginning reading materials and methods. For example, au is /ɔ/ as in cause in 90% of common English words but is /æ/ as in laugh in only two words and their derivations. Since laugh and aunt are frequently taught as "sight" words in early reading, it might be expected that children develop a false generalization about au which they apply to unfamiliar words they encounter--even though letters and sounds are not stressed with "sight" words. The present study will reveal the extent to which this happens with the different types of children studied.

Definition of Terms

Several terms are important to an understanding of this study. They are defined as follows:

Vowel Cluster: a group of letters composed of two or more contiguous vowel graphemes. It is used synonomously with compound vowel, vowel pair, or vowel digraph. There are 61 different vowel clusters in the corpus of 20,000 common English words used by Venezky, 1963; some occur in only one word and one occurs in more than 1,000 words. (aie occurs in one word--gaiety, while io occurs in 1,293 words--action, ratio, lion, etc.)

Letter-Sound Correspondence: a grapheme-phoneme relationship. Many letters, like d, f, l and z have invariant or nearly invariant pronunciations (d—>/d/, f—>/f/ except in of where f—>/v/) and are, therefore, predictable. Some letters, like b, c, and m have variant letter-sound correspondences which can be predicted on the basis of grapheme environment (c—>/s/ before e, i, or y--cell, city, cyst, otherwise c—>/k/). Each single letter vowel, (a-e-i-o-u-y), has two major correspondences (long and short as in cape and cap), which are usually predictable, and other less predictable pronunciations. Vowel clusters generally have several letter-sound correspondences (ou—>/au/, /ʊ/, and /u/ as in found, would and you) which, in most instances, are not predictable.

20,000 Word Corpus: a 1963 modification of the Thorndike list of 20,000 most frequent English words (Venezky, 1963). The original Thorndike list was revised by Venezky primarily through deletion of archaic words and addition of new words. The revised list was programmed for a computer analysis of the letter-sound correspondence therein. The output included a complete tabulation of spelling-to-sound correspondences, along with word lists for each correspondence. In addition, a sound-to-spelling correspondence listing, a reversed spelling listing, and a reversed pronunciation listing was obtained. An analysis of the corpus by this investigator revealed 61 different vowel clusters, representing 92 different phonemes and phoneme strings for a total of more than 300 letter-sound correspondences in over 6,000 words.

1,000 Word Corpus: the 1,003 most frequent English words derived from a corpus of 1,000,000 running words by Kučera and Francis (1967). This corpus is the most recent and certainly the most exhaustive computer tabulation of word frequencies to date. The 1,000 word corpus used in this study contains the 1,003 most frequently written American English words, ranging from the most frequent, the, which occurs 69,971 times per million running words, to the 1,003rd most frequent (11 words, each occurring 106 times per 1,000,000 running words--applied, reach, etc.).

Type: a "distinct word," viewed as one word regardless of how frequently the word appears. (the and applied are considered distinct words regardless of their frequencies.)

Token: an "individual word" considered for this study in terms of frequency of its appearance. In this study pronunciation frequencies of words in the 1,000 word corpus are based on tokens, while pronunciations in the 20,000 word corpus are based on types. The following example is offered for clarification. In the 1,000 word corpus there are five words which contain the au spelling. Of these five, four have the /ɔ/ pronunciation (because, etc.), while one has the /æ/ pronunciation (laugh). With a token description--based on number of occurrences of each word in 1,000,000 running words, /ɔ/ is equal to 91.34% and /æ/ 8.57%. By comparison, if the pronunciation frequencies were based on types, /ɔ/ would equal 80% and /æ/ 20%.

Reading Ability: performance on a standardized reading test. All subjects in this study received a grade-level equivalent reading score, on such standardized reading tests as the Metropolitan Achievement or the Iowa Test of Basic Skills (see Appendices D, E, and F).

Intelligence: performance on a standardized group intelligence test (see Appendices D, E, and F).

Distractor: a multiple choice response item.

Chapter II

SELECTION OF VOWEL CLUSTERS

This study was designed to achieve two broad objectives: (1) to investigate elementary pupils' pronunciations of vowel clusters in unfamiliar words, and (2) to analyze the observed pronunciations in relation to existing letter-sound correspondences of vowel clusters in common English words. This chapter contains a description of the analysis of the vowel clusters in 20,000 common English words, and the procedures followed in the selection of representative vowel clusters used in this investigation. The symbol-sound correspondences of the selected vowel clusters as they occur in the 1,000 most frequent English words are presented also.

Analysis of Vowel Cluster Letter-Sound Correspondences

As part of an inter-disciplinary study of the reading process begun at Cornell University in 1961, Venezky developed a computer program to derive and tabulate letter-sound correspondences in a corpus of 20,000 common English words (Venezky, 1963). The 20,000 word corpus was a modification of the most common 20,000 words according to the Thorndike frequency count (Thorndike, 1941). Venezky omitted many archaic and low frequency words, particularly proper nouns, and added a number of words in their place. Along with other information, the computer analysis provided an inclusive tabulation of letter-sound correspondences found in the corpus as well as totals and percentages for each pronunciation in each word position, and a complete word list for each correspondence. A Pronouncing Dictionary of American English (Kenyon & Knott, 1953) was used to determine the pronunciation of most words in the corpus.

The principal purpose of this analysis and later research by Venezky and Weir was, ". . . to construct a theoretical framework for deriving sound from spelling and to search for the most plausible linkages for fitting these relationships into the total language structure" (Venezky, 1967, p. 80). Later work by Venezky and others was concerned with whether or not readers use these theoretical patterns of symbol-sound relationships when reading.

Venezky's unpublished computer print-out of spelling-to-sound correspondences in 20,000 English words was made available to this investigator during the academic year 1968-1969. An analysis of the vowel cluster letter-sound correspondences in this print-out disclosed the following:

1. There were 61 vowel clusters, including those containing the semi-vowels w and y in the corpus.

2. The 61 vowel clusters represented 92 different single vowel phonemes and phoneme strings, producing more than 300 symbol-sound correspondences.

3. The 61 vowel clusters appeared 6,272 times in the 20,000 word corpus.

4. There was great variance in the frequency of occurrence of the 61 vowel clusters in the 20,000 word corpus. As shown in Table 2:01, one occurred in more than 1,000 words while 25 occurred in three words or less.

Table 2:01

Frequency of Vowel Clusters in 20,000 Word Corpus

Number of Clusters	Number of Words
1	over 1000
2	500 - 999
14	100 - 499
9	50 - 99
4	10 - 49
6	4 - 9
25	1 - 3

5. Vowel clusters varied greatly in the number of individual phonemes or phoneme strings they represented. Table 2:02 indicates that some represented only one sound while one represented 17 sounds.

Table 2:02

Frequency of Occurrences of Vowel Clusters and the Numbers of Sounds they Represent

Vowel Cluster	Number of Words in Which it Occurs	Number of Sounds it Represents
io	1293	10
ea	599	17
ia	581	15
ou	475	11
ee	319	6

Table 2:02 (cont.)

Vowel Cluster	Number of Words in Which it Occurs	Number of Sounds it Represents
oo	312	7
ai	303	9
ie	274	15
ow	256	3
au	191	6
ay	159	8
iou	139	5
oi	130	7
oa	125	7
ue	108	16
ua	104	13
ui	102	8
ei	94	8
ey	92	5
aw	88	3
ew	82	3
eo	75	13
iu	56	4
oy	56	2
oe	52	10
eu	51	8
eou	33	2
uou	27	3

Table 2:02 (cont.)

