DOCUMENT RESUME

ED 300 820 CS 211 578

AUTHOR Farr, Roger; And Others
TITLE An Examination of the Writing Vocabulary of Children in Grades Two through Eight: A Study of a National Sample of Children's Writing.
INSTITUTION Indiana Univ., Bloomington. Center for Reading and Language Studies.
SPONS AGENCY Harcourt Brace Jovanovich, Inc., New York, N.Y.
PUB DATE 88
NOTE 79p.
PUB TYPE Reports - Research/Technical (143) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Elementary Education; *Spelling; *Spelling Instruction; *Word Lists; Writing Research; *Writing Skills
IDENTIFIERS Spelling and Writing Patterns; Spelling Patterns

ABSTRACT

To determine the words that children misspell as they write and then to analyze those words as they represent different spelling patterns at different grade levels, for different proficiencies of writing, and in the context of the essay in which the misspelling occurred, this study examined writing samples from across the country. The study was based on the belief that it is the words that children actually misspell in their writing that represent the words and patterns that need to be taught. The samples from this study came from three national writing assessment and four state writing assessment programs and included 21,876 writing samples from children in grades 2-8. Findings produced a set of seven word lists at grades 2 through 8 and the "Spelling List Concordance: Words Taught in Seven Major Spelling Programs." To investigate spelling avoidance in the context of composing stories a second pilot study explored the extent to which students avoid writing words because they are unsure of how to spell them. Subjects, 27 students from grades 1-5 in rural schools near a major university town in the Midwest, were asked both to tell and to write the same story in response to a picture prompt. Results indicated that most of the time there was no evidence of direct word or phrase substitution at all. However, the conclusions leave unanswered questions about spelling avoidance, and this may be because of the study's design. (Eleven tables of data are included, and one appendix, a spelling avoidance study protocol, and 38 references are attached.) (MS)

AN EXAMINATION OF THE WRITING VOCABULARY OF CHILDREN IN GRADES TWO THROUGH EIGHT

A Study of a National Sample of Children's Writing

Roger Farr
Caroline Beverstock
Bruce Robbins

Center for Reading and Language Studies
Indiana University 1988

This study was conducted under a grant from Harcourt Brace Jovanovich, Inc.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

The authors would like to thank a number of people who worked many long hours collecting writing samples, reading essays, coding vocabulary, entering data, and conducting computer analyses.

Sandra Mangurian, Director of Research at Harcourt Brace Jovanovich, provided advice, support, and encouragement throughout the entire project. A National Advisory Panel provided counsel in conceptualizing the study and offered many useful suggestions regarding the analyses.

NATIONAL ADVISORY PANEL	
Michael Beck	**Beck Evaluation and Testing Associates**
Dianne Bloom	**New Jersey Department of Education**
Wendy Littlefair	**Measurement Incorporated**
Ina Mullis	**National Assessment of Educational Progress**
Pat Porter	**Texas Department of Education**

In addition to serving on the Advisory Panel, Mike Beck helped to "dream up" the study, assisted with the collection of writing samples, and made numerous recommendations regarding data analyses.

The writing samples were collected from a variety of sources. The Psychological Corporation provided all of the writing samples used in the norming of the METROPOLITAN ACHIEVEMENT TEST. Tom Williamson, President of The Psychological Corporation, and Joan Sonnenschein, Senior Project Manager, were key people in providing these samples which helped to ensure the national representation of the writing samples. The National Assessment of Educational Progress (NAEP) provided data tapes of the NAEP writing assessment. These data also enhanced the national representativeness of the writing samples. Ina Mullis, Senior Consultant at NAEP, and a member of the National Advisory Panel assisted in securing these data.

We would also like to thank those directors and staff members of state and local school district assessment programs who helped to secure writing samples. In addition to the MAT and NAEP essays, additional essays were provided from state and local school district assessment programs in Connecticut, Georgia, Rhode Island, and Texas.