Vowel Cluster	Number of Words in Which it Occurs	Number of Sounds it Represents
ae	21	7
eau	14	3
ao	6	3
ieu	5	2
iew	5	1
oui	5	3
aeo	4	4
uo	4	3
uy	3	1
uoy	3	1
aa	2	1
oia	2	1
uay	2	1
eea	1	1
aea	1	1
eia	1	1
iaow	1	1
ii	1	1
oau	1	1
eow	1	1
ioa	1	1
uia	1	1
eoi	1	1

Table 2:02 (cont.)

Vowel Cluster	Number of Words in Which it Occurs	Number of Sounds it Represents
eei	1	1
oeu	1	1
oie	1	1
oua	1	1
eue	1	1
aiia	1	1
aii	1	1
aie	1	1
oue	1	1
uu	1	1

6. Most vowel cluster pronunciations are unpredictable; their sounds cannot be predicted from their spellings.*

7. Of the 61 vowel clusters, 30 occurred in 10 or more words in Venezky's modification of Thorndike's list of 20,000 common words. Of the 30, 23 occurred in words in which the vowel cluster is sometimes disyllabic. Only six of these vowel clusters were disyllabic more often than monosyllabic. Thus, the 30 vowel clusters, occurring in more than 6,000 words, represented single vowel phonemes about 80% of the time and two or more phonemes about 20%. This is shown in Table 2:03.

*A symbol-sound correspondence is considered predictable if it can usually be determined within a consonant environment. For example, g is usually /g/ before a, o, and u, as in game, go, and gum. On the other hand, ea may be either /i/, /ɛ/, or /e/ before /t/, as in heat, threat, and great. Therefore, ea is considered unpredictable.

Table 2:03

Monosyllabic and Disyllabic Status of the

30 Most Common Vowel Clusters

	One Syllable		Disyllabic	
Vowel Cluster	Number of Words	Per Cent	Number of Words	Per Cent
ae	18	85.7%	3	14.3%
ai	298	98.3%	5	1.7%
au	191	100.0%	0	0.0%
aw	86	87.7%	2	2.3%
ay	158	99.4%	1	0.6%
ea	486	81.1%	113	18.9%
ee	310	97.2%	9	2.8%
ei	68	72.3%	26	27.7%
eo	19	25.3%	56	74.6%
eou	8	24.2%	25	75.8%
eu	37	72.5%	14	27.5%
ow	82	100.0%	0	0.0%
ey	92	100.0%	0	0.0%
ia	150	25.8%	431	74.2%
ie	184	67.1%	90	32.9%
io	1,141	88.2%	152	11.8%
iou	79	56.8%	60	43.2%
iu	4	7.1%	52	92.9%
oa	104	83.2%	21	16.8%

Table 2:03 (cont.)

	One Syllable		Disyllabic	
Vowel Cluster	Number of Words	Per Cent	Number of Words	Per Cent
oe	30	57.7%	22	42.3%
oi	108	83.1%	22	16.9%
oo	305	97.8%	7	2.2%
ou	475	100.0%	0	0.0%
ow	256	100.0%	0	0.0%
oy	56	100.0%	0	0.0%
ua	1	1.0%	103	99.0%
ue	76	70.3%	32	29.7%
ui	68	62.7%	34	37.3%
uou	0	0.0%	27	100.0%
TOTAL	4,904		1,307	

Perhaps the best way to exemplify the variety of possible pronunciations of the vowel clusters is to list the most common clusters and their most common pronunciations. The following tables, 2:04 through 2:20, list each of the 17 vowel clusters which occurred in more than 100 words in the corpus. For each cluster the four most common pronunciations are included.

Table 2:04

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster <u>ai</u>

Phoneme	Number of Words	Percentage	Example
/e/	260	85.8%	bait
/ə/	20	6.6%	villain
/ɪ/	6	2.0%	captain
/ai/	5	1.7%	aisle
5 others	12	3.9%	plaid

Total Occurrences - 303 words

Table 2:05

Frequency of Occurrence of the 4 Most Common

Pronuncuations of the Vowel Cluster <u>au</u>

Phoneme	Number of Words	Percentage	Example
/ɔ/	175	91.6%	cause
/o/	6	3.1%	chauffeur
/æ/	5	2.6%	laugh
/au/	3	1.6%	kraut
2 others	2	1.1%	gauge

Total Occurrences - 191 words

Table 2:06

Frequency of Occurrence of the 4 Most Common Pronunciations of the Vowel Cluster ay

Phoneme	Number of Words	Percentage	Example
/e/	142	89.3%	day
/ɪ/	10	6.3%	always
/ai/	2	1.3%	aye
/ɛ/	1	0.7%	says
4 others	4	2.5%	picayune

Total Occurrence - 159 words

Table 2:07

Frequency of Occurrence of the 4 Most Common Pronunciations of the Vowel Cluster ea

Phoneme	Number of Words	Percentage	Example
/i/	318	53.1%	each
/ɛ/	135	22.6%	breakfast
/iə/	45	7.5%	cereal
/ɪə/	24	4.0%	area
13 others	77	12.8%	ocean, great

Total Occurrence - 599 words

Table 2:08

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster ee

Phoneme	Number of Words	Percentage	Example
/i/	293	91.8%	bleed
/ɪ/	12	3.8%	been
/iɛ/	8	2.5%	preempt
/e/	3	1.0%	matinee
2 Others	3	0.9%	reelection
Total Occurrence - 319 words			

Table 2:09

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster ia

Phoneme	Number of Words	Percentage	Example
/ɪə/	238	41.0%	alias
/ə/	124	21.3%	special
/ɪe/	77	13.2%	humiliate
/aɪə/	56	9.7%	giant
11 Others	86	14.8%	piano
Total Occurrence - 581 words			

Table 2:10

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster ie

Phoneme	Number of Words	Percentage	Example
/i/	73	26.7%	movie
/ɪ/	42	15.3%	sieve
/ɪə/	33	12.0%	audience
/aiə/	27	9.9%	diet
11 Others	99	35.1%	friend, lie
Total Occurrence - 274 words			

Table 2:11

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster io

Phoneme	Number of Words	Percentage	Example
/ə/	1,138	88.0%	action
/ɪə/	33	2.5%	idiot
/jə/	30	2.3%	onion
/aiə/	29	2.2%	lion
8 Others	63	5.5%	trio
Total Occurrence - 1,293 words			

Table 2:12

Frequency of Occurrence of the 5 Most Common

Pronunciations of the Vowel Cluster iou

Phoneme	Number of Words	Percentage	Example
/ə/	75	54.0%	delicious
/ɪə/	59	42.4%	furious
/jə/	3	2.2%	rebellious
/u/	1	0.7%	Sioux
/aijə/	1	0.7%	pious
Total Occurrence - 139 words			

Table 2:13

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster oa

Phoneme	Number of Words	Percentage	Example
/o/	94	75.2%	oat
/oə/	13	10.4%	coalition
/ɔ/	9	7.2%	broad
/oæ/	6	4.8%	coagulate
3 Others	3	2.4%	oasis
Total Occurrence - 125 words			

Table 2:14

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster oi

Phoneme	Number of Words	Percentage	Example
/ɔi/	104	80.0%	coin
/oɪ/	18	13.8%	coincide
/ə/	3	2.3%	porpoise
/uɪ/	2	1.5%	doing
3 Others	3	2.4%	chamois

Total Occurrence - 130 words

Table 2:15

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster oo

Phoneme	Number of Words	Percentage	Example
/u/	194	62.2%	boot
/ʊ/	84	26.9%	foot
/ə/	23	7.4%	flood
/oa/	6	1.9%	zoology
3 Others	5	1.6%	brooch

Total Occurrence - 312 words

Table 2:16

Frequency of Occurrence of the 4 Most Common Pronunciations of the Vowel Cluster <u>ou</u>

Phoneme	Number of Words	Percentage	Example
/au/	238	50.1%	ounce
/ə/	181	38.2%	touch
/u/	30	6.3%	soup
/o/	13	2.7%	soul
6 Others	13	2.7%	should
Total Occurrence - 475 words			

Table 2:17

Frequency of Occurrence of the 3 Most Common Pronunciations of the Vowel Cluster <u>ow</u>

Phoneme	Number of Words	Percentage	Example
/o/	131	51.2%	own
/au/	122	47.7%	cow
/a/	3	1.1%	knowledge
Total Occurrence - 256 words			

Table 2:18

Frequency of Occurrence of the 4 Most Common Pronunciations of the Vowel Cluster ua

Phoneme	Number of Words	Percentage	Example
/uə/	44	42.3%	actual
/ue/	14	13.5%	fluctuate
/jue/	13	12.5%	evacuate
/juə/	11	10.6%	annual
9 Others	22	21.1%	language
Total Occurrence - 104 words			

Table 2:19

Frequency of Occurrence of the 4 Most Common Pronunciations of the Vowel Cluster ue

Phoneme	Number of Words	Percentage	Example
/u/	25	23.1%	blue
/ju/	24	22.2%	value
/#/	23	21.3%	tongue
/uə/	14	13.0%	cruel
12 Others	22	20.4%	guess
Total Occurrence - 108 words			

Table 2:20

Frequency of Occurrence of the 4 Most Common

Pronunciations of the Vowel Cluster ui

Phoneme	Number of Words	Percentage	Example
/ju/	21	20.6%	ambiguity
/u/	19	18.6%	fruit
/ɪ/	18	17.6%	build
/wɪ/	18	17.6%	penguin
4 Others	26	25.6%	ruin
Total Occurrence - 102 words			

Selection of Vowel Clusters for Study

Rather than study all 61 vowel clusters, it was decided that a representative subset of the total array of vowel clusters would permit sufficient analysis of children's vowel cluster pronunciation behavior. The two principal criteria used for selection of the appropriate vowel clusters to include were frequency of occurrence and phonemic representation.

To begin, all vowel clusters occurring in fewer than 100 words were eliminated; these totaled 44. The remaining 17 were analyzed to determine the range of their sound correspondences. To test children's pronunciations of the spectrum of vowel clusters it was deemed necessary to include: (1) some clusters which have one principal pronunciation, such as ai—>/e/ (gain), oa—>/o/ (boat), and au—>/ɔ/ (pause); (2) clusters which have two principal pronunciations such as oo—>/u/ (food) or /ʊ/ (good), and ow—>/o/ (grow) and /au/ (plow), and ou->/au/ (proud) and /ə/ (famous); and (3) clusters with more than two main pronunciations such as ie->/i/ (movie), /I/ (sieve) and /ai/ (die). The cluster ay was included because of its alternation with ai in word position, and ea was included because of its frequency. Though io was the most frequent vowel cluster, it was omitted because nearly 90% of the time it occurs in /ʃən/ syllables as in nation and passion.

Based upon the preceding criteria, these nine vowel clusters, ai, au, ay, ea, ie, oa, oo, ou, and ow, appeared to comprise a representative cross-section of all vowel clusters. Further, they accounted for nearly half of all the occurrences of all 61 vowel clusters in the modified Thorndike 20,000 word corpus. By testing each of them in a variety of environments, a manageable instrument could be constructed.

Letter-Sound Correspondences of the Nine Selected Vowel Clusters in the 1000 Most Frequent English Words

The pronunciation frequencies of the modified Thorndike 20,000 word corpus discussed and tabled previously, were based on word types. That is, each word received the same weight and was counted only once regardless of its frequency in the sample of written words from which the corpus was selected. Common vowel cluster words such as would, could, and should affected the pronunciation proportions no more than such rarely used words such as brooch and ooze.

To provide another basis for the analysis of children's pronunciations of vowel clusters in relation to actual pronunciation frequencies, an analysis of token word frequencies was required. In 1967 Kučera and Francis published an exhaustive computational tabulation of English words. The corpus consisted of 1,014,232 words of natural-language text in 15 different genre, and included 50,406 distinct words (types). Their analysis ranked these fifty thousand words on the basis of their frequencies in the total sample. For example, <u>the</u> was the most frequent word, occurring 69,971 times while <u>accordian</u> was one of the most infrequent, occurring only once.

One possible influence on children's pronunciations could be a large number of words with the same vowel cluster letter-sound correspondence, and another influence could be highly frequent words with a different correspondence. Assuming a child knows five words with <u>ai</u> spellings, <u>maid</u>, <u>lain</u>, <u>paid</u>, <u>pain</u>, and <u>said</u>, would his pronunciation of <u>ai</u> in an unfamiliar word be more greatly influenced by the first four relatively infrequent words, or by the highly frequent word <u>said</u>? For example, <u>ou</u> is /au/ (ounce) in 50% of the words in which it occurs and is /ʊ/ (could) in only 1%. Yet the /ʊ/ pronunciation occurs in three highly frequent words, <u>would</u>, <u>could</u>, and <u>should</u>. Would children's pronunciations of vowel clusters in unfamiliar words be more closely related to the vowel cluster pronunciation proportions on the type corpus or the token corpus?*

---

*In subsequent analyses of the data, the pronunciation proportions of both the type corpus and the token corpus were used in relation to the pronunciation proportions of the subjects. All hypotheses tested in this investigation are based on either type or token vowel cluster pronunciation proportions.

This investigation analyzed the 1000 most frequent words in the Kučera-Francis Corpus to determine the frequency of pronunciation of the nine vowel clusters based on tokens. Pronunciations were derived from Kenyon and Knott's A Pronouncing Dictionary of American English. Words in this subset of 1000 words occurred from 106 to 69,971 times per million running words. It was found that approximately 20% of the words in this (token) corpus contained vowel clusters, compared with a third of the words in the 20,000 word (type) corpus.

Table 2:21 compares the pronunciation proportions of the type and token corpora for the vowel clusters selected for this study. Further, it shows that for some vowel clusters (e.g., ow, au) there was little difference between type and token pronunciation frequencies, while for others (e.g., ou, ie) the differences were considerable. These differences were an important aspect of this study. For each vowel cluster, the principal and secondary pronunciation proportions on each corpus was determined. For example, on the type corpus the principal pronunciation of ai was /e/ at .86 and the secondary was /ε/ at .07. On the token corpus the principal pronunciation of ai was /ε/ at .38 and the secondary was /e/ at .27. In the analyses reported in Chapters 3 and 4, subjects' pronunciation proportions were related to the proportions on each corpus.

Table 2:22 presents the words' position percentages for each of the selected vowel clusters. These percentages became the basis of word positions of the vowel clusters in the synthetic words used in the study. Construction of these synthetic words is discussed in Chapter 3.

Table 2:21

A Comparison of Vowel Cluster Pronunciation Proportions

in the 20,000 Word and 1,000 Word Corpora

Vowel Cluster	Phoneme	Type Corpus (20,000)		Token Corpus (1,000)		Example
		Words	Per Cent	Words	Per Cent	
ai	/e/	260	85.8%	10	27.4%	bait
	/ə/	20	6.6%	2	9.1%	villain
	/ɪ/	6	2.0%	2	26.1%	captain
	/ɛ/	3	1.0%	1	38.3%	said
	others	10	4.6%	0	0.0%	
ay	/e/	142	89.3%	14	89.5%	days
	/ɪ/	10	6.3%	1	7.3%	always
	/ai/	2	1.3%	0	0.0%	aye
	/ɛ/	1	0.7%	1	3.2%	says
	others	4	2.5%	0	0.0%	

Table 2:21, cont.