After the writing samples were collected, the task of coding and entering the data into computer files was begun. The management of 24,000 writing samples and a staff to complete the task was very complicated. Cheryl Logsdon was superb in managing these responsibilities. Her staff included the following coders, readers, and typists:

Staff Manager

Cheryl Logsdon

Coders and Readers

Julie Anderson
Shelley Criswell
Brian Eldridge
Kris Hart
Becky Hart
Amy Klingenberger
Donna Lockett
Craig McCartt
Pete McCluskey
Meg McCluskey
Yalonda McTush
Kelley Miller
Tiffany Moore
Brian Townsend
Theresa Weaver

Typists

Linda Applegate
Rose Bault
Marian Brazzell
Ruth Eppele
Dawn Hurley
Sharon Kale
Jin Kang
Brenda Lakatos

Cheryl Kelleher and Katherine Lee joined the research team in the second year of the project and added valuable insights and managed several aspects of the final analyses.

The computer analysis of a data base as large and cumbersome as this was a major assignment. It necessitated the development of new programs for data entry and analyses as well as the development of computer routines to manage the data. The Indiana University Bloomington Academic Computing Services (BACS) provided excellent services which went far beyond the specifications of the job. We are especially grateful to Polley A. McClure, Dean of Academic Computing, and Doug Grover, Manger of Large Systems Support Services. Bob Eckert, Senior Project Manager at BACS, carried out all of the analyses and developed the data base. It is difficult to imagine how we could have ever completed the project without Bob.

Finally, we want to acknowledge the Center for Reading and Language Studies at Indiana University which provided substantial support services and assisted with maintenance of the collection of writing samples. We are especially grateful to the Center for always saving the window seat locations for the spelling project staff.

INTRODUCTION

Spelling often seems to be at the center of the public's attempts to judge the quality of education. Each year the winner of the National Spelling Bee is presented at the White House, on the Johnny Carson show, and to the nation with all the reverberations of a national hero, and the impression is given that spelling is the epitome of quality education. Local media provide similar recognition to regional spelling bee winners. It would not be difficult for a foreigner to conclude that spelling is one of the most important subjects taught in our schools.

Almost in direct opposition to the belief in the importance of spelling are those educators who have studied writing and have been instrumental in increasing attention to the teaching of writing. They proclaim, almost without exception, that attention to spelling can stifle good writing, that spelling should not be taught as a separate subject, that children will learn to spell if they are given more opportunities to participate in reading and writing activities--and if they don't learn to spell through these activities--they are probably destined to be poor spellers because there really is no way to teach spelling. Furthermore, they argue that spelling should be considered only when a piece of writing is being edited and prepared for publication.

How can we explain these differences? Why is spelling held up by the public as something that is critical to good education and yet hear from the writing experts that spelling is not very significant. While *the truth* about such complex issues is not the focus of this study, it is important to establish *six beliefs about spelling which gave impetus to this study*.

Is Spelling Important?

It is almost certain that the general public gives too much attention to spelling as an indicator of the quality of education. Perhaps that is because of the American need for competition and winners. Spelling contests may be responsive to that need. Furthermore, to most people spelling *seems* like something that ought to be taught in schools, and it is something that can be easily evaluated by those who employ the schools' graduates. Certainly many employment applications have been cast aside because of misspelled words--sometimes regardless of the other qualifications of the job applicant. "Can't they even teach them to spell in our schools?" is a constant lament of many business people.

Yet, despite the misplaced emphasis on correct spelling, there is something about good spelling that should not be overlooked. Correct spelling, especially in writing that is to be read by others--as most writing is destined to be--suggests that the writer cared enough to take the effort to see that the writing was of good quality. He or she chose good quality paper, wrote with attention to correct usage, attempted to formulate his or her ideas and feelings so they would be communicated to the reader, and *checked to be sure that the words were spelled correctly*. Certainly we can live with poor spelling, but we can't get very far without attention to quality--and spelling is one of those things that is used as an indication of attention to quality. Spelling may be a partial stand-in for determining whether one cares about quality.

Can Spelling Be Learned?

There is no question that a child can learn to be a better speller. There are, of course, many arguments about how spelling can best be learned. The debates focus

on whether lists of words should be used, whether word patterns should be learned, and whether spelling should be learned separate from reading and writing. But the fact that spelling can be learned is not debatable. Even though some people never seem to learn to spell very well, they have learned to spell many words correctly, and there certainly was a time in their lives when they could not spell any words correctly. These people did *learn* to spell. They just seem not to be able to spell, or don't care to pay attention to spelling correctly some of the words they use in their writing.