Vowel Cluster	Phoneme	Type Corpus (20,000)		Token Corpus (1,000)		Example
		Words	Per Cent	Words	Per Cent	
au	/ɔ/	175	91.6%	4	91.4%	cause
	/o/	6	3.1%	0	0.0%	chauffer
	/æ/	5	2.6%	1	8.6%	laugh
	/au/	3	1.6%	0	0.0%	sauerkraut
	others	2	1.1%	0	0.0%	
ea	/i/	318	53.1%	22	57.4%	each
	/ɛ/	135	22.6%	9	22.8%	bread
	/iə/	45	7.5%	3	8.1%	cereal
	/e/	12	2.0%	2	10.2%	great
	others	89	14.8%	1	1.5%	
ie	/i/	73	26.7%	8	46.6%	movie
	/ai/	25	9.1%	2	10.7%	lie
	/ɛ/	5	1.8%	2	11.4%	friend

Table 2:21, cont.

Vowel Cluster	Phoneme	Type Corpus (20,000)		Token Corpus (1,000)		Example
		Words	Per Cent	Words	Per Cent	
(cont.)						
ie	/aiə/	4	1.5%	2	14.3%	diet
	others	166	60.9%	3	16.7%	
oa	/o/	94	75.2%	2	100.0%	boat
	/oə/	13	10.4%	0	0.0%	coalition
	/ɔ/	9	7.2%	0	0.0%	broad
	/oae/	6	4.8%	0	0.0%	coagulate
	others	3	2.4%	0	0.0%	
oo	/u/	194	62.2%	8	47.8%	boot
	/ʊ/	84	26.9%	7	50.0%	foot
	/ə/	23	7.4%	1	2.2%	flood
	/oa/	6	1.9%	0	0.0%	zoology
	others	5	1.6%	0	0.0%	

Table 2:21, cont.

Vowel Cluster	Phoneme	Type Corpus (20,000)		Token Corpus (1,000)		Example
		Words	Per Cent	Words	Per Cent	
ou	/au/	238	50.1%	15	36.4%	ounce
	/ə/	181	38.2%	7	7.9%	touch
	/u/	30	6.3%	4	22.5%	soup
	/ʊ/	6	1.3%	5	25.9%	should
	others	20	4.1%	4	7.2%	
ow	/o/	131	51.2%	15	46.7%	own
	/au/	122	47.7%	6	51.4%	cow
	/a/	3	1.1%	1	1.9%	knowledge
	others	0	0.0%	1	0.0%	

Table 2:22

Word Positions of the 9 Selected Vowel Clusters in the 20,000 Word Corpus

Vowel Cluster	Number of Words	Initial Position		Medial Position		Final Position	
		Words	Per Cent	Words	Per Cent	Words	Per Cent
ai	303	7	2.3%	294	97.0%	2	0.7%
ay	159	1	0.6%	48	30.2%	110	69.2%
au	191	53	27.7%	136	71.2%	2	1.0%
ea	599	16	2.7%	559	93.3%	24	4.0%
ie	274	0	0.0%	247	88.0%	33	12.0%
oa	125	7	6.0%	112	90.0%	6	4.0%
oo	312	2	0.6%	294	94.2%	16	5.1%
ou	475	37	7.8%	433	91.2%	5	1.0%
ow	256	4	1.6%	161	62.9%	91	35.6%

Chapter III

DEVELOPMENT OF THE INSTRUMENT AND PROCEDURES OF THE STUDY

The two broad purposes of this study were: (1) to investigate the pronunciation of vowel clusters in synthetic words by elementary school children, and (2) to analyze the observed pronunciations in relation to existing letter-sound correspondences of vowel clusters in common English words.

This chapter deals with the development of experimental oral and multiple-choice instruments, the two pilot studies (A, which was concerned with testing procedures, and B, which was used to refine the instrument), the final instrument, the procedures of the study, and the design and statistical analyses of the study.

Development of the Experimental Oral and Multiple-Choice Instruments

Test Items

To adequately measure children's pronunciations of vowel clusters, it was essential that real words not be used. Had real words been used it was likely that most subjects would have been familiar with some of them, and consequently the results would have been clouded. The dependent variable, based on pronunciation of familiar vowel clusters in unfamiliar contexts, could be assessed accurately only by constructing synthetic words containing the nine vowel clusters. It was determined that each vowel cluster should be tested in ten different synthetic words to enable any pronunciation patterns to emerge.

The principal guideline followed in the construction of the synthetic words was linguistic plausibility. To assure content validity it was essential that the synthetic words resemble real words in both appearance and sound. For example, many consonant clusters appear only in initial word positions in modern English spelling, dr, fl, fr, gl, gr, sm, etc., while others occur only in final positions; ck, nt, ll, etc. To be plausible, synthetic words had to be constructed upon the patterns of English spelling.

The word positions of the vowel clusters included were controlled to reflect their position frequencies in the 20,000 word corpus. These positions were ascertained from the analysis presented in Table 2:22, found on page 42.

In the construction of the synthetic words, the choice of preceding and following consonants was based on further examination of the 20,000 word corpus. For example, since ee is never followed by g nor is ie preceded by c in English, such sequences were avoided.

The first draft of the synthetic word list, containing 10 synthetic words for each of the 9 vowel clusters was submitted to a linguist, a psycholinguist, a reading specialist and a psychologist as a further check on content validity. As a result of their evaluation, several items were deleted because of their high similarity to real words in either appearance or sound, and additional synthetic words were added.

In addition to the 90 items made up of synthetic words containing vowel clusters (10 each of the 9 vowel clusters: ai, ay, au, ea, ie, oa, oo, ou and ow) 10 check items were included to determine reliability. Five of these items were real words and five were synthetic words with predictable letter-sound correspondences (e.g., pid p—>/p/). By including 10 check items, the accuracy of subjects' performance on the instrument could be established. If a subject's responses were unreliable, that is, if he simply guessed or checked responses randomly, he could be expected to miss many of the check items. The reason for demonstrating the reliability of the instrument in this fashion was drawn from the work of Kerlinger (1957, pp. 429-430).

Kerlinger defines reliability as the accuracy or precision of a measuring instrument, and he advances several synonyms for reliability: accuracy, consistency, dependability and predictability. Of his approaches to reliability, one seemed most suitable for this instrument: "Are the measures obtained from a measuring instrument the "true" measure of the property measured?". Implicit in this question is the notion of accuracy. Each of the vowel clusters included in this study has several phonemic correspondents in common English words. For example, the vowel cluster ea is /i/ in bead, /ε/ in bread, /e/ in break, /ə/ in ocean. Likewise, it is /i/ in read, and lead, and /ε/ in read and lead. Because there were no right or wrong answers to the 90 vowel cluster items, other means of determining reliability were deemed less appropriate than assessing accuracy of response through the 10 check items.

Other suggestions by Kerlinger for the improvement of reliability of the instrument were incorporated (442-443). The items were unambiguous; each item was simply a synthetic word. Care was taken to assure that the written instructions were clear (see Appendix B). In addition, the instructions were given orally by the investigator.

Nature of the Instruments

Using the items described, two experimental tests were developed, an oral pronunciation test, and a written multiple-choice test containing the 100 test items and real word response items similar in sound to the pronunciation of the vowel cluster being tested. Both tests contained the same 100 items. The 100 items were divided into two halves (labeled A and B), each half was composed of five synthetic words for each of the nine vowel clusters, and five check items. Using a table of random numbers, each 50 items subtest was arranged in two orderings. The four orderings were designated A1, A2, B1, and B2. On the four oral subtests, each item was printed on a flash card; on the four multiple-choice subtests, the test items and response choices were duplicated on two pages (see Appendix B).