Does Misspelling Interfere With Communication?

There is no question that any particular misspelling is a writer's best representation of word the writer intended to use. Thus, **incorrect spelling rarely changes a writer's intended message**. What may be changed is the **reader's understanding of the message**. There are certainly times when a misspelled word results in a misunderstanding on the part of the reader. It is also true that large numbers of misspelled words cause a reader to have to slow down and focus attention on determining the word a writer wanted to use. When a reader is paying attention to word recognition, it is quite probable that less attention is being paid to understanding the author's ideas. It is almost akin to a beginning reader who has difficulty decoding words and, therefore, has little attention left to understand the ideas. Therefore, though misspellings sometimes cause interference for the reader, spelling makes only a slight difference in most cases.

Is Spelling Separate from Other Aspects of Good Writing?

In addition to making the reader's task easier, good spelling is intertwined with correct usage and good writing. One cannot teach the spelling of homonyms such as

there and *their* without teaching correct usage. The use of spelling checkers on computers and the use of a dictionary to check spelling errors is dependent on a writer's knowledge of correct usage. Many writing samples have gone through a spelling checker on a computer and the computer has indicated no spelling errors. Yet, when the writing is examined there are many spelling errors resulting from incorrect grammatical usage or from faulty understanding of a word's meaning.

What is the Best Source of Spelling Words?

One aspect of spelling that almost everyone agrees on is that when spelling is taught the words to be taught should be those that children have difficultly spelling correctly. Many spelling authorities argue that a child's spelling words should come from the words he or she has misspelled in writing. The focus on writing as the source of words to be taught is well established.

Are Words or Patterns to be Taught?

The focus should not just be on the specific words to be taught but rather should be on the patterns that are represented by those words. Furthermore, the spelling patterns go far beyond mere phoneme/grapheme matching. Spelling patterns are based on syntax and semantics as well. There is still much to learn about misspelling patterns. This study is one attempt to advance our knowledge about spelling and writing.

OVERVIEW OF THE STUDY

This study was primarily an attempt to determine the words that children misspell as they write and then to analyze those words as they represent different spelling patterns at different grade levels, for different proficiencies of writing, and in the context of the essay in which the misspelling occurred. The major design questions for the study focused on the types of writing samples to collect and the procedures to be used to analyze and record the writing samples.

The belief that children's writing should be the source of words to be taught is well established in the literature. Gentry (1988), for example, says that, "Common sense tells us that a spelling program must teach words that students actually use in writing." It is even more obvious that some of the words that children use in their writing present more spelling difficulties than others. This study is based on the belief that it is the words that children actually misspell in their writing that represent the words and patterns that need to be taught.

There are, however, numerous spelling lists which already exist and many of these lists are based, at least in part, on samples of children's writing. Most of these lists were, however, developed many years ago (Thorndike and Lorge, 1944; Kucera and Francis, 1961). Other lists, which have been collected more recently, are based on essays which have been collected conveniently and do not adequately represent a national sample (Smith and Ingersoll, 1984).

Furthermore, this study was based on the conviction that only by carefully examining changes in spelling patterns and spelling improvement across grades and

across levels of writing could one arrive at valid conclusions about misspellings--and actually determine the words that need to be taught to children. The mistake could be made that a particular word that is often misspelled by second grade children is a word that should be taught at that grade. This study postulated that a particular misspelled word can be understood only when it is examined in comparison to the variety of ways that word is misspelled and by examining the changes in the spelling and the patterns of misspelling as that word is used in the writing of children at different development levels. Thus, it was important to collect and analyze writing samples collected from students at all grade levels and from students who were good, average, and poor writers.

The inclusion of students who represent a range of writing abilities was so important that it forced the researchers to use only writing samples which were collected from students under known conditions and essays that had been scored by a system for which there was known reliability and validity. The use of essays which represent a convenience sample submitted by teachers could easily result in teachers sending in only the best essays from a class--or only the worst essays! Furthermore, the quality of writing of the essays would not be known under these conditions, and to set up a valid and reliable grading process to determine the quality of writing of each writing sample would be very difficult and very costly. Rather, it was decided to collect writing samples from large scale writing assessment projects. The writing samples had to be already evaluated and had to include the following: 1) the specific writing prompts and directions, 2) the scoring procedures that were used, and 3) the validity and reliability of the scoring procedures.

In addition to carefully defining the data base in terms of the levels of writing, the study was designed to include the entire writing sample in the analysis. This was based on the already stated belief that misspelled words and misspelling patterns can be understood more thoroughly if they are examined in context. This study sought to answer such questions as :

> whether a writer continued with the same misspelling pattern throughout a piece of writing
>
> whether a misspelling pattern was related to other words in the writing
>
> whether the misspelling pattern was based on a particuiar usage or conceptual difficulty

Thus, it was necessary to retain the entire writing sample, including the misspellings, for further analysis.

The use of writing assessment essays provided an additional value for this study. Most of these essays are one draft writing. That is, the students do not have an opportunity to revise their writing. Therefore, they could not check with classmates, teachers, parents, or dictionaries as to the correct spelling of words. Students know they are writing as part of a test and that their writing will be evaluated. In this kind of writing, students are expected to do their best work. Therefore, we assume the misspellings in these essays are the writer's best attempts at spelling. In contrast, when informal, everyday writing is collected, it is quite possible that little attention is given to correct spelling.

Finally, it was essential that the study include writing samples from across the entire country. The use of writing samples from state and national writing assessments

provided an opportunity for this study to assure that the writing samples provided a nationally representative sample. Several design features of this study assured that the sample of essays was nationally representative of students' writing in grades two through eight.

WORD LIST REVIEW

Word listing represents an effort to approach spelling and vocabulary empirically. The purpose of the generation of the lists is to provide a solid base for the decisions about curriculum. Despite the many scientific efforts, development of the definitive list is elusive. These questions face the developer/researcher: Do children need to learn the words that children use or the ones adults use? How much overlap is there between adult and child vocabularies? Can a list be developed once and for all or does it require periodic adjustments due to the changing nature of the language? If so, how often does it need to be updated?

The word lists have been compiled from three different sources. One major set of lists has been developed from **writing vocabularies** (Jones, 1913; Bauer, 1916; Tidyman, 1921; Horn, 1926; Fitzgerald, 1931; Rinsland, 1945; Hillerich, 1966). The second source of lists was developed from **printed vocabularies** (Thorndike, 1921; Thorndike, 1931; Gates, 1935; Dolch, 1936; Thorndike and Lorge, 1944; Fry, 1957; Kucera and Francis, 1967; Carroll, Davies, and Richman, 1971; Harris and Jacobsen, 1972 & 1975; Sakiey and Fry, 1979). The third and least used source is **speaking vocabularies** (Dale, 1931; Dale and Chall, 1948; Murphy, 1957; Jones and Wepman, 1966; Wepman and Hass, 1969; Moe and Hopkins, 1975).

One analysis of word lists has been the comparison of the words chosen for different spelling series and the grades at which the words are presented. The *Spelling List Concordance* (Indiana University, 1988) presents seven of the major publishers lists, the placement for each word, mean grade placement, grade placement range,

and an agreement of placement index. There have been previous efforts at comparison, but with earlier editions of the programs (Hagerty, 1981).

The Horn list, though now over sixty years old, remains a major citation in the development of spelling lists. Horn (1926) listed 10,000 words most often used outside of school from 65 types of writing and 5,136,816 running words. He included business correspondence; personal correspondence; vocabulary of the letters of people of more than average literacy ability including well-known writers; vocabulary of letters printed in magazines and metropolitan newspapers; letters of application and recommendation; vocabulary of adult writing needs other than correspondence such as the vocabulary of minutes, resolutions, committee reports, the vocabulary of excuses written to teachers by parents; and the vocabulary of the letters of a single individual.