Three real words were offered as multiple-choice distractors for each synthetic word used as a stimulus. The three (distractors) contained at least two of the most frequent pronunciations of the vowel cluster in the modified Throndike 20,000 word corpus. Furthermore, the distractors were selected from Clarence R. Stone's Revision of the Dale List of 769 Easy Words (Spache, 1960), words which, purportedly, most children can read by the end of the second grade. In no case were the vowel sounds in the real words spelled the same as the vowel cluster in the synthetic word being tested. To control for order effects, the distractors for each vowel cluster were randomly assigned to each subtest ordering. As an example, Table 3:01 presents two synthetic words used to test the vowel cluster ea, and shows their test form, item number and response sequence.

Table 3:01

An Example of Test Form, Item Number and Response Sequence of Two Synthetic Words

Test Form	Item Number	Synthetic Word	Response Sequence		
A1	3	polead	be	bed	baby
A2	50	polead	be	baby	bed
B1	28	deach	bed	be	baby
B2	15	deach	baby	bed	be

In summary, there were 100 test items of which 90 were synthetic words containing 10 each of the 9 vowel clusters, 5 test items were real words, and 5 were synthetic words with predictable letter-sound correspondences. The five real words and the five predictable synthetic words were included as reliability control items.

Pilot Studies

Two pilot studies were conducted and were designated Pilot Study A and Pilot Study B. Both pilot studies were done at Waterloo Elementary School, Waterloo, Wisconsin. The essential purpose of Pilot Study A was to refine the testing procedures. Pilot Study B was designed to secure information which would contribute to the final testing instruments used in the study.

Testing Procedures for Pilot Studies A and B

During Pilot Study A and Pilot Study B, both oral and multiple-choice forms of the test were used. On the multiple-choice test, each synthetic stimulus word was followed by three distractors. All three distractors for each synthetic word contained phonemes represented by that vowel cluster in the 20,000 word corpus (see Appendix B). The multiple-choice test was designed to be administered either individually or to groups. The pupils' task was to circle a real word from a choice of three whose underlined letters were, he felt, closest in sound to the underlined letters in the synthetic stimulus word.

In addition, each synthetic word was typed on a 5" x 7" flash card using primary type, lower case letters. The flash cards were arranged in sequences identical to tests A1, A2, B1, and B2, and were designed for oral pronunciation use in Pilot Study A and Pilot Study B. The oral pronunciation test was an individual test. Each subject viewed each synthetic word on a flash card and pronounced into a tape recorder. Later, phonemic transcriptions of the tape recording were made.

Pilot Study A

Pilot Study A was conducted to refine the testing procedures. The pilot sample consisted of three second, three fourth, and two sixth grade pupils at Waterloo Elementary School, Waterloo, Wisconsin. Four of the subjects were girls, one second grader, two fourth graders, and one sixth grader; four were boys, two second graders, one fourth grader and one sixth grader. On the basis of the Gates McGinty Primary

A reading test, two second grade readers of low reading ability and one of high ability were randomly selected. Scores on the Nelson Reading Test, Form A, were used to randomly select two fourth graders and one sixth grader of low reading ability and one fourth and one sixth grade pupil of high reading ability. All subjects were randomly selected from the two halves of each class based on achievement test median splits. Pupils were given one oral and one multiple-choice form of the test next. Thus, they responded to each of the 100 items twice, once in oral form and once in multiple-choice form. Pilot Study A indicated that the subjects could comprehend the instructions and perform the tasks satisfactorily. Thus, no significant changes in the testing instrument or procedures were made.

Pilot Study B

Pilot Study B was designed to examine the relationship between subjects' oral pronunciations of synthetic words containing vowel clusters, and their multiple-choice response to the same synthetic words. The reason for determining this relationship was to gain information that would contribute to the construction of the final multiple-choice instrument to be used in the study. For example, if there were no differences in the subjects' oral and multiple-choice test performance, it could be assumed that the multiple-choice test was adequate in its present form. If there were differences, on the other hand, the final multiple-choice test would reflect such findings.

Subjects

Forty-eight subjects were selected from Waterloo Elementary School, Waterloo, Wisconsin. Waterloo is a city of 2,000 residents and it is somewhat of a composite of community types. It is near enough to Madison, Wisconsin, the State Capitol, to be considered a suburb, yet it is in a rural area, and it has some light industry. Thus, some of the school children are from farm families, others have fathers who commute to Madison, and other parents are employed in local industry.

The 48 subjects included 16 second, 16 fourth, and 16 sixth graders, Each group of 16 included 8 boys and 8 girls. Each subgroup of 8 boys or 8 girls contained 4 who were designated high in reading ability, and 4 designated low in reading ability. The reading level split was based on class median scores on the Gates McGinity Reading Test Form A in grade two, and on the Nelson Reading Test, Form A, in grades four and six. Four boys and four girls from each side of the Median Score were randomly assigned to the sub-groups to be tested. Table 3:02 shows the mean reading score for each cell in the design by grade level, by sex and by reading ability. (For a description of all subjects, see Appendix D. A listing of all reading achievement tests is found in Appendix F.)

Table 3:02

Pilot Study B

Mean Reading Score for Each Cell by Grade Level, Sex and Reading Ability

There were 4 subjects in each cell

		Grade 2	Grade 4	Grade 6
Boys	High	3.2	6.1	8.5
	Low	2.2	3.7	6.3

Table 3:02 (cont.)

		Grade 2	Grade 4	Grade 6
Girls	High	3.3	5.4	9.2
	Low	2.4	3.9	6.0

Testing Procedures

Both oral and multiple-choice forms of the four test orderings were used. Combinations of four tests, two multiple-choice and two oral, were administered to each of the subjects.

To control for order and test-type effects, four two-day testing sequences were devised and one subject from each cell was tested with each of the four sequences. Each subject was tested with one oral and one multiple-choice test on one day, and another oral and multiple-choice test on the following day. Table 3:03 presents the four testing sequences used during Pilot Study B.

Table 3:03

Testing Sequences - Pilot Study B

First Day	Second Day
A1 Written - B1 Oral	A2 Oral - B2 Written
B1 Written - A1 Oral	B2 Oral - A2 Written
A2 Oral - B2 Written	A1 Written - B1 Oral
B2 Oral - A2 Written	B1 Written - A1 Oral

Each of the four testing sequences was administered to one subject from each cell (see Table 3:02). The oral tests were administered individually, and the multiple-choice tests were administered to each of the four groups of 12 subjects. All testing was done during an eight day period in March, 1969.

In summary, each of the 48 subjects in Pilot Study B responded to all 100 test items in two ways: orally, and through a multiple-choice test. Phonemic transcriptions of the tape recorded oral pronunciations were made by the investigator. All data were key punched for computer analysis. A computer program was written to tabulate the pronunciation proportions, and to test the statistical significance of the results using analysis of variance.

Results of Pilot Study B

The analysis examined the specific agreement of the oral and multiple-choice responses of each subject to each synthetic word. For computer purposes a 1 was assigned to each response pair (oral/multiple-choice) which was the same, and a 0 to each that was different. There were 320 responses to each vowel cluster at each of three grade levels; second, fourth and sixth.

The hypothesis tested was: There are no differences in subjects' oral (O) and multiple-choice (M-C) responses to synthetic words containing vowel clusters, that is, H_0: $\mu_O = \mu_{MC}$ $\alpha = 0.01$.

The dependent variable for this analysis was the specific agreement of each subject's oral and multiple-choice responses to each of the 10 synthetic words used to test each of the 9 vowel clusters. Using the ANOVA H computer program, a 10 x 2 x 2 x 3 analysis of variance, in which the main effects were 9 vowel clusters (and check items) sex, two reading levels and three grade levels, was performed on the oral/ written agreement scores. The main effects and interactions in this analysis, together with their F values, are given in Table 3:04.

As Table 3:04 shows Vowel Cluster, Reading Level and Grade Level were significant main effects ($p < .01$) and there was a significant interaction between Vowel Cluster and Grade Level ($p < .01$). Thus, there was deemed to be a significant difference in the subjects' oral and multiple-choice responses and the hypothesis could not be accepted.

Table 3:04

F Values of Main Effects and Interactions for Pilot Study B

Source of Variation	Degrees of Freedom	Mean Square	F Value	p <
Vowel Cluster	9,324	135.568518	44.05	.01
Sex	1,36	7.50	1.37	NS
Reading Level	1,36	407.008333	19.94	.01
Grade Level	2,36	117.352083	5.75	.01
Cluster X Sex	9,324	2.712963	0.88	NS
Cluster X Rdg. Level	9,324	3.239815	1.05	NS
Sex X Rdg. Level	1,36	1.008333	0.05	NS
Cluster X Grade	18,324	7.099769	2.31	.01
Sex X Grade	2,36	38.268750	1.87	NS
Rdg. Level X Grade	2,36	12.152083	0.60	NS

Table 3:04 (cont.)

Source of Variation	Degrees of Freedom	Mean Square	F Value	p <
Cluster X Sex X Rdg. Level	9,324	4.517593	1.47	NS
Cluster X Sex X Grade	18,324	6.252546	2.03	NS
Cluster X Rdg. Level X Grade	18,324	3.918287	1.27	NS
Sex X Rdg. Level X Grade	2,36	10.502083	0.51	NS
Cluster X Sex X Rdg. Level X Grade	18,324	3.032176	0.99	NS

To illustrate the main effects found significant in this analysis three tables are included. Table 3:05 shows that there was considerable variation in oral-multiple-choice agreement scores by vowel cluster. There was total agreement on from five to seven synthetic words for five vowel clusters: ai, ou, oa, ea, and ow, while for one, ou, there was agreement on only three synthetic words.

Agreement in oral and multiple-choice responses to synthetic words containing vowel clusters was also a factor of reading level and grade level as shown in Tables 3:06 and 3:07. Subjects of high reading level agreed on nearly seven of ten words, as compared to five of ten for subjects of low reading level. There was an upward progression in oral-multiple-choice agreement on nearly five synthetic words by second grade subjects, on nearly six words by fourth grade subjects, on more than six words by sixth grade subjects.

Further, the analysis showed that the instrument was reliable (Table 3:05). The mean oral and multiple-choice agreement score for the ten check items was 8.667. This mean is very high considering five of the check items were also synthetic words, though with predictable letter-sound correspondence. Subjects reliably responded to the instrument.

Table 3:05

Oral-Multiple-Choice Agreement

Means by Vowel Cluster Where Ten Items Were Used for Each Cluster

ou	au	ie	oo	ai	ow	oa	ea	ay
3.021	3.854	4.625	4.729	5.667	5.833	6.512	6.729	7.167
		Grand Mean			Check Items Mean			
		5.683			8.667			

Table 3:06

Oral-Multiple-Choice Agreement

Means by Reading Level Where Ten Items Were Used for Each Vowel Cluster

Grand Mean	High Reading Level	Low Reading Level
5.683	6.604	4.762

Table 3:07

Oral-Multiple-Choice Agreement

Means by Grade Level Where Ten Items Were Used for Each Vowel Cluster

Grand Mean	Second Grade	Fourth Grade	Sixth Grade
5.683	4.737	5.906	6.406

Development of the Final Instrument

The final multiple-choice test was constructed on the basis of the results of Pilot Study B. The results indicated that subjects used a wider variety of responses on the oral test than on the multiple-choice test. These oral preferences were considered in the development of the final instrument by the selection of an additional alternate choice for each item for each of the nine vowel clusters. Thus, the same 100 multiple-choice test items were retained, but four alternative choices rather than three were made available for each item in the final instrument.

Tables 3:08 and 3:09 are presented to show this modification. Table 3:08 presents the three most frequent oral pronunciations for each vowel cluster made by the Pilot Study B subjects. Each of the three pronunciations of each vowel cluster except ai—>/æ/ (because of its low frequency on both corpora) was included in the final multiple-choice instrument. In addition, examination of the raw data revealed four other oral pronunciations from the "other" categories which were given frequently enough to be included in the final instrument.

They were: au—>/o/, ie—>/ɛ/, oo—>/ə/, and ou—>/ə/.

Table 3:09 shows the four phonemic response choices included in the final instrument used in the study. These response choices not only reflected the common oral pronunciations of vowel clusters given in Pilot Study B, but included the principal and secondary pronunciations of each vowel cluster on both the type and token corpus, when those pronunciations were monosyllabic. (See also Table 3:10 for principal and secondary pronunciation proportions on the type and token corpus, and Table 2:21 for a more complete listing of type and token vowel cluster pronunciation proportions.) Since the nine vowel clusters were nearly always monosyllabic (see Table 2:04 to 2:21) only monosyllabic response words were included. All forms of the final multiple-choice instrument are presented in Appendix C.

Table 3:08

The Three Most Frequent Oral Pronunciations of Each Vowel Cluster made by Pilot Study B Subjects

Vowel Cluster	Phoneme	Percentage	Vowel Cluster	Phoneme	Percentage
ai	/e/	71.3%	oa	/o/	75.6%
	/ɪ/	5.3%		/ɔ/	8.6%
	/æ/	5.0%		/au/	6.3%
	others	18.4%		others	9.5%

Table 3:08 (cont.)

Vowel Cluster	Phoneme	Percentage	Vowel Cluster	Phoneme	Percentage
au	/ɔ/	52.6%	oo	/u/	58.3%
	/æ/	12.0%		/ʊ/	19.6%
	/au/	11.0%		/o/	11.0%
	others	24.4%		others	11.1%
ay	/e/	82.3%	ou	/au/	43.4%
	/ɪ/	6.1%		/ʊ/	10.5%
	/ai/	3.2%		/u/	10.0%
	others	8.4%		others	36.4%
ea	/i/	75.3%	ow	/au/	48.3%
	/ɛ/	11.3%		/o/	43.6%
	/e/	7.0%		/ɔ/	4.0%
	others	6.4%		others	4.1%
ie	/ai/	42.4%			
	/i/	32.5%			
	/ɪ/	8.5%			
	others	16.8%			

Table 3:09

The Four Phonemic Response Choices for Each Vowel Cluster on the Modified Multiple-Choice Instrument

Vowel Cluster	Phonemic Response Choice in the Study			
ai	/e/	/ɛ/	/ai/	/ɪ/
au	/ɔ/	/æ/	/au/	/o/
ay	/e/	/ɛ/	/ai/	/ɪ/
ea	/i/	/ɛ/	/e/	/ə/
ie	/ai/	/i/	/ɪ/	/ɛ/
oa	/o/	/ɔ/	/au/	/a/
oo	/u/	/ʊ/	/ə/	/o/
ou	/au/	/ʊ/	/ə/	/u/
ow	/au/	/o/	/ɔ/	/[illegible]/

Table 3:10

Principal and Secondary Pronunciation Proportions of the Nine Vowel Clusters on the Type and Token Corpora

		Type Corpus		Token Corpus	
ai	principal	/e/	.86	/ɛ/	.39
	secondary	/ai/	.07	/e/	.27
au	principal	/ɔ/	.92	/ɔ/	.91
	secondary	/o/	.03	/æ/	.09

Table 3:10 (cont.)