Horn found that certain words would fall in the first 100 in frequency no matter how much material was tabulated: I, THE, AND, TO, A, OF, BE, IN, WE, HAVE. In his study, all words were recorded including slang, colloquial, and supposedly obsolete words but not proper names, and words of less than four letters "since these have relatively small spelling difficulty,"

Another of the influential writers in this area, Fitzgerald (1931), is still listed as a reference by many of the major publishers of spelling series. His commentary on the word lists and particularly on the use of the word lists, remains valuable. He has written:

> It seems certain now that the course of study should be scientifically planned in order than children can be guided in a well-organized curriculum to learn the spelling or words that they need to write effectively in the situations that call for writing.

> One of the most important causes of poor spelling in the past, and to a great extent even in the present day, is invalid word lists. In many curriculums words have been chosen without careful attention to the validity of the sources. Words have been taken frequently from vocabularies which, although worthwhile for the purposes of the investigators, are not highly valid for writing in the elementary school. Too many words are presented which are of too little value. Too large a number of words to learn, many of which are not recognized as useful by the child, is a cause for discouragement and failure. In some curriculums, as much emphasis is given to the presentation of words which a child will write infrequently as to the teaching of those which he will write frequently.

These concerns face the users of word lists, the developers of curriculum and texts: the sources of words, their validity for the elementary school, the number of words to be presented, whether the words make sense for students to learn or seem to them to be nonsense, whether the, frequency of word use is taken into account, whether there is differentiation for "gifted, average, slow-learning," whether words are presented at the appropriate grades. Fitzgerald states:

> A most important fundamental for the spelling curriculum is a basic word list, scientifically selected and properly graded. (p. 3)
>
> The investigations most useful for word selection in spelling are those based upon real writing of child and adult. However, studies of oral and reading vocabularies are helpful in evaluating words for grade placement. (p. 5)

Within the Indiana University List we have developed both vocabulary and analysis of misspellings in one list, unlike many earlier efforts. "Data on the misspelling of words are of particular importance in selecting spelling words for various grades." (Fitzgerald, pg 39)

> In brief, the procedure for word selection involves the use of valid research to determine words common to child and adult writing, to discover words common to child and adult writing, to discover words highly crucial in adult writing which become progressively more valuable to children as they go from level to level, and to identify important words frequently used in child letter writing and school work but of low frequency in adult letter writing. (p. 39)

In deciding which words to use we must balance between adult utility and child utility of vocabulary. Some vocabulary is specific to children's contexts like TEACHER'S, HALLOWEEN, VALENTINES, HANDKERCHIEFS, SANTA CLAUS, ARITHMETIC, SPELLING, DOLL, AIRPLANE, BICYCLE, CLASSMATE. This extends the level of selection we had been calling "prompt specific" and we need to compare to adult list(s) to find these school or childhood specific words. We need to look for game specific language, toy specific language and other language specific to children..

Another issue is the permanency in the language of certain words. Horn looked at 5000 words, but he found that only 16 had entered the language after 1865 at the time of his study. Two examples are AUTOMOBILE and BASKETBALL. He did note that trade names were apparent on his lists. This may be sustained although many more new words may have become current given the technological gains in the first sixty years. Meanings and popularity of words also change over time without word extinction, as in the word COMPUTER, which not long ago was an infrequently used word meaning mathematician.

The existing corpus of word lists represents both a very useful tool and something which needs to be critically examined to be sure that lists and word choices are the best possible. New lists, such as the Indiana List, can provide new insight and perspective on the traditional words as well as new candidates for the lists.

SOURCES OF THE DATA

There were several requirements the writing samples had to meet to be included in this study. First, they had to be from national or state assessments of direct writing proficiency. These are called "direct" assessments of writing skill since the assessment is based on a piece of writing from the students as opposed to the "indirect" method, generally multiple-choice questions about writing correctness. Second, the samples as a whole were to be representative of the range of socioeconomic status, region, and gender across the United States. Third, the samples had to have been holistically scored under known conditions of validity and reliability. Finally, the samples had to have been written in response to specific prompts and with known administration protocols.