		Type Corpus		Token Corpus	
ay	principal	/e/	.89	/e/	.90
	secondary	/ɪ/	.06	/ɪ/	.07
ea	principal	/i/	.53	/i/	.57
	secondary	/ɛ/	.23	/ɛ/	.23
ie	principal	/i/	.27	/i/	.47
	secondary	/ɪ/	.15	/ə/	.14
oa	principal	/o/	.75	/o/	1.00
	secondary	/oa/	.10		
oo	principal	/u/	.62	/ʊ/	.50
	secondary	/ʊ/	.27	/u/	.48
ou	principal	/au/	.50	/au/	.36
	secondary	/ə/	.38	/ʊ/	.26
ow	principal	/o/	.51	/au/	.51
	secondary	/au/	.48	/o/	.47

The Study

The study was planned to examine several questions about factors related to the pronunciation of vowel clusters:

1. How well do children's pronunciations of vowel clusters in synthetic words approximate the actual pronunciation frequencies of the same vowel clusters in both a type and a token corpus?

2. What differences are there in the vowel cluster pronunciation frequencies of good and poor readers?

3. Do boys and girls differ in their pronunciations of vowel clusters?

4. What differences are there in the vowel cluster pronunciations of second, fourth, and sixth grade subjects?

5. Do children of different community types differ in their pronunciations of vowel clusters?

In addition to testing hypotheses examining these relationships, the study was designed to provide information about three further questions.

1. Will subjects' pronunciations of vowel clusters be more closely related to the letter-sound correspondences on the type corpus or on the token corpus?

2. Will consonant environment affect the pronunciation of vowel clusters in synthetic words?

3. Will word position affect the pronunciation of vowel clusters in synthetic words?

These three questions were not tested statistically, but the raw data was examined. A discussion of these questions is contained in Chapter 4.

Selection of Subjects

The school authorities of three distinct community types (rural, suburban, urban) agreed to participate in the study; Seneca, Cedarburg, and Racine, Wisconsin. Seneca is a rural village in Western Wisconsin with a population of 137 (rural). Ninety-seven per cent of the district's 547 pupils are bussed to school from surrounding farms. Cedarburg, a community of 10,000, is a northern suburb of Milwaukee (suburban). Many of its residents cummute to Milwaukee for their employment, and Cedarburg is in one of the fastest growing counties in the United States. Racine is an urban city of 100,000 and is considered the most industrial community in Wisconsin. Many of its residents are factory employees.

Two classrooms at each of the three grade levels, second, fourth, and sixth, in each of three school systems were selected for the study. This resulted in an initial sample of 453 elementary school pupils.

Seventeen of these subjects were omitted because they were not present during one of the two days of testing. It was determined that the loss of such a small number of subjects would not affect the outcome of the study. On the other hand, had the subjects been retained and tested at a later date, the effects of these delayed responses would have been uncertain. Thus, 436 subjects, all of whom were tested on two consecutive days, were included in the study. The distribution of these subjects is summarized in Table 3:11.

Table 3:11

Distribution of Subjects by Community, Grade and Sex

		Rural	Suburban	Urban	Totals
Grade 2	Male	21	23	27	133
	Female	16	25	21	
Grade 4	Male	20	30	32	142
	Female	16	24	20	
Grade 6	Male	16	32	39	161
	Female	20	28	26	
	TOTAL	109	162	165	436

There were 109 rural, 162 suburban, and 165 urban pupils included in the sample. Of these, 133 were in second, 142 in fourth and 161 in sixth grade. The sample consisted of 240 boys and 196 girls. Reading level was determined by a median split for each sex in each classroom. (The reading achievement tests used are listed in Appendix E.) This resulted in 202 subjects of low reading level and 234 subjects of high reading level.

The participating classes at each grade level in each community were selected randomly from a list of all classes at these grade levels in each district. In the case of Seneca, rural, however, there were only two classes at each grade level, so the sample there consisted of all second, fourth and sixth grade pupils who were not absent during the testing. At Cedarburg, suburban, the classes were selected randomly from a minimum of eight classes at each grade level. The Racine school system, urban, is a unified district encompassing urban, suburban and rural schools. Because suburban and rural pupils were being tested in Seneca and Cedarburg, Racine school authorities randomly assigned classes from schools designated "inner city" or "urban." Table 3:12 summarizes the class identifications, grades, schools, school districts, and median reading scores for the study. For a description of all subjects, see Appendix E.

Table 3:12

The 436 Subjects in the Study by Class, Grade, School,

District, and Median Reading Scores by Sex

Class	Grade	School	District	Male Median Reading Score	Female Median Reading Score
A	2	South	Seneca	2.7	2.8
B	4	South	Seneca	4.3	6.2
C	6	South	Seneca	7.7	8.7
D	6	Lynxville	Seneca	6.5	7.2
D	4	Seneca	Seneca	4.7	4.8
F	2	Seneca	Seneca	2.8	2.8
G	2	Westlawn	Cedarburg	3.1	3.3
H	2	Hacker	Cedarburg	1.8	2.2
I	4	Westlawn	Cedarburg	4.4	4.9
J	4	Lincoln	Cedarburg	4.8	5.0
K	6	Washington	Cedarburg	6.6	7.2
L	6	Washington	Cedarburg	6.0	7.0
M	4	Janes	Racine	2.3	4.0
N	6	Janes	Racine	5.2	4.9
O	2	Janes	Racine	1.5	1.6
P	2	McKinley	Racine	1.8	1.7
Q	6	McKinley	Racine	6.7	6.6
R	4	McKinley	Racine	3.8	3.6

Measurement Procedures Used in the Study

In this investigation pupils responded to the final multiple-choice test. Since the multiple-choice test was constructed to reflect the oral preferences of Pilot Study B the oral test was omitted. Each pupil responded to all 100 test items over a two-day period, one randomization of 50 items the first day, and a randomization of the other 50 items the next day. As with Pilot Study B, each of the two subtests of 50 items was arranged in two orderings: A1, A2, B1, and B2. Eight two-day testing sequences were possible; the sequences were labeled A through H, as shown in Table 3:13.

Each of the 436 subjects was assigned a code number, then the eight testing sequences were assigned sequentially to the subjects. That is, subject #1 followed sequence A; #2, B; #3, C, etc. The tests were administered to class groups on two consecutive days in each community during late April and early May, 1969.

In summary, all 436 subjects responded to the same 100 synthetic words, 50 items a day on two consecutive days. Both boys and girls at each grade level and in each community followed each of the eight testing sequences.

Table 3:13

Testing Sequence During the Study

Sequence Label	First Day	Second Day
A	A1	B1
B	A1	B2
C	B1	A1

Table 3:13 (cont.)

Sequence Label	First Day	Second Day
D	B1	A2
E	A2	B1
F	A2	B2
G	B2	A1
H	B2	A2

Hypotheses and Statistical Analyses for the Study

The study was designed to examine the relationships between vowel cluster pronunciations by subjects of high and low reading level, male and female subjects, second, fourth, and sixth grade subjects, and urban, suburban and rural subjects, and to analyze the observed pronunciations in relation to existing letter-sound correspondences of vowel clusters in common English words.

For this investigation, two sources of letter-sound correspondences of vowel clusters were used: the type corpus and the token corpus. As defined on page 16, the type corpus is a body of 20,000 common words; a 1963 revision of the Thorndike frequency count. In this corpus, each of the 20,000 words was considered a type. That is, each word (type) received the weight of one regardless of its frequency of occurrence in the written materials analyzed to determine the corpus.

The token corpus contained the 1000 most frequent English words according to the 1967 Kučera-Francis study. This study provided a ranked listing of more than 50,000 words and listed the total occurrence of each word in a sample of 1,014,232 words of natural language text. A token was considered an "individual" word and was counted each time it occurred. The present investigator analyzed the 1000 most frequent words, in the token corpus, to determine the frequency of pronunciation of the nine vowel clusters based on word tokens. That is, each word was multiplied by its number of occurrences in the sample of words analyzed by Kučera and Francis.