Acceptance of indirect measures of writing ability has decreased over recent years. The National Council of Teachers of English support the assessment of writing from samples rather than in indirect, multiple choice tests because they feel that the direct assessment leads more directly to an understanding of students' writing skills and planning of appropriate instruction. Direct assessment of writing includes several alternatives for scoring the writing samples -- primary trait, analytic, holistic, computer analysis, T-unit among others. Holistic scoring is preferable because it is the most researched, has been widely adopted and is the simplest to use. (Sopko, 1988) According to Jongsma and Dean (1988),

> Holistic scoring is best suited for reporting group performance, particularly in large scale evaluations such as district-wide or state-wide assessments. As such, holistic scoring offers a relatively simple way of assessing growth in writing proficiency over time, of measuring group

> differences in writing performance, and of determining the effectiveness of a writing program. (p. 12)

As such, the holistic scores of the writing samples in this study allow comparisons of the writing vocabularies of students of differing writing proficiencies and at different grade levels.

The samples for this study came from three national writing assessments and four state writing assessment programs. The national assessments were two sets of the *Writing Test* from the Metropolitan Achievement Tests and, for comparison, a set of papers from the National Assessment of Educational Progress's 1984 Assessment of the Conventions of Writing. The individual states contributing samples to the state-level portion of the study were Connecticut, Georgia, Rhode Island and Texas. Grateful appreciation is expressed to the cooperating state assessment offices for permitting the use of their samples in this study.

Two tables, **Sources of the Samples**, and **Sources of Samples by Grade Level** follow.

TABLE ONE: SOURCES OF THE SAMPLES

Assessment	Time of Testing	Grade(s)	Number	% of total
Metropolitan	Spring/85	2-8	6,692	30.6
Metropolitan	Spring/87	3, 6	10,054	46.0
Connecticut		4	306	01.4
Georgia		6, 8	1,220	05.6
Rhode Island	Spring/87	3	940	04.3
Texas	10/85	3, 5, 7	1,858	08.5
Texas	9/87	4, 6	806	03.7
TOTAL			21,876	
NAEP	1984	4, 8	870	

*not included in this table. NAEP will be used for comparison and later addition to the data base.

TABLE TWO: SOURCES BY GRADE LEVEL

Grade	Assessment	Number	% of grade	Total/Grade
Two	Metropolitan	962	100.0	962
Three	Metropolitan	1058	13.9	7637
	Metropolitan	4990	65.3	
	Texas	649	8.4	
	Rhode Island	940	2.0	
Four	Metropolitan	935	46.0	2045
	Texas	803	39.0	
	Connecticut	307	15.0	
	NAEP	*418		
Five	Metropolitan	796	73.0	1093
	Texas	297	27.0	
Six	Metropolitan	1064	16.0	6634
	Metropolitan	5083	77.0	
	Georgia	503	8.0	
	Texas	4	> 1.0	
Seven	Metropolitan	912	51.0	1796
	Texas	884	49.0	
Eight	Metropolitan	813	53.0	1530
	Georgia	717	47.0	
	NAEP	*452		
TOTAL		22970		

*Not included in this table. NAEP will be used for comparison and later addition to the data base.

TABLE THREE: GENDER

	Number	Percent
Female	9242	49.92
Male	9272	50.08
Missing	3362 (data were unavailable for Georgia, Rhode Island, and some of the *Metropolitan Tests*)	

Sample Description

In addition to the grade, specific assessment source, and gender, all the samples were coded for socioeconomic status, geographic region, and with standard scores which allow comparison across assessments. Socioeconomic status, geographic region, and the standard scores were designated as follows:

Socioeconomic Status: This variable was determined for each school district contributing student papers to the study. Socioeconomic status (SES) was defined by an index formed from a composite of the median family income and the percent of adults in the school district who were high school graduates, weighting each of these variables approximately equally. The data from which this index was formed were based on 1980 census data grouped geographically by school system boundaries. Once the index was determined for each district, the districts were assigned an overall SES code ranging from 1 to 5. This code was developed by breaking the entire SES index distribution into quintiles. A code of 1 indicates the highest quintile, 3 the middle, and 5 the lowest quintile of district SES as defined in this study.