Thus, the type corpus contained 20,000 words and the letter-sound correspondences of the vowel clusters reflected a single occurrence of each word containing a vowel cluster spelling. The token corpus contained only the 1000 most frequent words, and the letter-sound correspondences were based on the total occurrences of each word containing a given vowel cluster spelling. For example, one correspondence of the vowel cluster ou was /ʊ/. On the type corpus this correspondence was true in 1.3% of its occurrences, while on the token corpus this correspondence occurred 25.9% of the time. This difference was due to three very common words, would, could, and should. On the type corpus each of these words was counted once, while on the token corpus each of these words was multiplied by its number of occurrences.

An underlying question of this investigation was whether subjects' vowel cluster pronunciation would be related to vowel cluster pronunciation proportions on either the type corpus or token corpus.

The study was constructed to test 12 null hypotheses about vowel cluster pronunciation scores. To test these hypotheses, two analyses were performed. For each subject, two frequency difference scores were calculated for each vowel cluster in each analysis. These frequency scores were obtained in the following manner. Of the four response choices to each item on the test, two pronunciations of each vowel cluster were used in each analysis. They were the principal and secondary pronunciation proportions on the type corpus in analysis one, and the principal and secondary pronunciation proportions on the token corpus in analysis two (see Table 3:10). Each subject's responses to the ten items for each vowel cluster were analyzed to determine the number of responses which were principal and secondary pronunciations on the type corpus, and the number of responses which were principal and secondary pronunciations on the token corpus. Then, the frequency differences were calculated as follows:

Analysis One: The principal pronunciation proportion for each vowel cluster on the type corpus minus the principal pronunciation proportion actually occurring, and the secondary pronunciation proportion for each vowel cluster minus the secondary pronunciation actually occurring. For example, the principal pronunciation of ai on the type corpus was /e/. Its proportion was .86. The secondary pronunciation was /ai/ at .07. Assuming a subject pronounced ai—>/e/ on eight test items, the difference score would be .06 (.86 - .80 = .06). If he had selected the secondary pronunciation, /ai/ on two items the difference score would be -.13 (.07 - .20 = .13).

Analysis Two: The prinicpal pronunciation proportions for each vowel cluster on the token corpus minus the principal pronunciation actually occurring, and the secondary pronunciation proportions for each vowel cluster minus the secondary pronunciation proportions actually occurring.

The scores used to test the hypotheses (the dependent variables for each analysis) were defined in each analysis as the sum of the differences between the subject's principal and secondary pronunciation proportions and the principal and secondary pronunciation proportions on the corpus. For example, continuing with the hypothetical subject discussed above, his difference score for ai would be .07 (.06 and -.13 = .07). These sums were used to test the following hypotheses:

Hypothesis One: There is no difference in the type corpus (TP) difference scores of second (G2), fourth (G4), and sixth (G6) grade subjects, that is: $H_{1(TP)}$: $\mu_{G2} = \mu_{G4} = \mu_{G6}$.

Hypothesis Two: There is no difference in the type corpus (TP) difference score of male (M) and female (F) subjects, that is: $H_{2(TP)}$: $\mu_M = \mu_F$.

Hypothesis Three: There is no difference in the type corpus (TP) difference scores of subjects of high (H) and low (L) reading levels, that is: $H_{3(TP)}$: $\mu_H = \mu_L$.

Hypothesis Four: There is no difference in the type corpus (TP) difference scores of subjects of suburban (S), urban (U), and rural (R) communities, that is: $H_{4(TP)}$: $\mu_S = \mu_U = \mu_R$.

Hypothesis Five: There is no difference in the type corpus (TP) difference scores of the eight vowel clusters, that is: $H_{5(TP)}$: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$.

Hypothesis Six: There is no difference in the type corpus (TP) difference score of principal (P) and secondary (S) response types, that is: $H_{6(TP)}$: $\mu_P = \mu_S$.

Hypothesis Seven: There is no difference in the token corpus (TK) difference scores of second (G2), fourth (G4), and sixth (G6) grade subjects, that is: $H_{7(TK)}$: $\mu_{G2} = \mu_{G4} = \mu_{G6}$.

Hypothesis Eight: There is no difference in the token corpus (TK) difference score of male (M) and female (F) subjects, that is: $H_{8(TK)}$: $\mu_M = \mu_F$.

Hypothesis Nine: There is no difference in the token corpus (TK) difference scores of subjects of high (H) and low (L) reading levels, that is: $H_{9(TK)}$: $\mu_H = \mu_L$.

Hypothesis Ten: There is no difference in the token corpus (TK) difference scores of subjects of suburban (S), urban (U), and rural (R) communities, that is: $H_{10(TK)}$: $\mu_S = \mu_U = \mu_R$.

Hypothesis Eleven: There is no difference in the token corpus (TK) difference scores of the seven vowel clusters, that is: $H_{11(TK)}$: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$.

Hypothesis Twelve: There is no difference in the token corpus (TK) difference score of principal (P) and secondary (S) responses, that is: $H_{12(TK)}$: $\mu_P = \mu_S$.

In addition to testing these hypotheses, the raw data were examined to provide answers to the following questions:

1. Will subjects' pronunciations of vowel clusters be more closely related to the letter-sound correspondences on the type corpus or on the token corpus?

2. Will consonant environment affect the pronunciation of vowel clusters in synthetic words?

3. Will word position affect the pronunciation of vowel clusters in synthetic words?

These questions are discussed in Chapter 4.

The ANOVA-FINN computer program, which treats unequal n's, was used for the two analyses of the data. A 3 x 2 x 2 x 3 x 8 x 2 analysis of variance (with repeated measures on the last two factors) in which the main effects were three grade levels, sex, two reading levels, three community types, eight vowel clusters* and two response types (principal and secondary) was performed on the type corpus frequency difference scores. A 3 x 2 x 2 x 3 x 7 x 2 analysis of variance (with repeated measures on the last two factors) in which the main effects were three grade levels, sex, two reading levels, three community types, seven vowel clusters* and two response types, (principal and secondary) was performed on the token corpus frequency difference scores. Geisser-

---

*In analysis one (type) the vowel cluster oa was omitted because its secondary pronunciation /oə/, being disyllabic, was not offered as a response choice on the instrument. In analysis two (token) the vowel cluster ie was omitted for the same reason, and oa was omitted because it has no secondary pronunciation on the token corpus (see Table 3:10). Both vowel clusters were included in the study, however, because of their frequency of occurrence and their extreme dissimilatory in principal and secondary pronunciation frequencies. These vowel clusters will be discussed in Chapter 4.

Greenhouse corrections on degrees of freedom for repeated measures were used. Duncan's New Multiple Range Test was used to make post hoc comparisons among the means of main effects found significant.

Before the analyses were run on the data, complete tabulations of all 436 subjects' responses were made and pronunciation percentages calculated. These tabulations are presented in Chapter 4.

Summary of the Study

The purpose of this study was to investigate elementary pupils' pronunciations of vowel clusters and to analyze factors which may relate to pronunciation preferences. The study was done in three stages: Pilot Study A, Pilot Study B and the Study.

An instrument was constructed to measure pupils' pronunciations of vowel clusters in synthetic words. During Pilot Study B, 48 subjects gave oral pronunciations to 90 synthetic words and ten check items, and completed a 100-item vowel cluster multiple-choice test containing the same items. After the pilot study, the final multiple-choice instrument was developed to reflect major oral pronunciations given.

During the Study 436 subjects, male and female second, fourth, and sixth graders of high and low reading level from suburban, urban, and rural communities, responded to the modified 100-item multiple-choice test.

The statistical technique of analysis of variance was used to analyze the data in an evaluation of 12 hypotheses.