TABLE FOUR: SES DISTRIBUTION

	Number	Percent
High	1112	6.84
High-Average	4029	24.80
Average	2028	12.48
Low-Average	4621	28.44
Low	4458	27.44
Missing*	5828	

*Socioeconomic status was not available for the samples from Connecticut, Rhode Island, and Georgia.

Geographic Region: All participating districts were divided into geographic regions of the country using the four regional categories used by the National Education Association: Northeast, Midwest, South, and West. Each district was appropriately coded using the state-by-region categories.

TABLE FIVE: REGIONAL DISTRIBUTION

	Number	Percent
East	4523	21.09
Midwest	3214	14.98
South	6707	31.27
West	7007	32.67
Missing	425 (Some regions were not available on the *Metropolitan*)	

Standard Scores: As described above for each set of papers included in the study, all essays had been previously scored by experienced, trained readers prior to inclusion in this study. The availability of this rating for each paper is a clear advantage of the present study over previous investigations of a similar nature. However, the use of several sources of papers, while providing a wide range of topics and student backgrounds, brought with it several scoring procedures and score scales. It was necessary to collapse these several scales into a single scale so papers from the various sources could be combined and generalizations drawn across sets of papers. This step was accomplished by transforming the several score scales into a single 1-99 scale. This transformed scale simply assigned a 1 to the lowest obtainable score on each original scale, a 99 to the highest possible such score, and intermediate values proportionately between these values. The transformed scale did not normalize the original scales or change the shape of their distributions in any way. The transformation simply permitted us to combine the several scales into a single combined scale metric.

TABLE SIX: STANDARD SCORES

Mean	S.D.	Range
50.874	19.39	98

Metropolitan

The Writing Test is one part of the sixth edition of the *Metropolitan Achievement Tests* (MAT6). The first set of samples used in this study comes from the norming set administered nationally in spring 1985.

> The MAT6 *Writing Test* was normed concurrently with the MAT6 Survey battery and the *Language Diagnostic Tests*. A two-stage sampling procedure was employed. During the first stage, a nationally representative sample was selected to take the Survey Battery and Diagnostic Tests. Participants were selected to represent the national student population in terms of geographic region, socioeconomic status, school system size, and ethnicity. In the second stage, a subsample was randomly selected to participate in the standardization of the MAT6 *Writing Test*. (Wiser, 1988, p. 7)

These data were carefully chosen to be nationally representative and provide an anchor across all the data in this study since each grade level from two to eight was included in the norming sample. The second set of *Writing Test* samples came from an equating study in spring of 1987 to add new prompts to the ones normed in 1985.

> A sample of 7547 students, representing 31 school districts and 20 states, were assessed. The students were stratified by grade level with 2269 third graders, 2738 sixth graders, and 2540 tenth graders participating. Because the study was for equating the prompts, not standardizing them, the sample was not selected to mirror national census data. However, an effort was made to sample school districts from a variety of geographic areas...(Jongsma and Dean, 1988, p. 5)

Since this was an equating study, the data from the second set of *Writing Test* samples present a special opportunity to compare the responses of a single writer to a pair of prompts. This comparison can be made both globally across the sample as well as on a writer-by-writer basis.

There are three levels of the *Writing Test*, Level 1 for Grades 2 through 4, Level 2 for Grades 5 through 8, and Level 3 for Grades 9 through 12. The test measures narrative/descriptive writing.

> For all prompts, students are directed to 'write a story about the picture...tell what is happening and what might happen next' in response to a pictorial stimulus. The narrative/descriptive type was selected in an effort to provide consistency across grade levels. That is, in the lower grades, nearly all writing is descriptive or narrative. However, in the upper grades, variety increases with a greater emphasis on expository and argumentative types, but not to the exclusion of narrative/descriptive. By using the narrative/descriptive type, the MAT6 *Writing Test* is able to provide a consistent measure of growth across grade levels. (Jongsma and Dean, 1988, pp. 2-3)

The *Writing Test* protocol requires standardized testing conditions. The manuals include directions for preparing parents, students, and the test administrators (suggested to be the classroom teacher to minimize the students' anxiety) for the test. The test takes a total of 30 minutes, 10 minutes for the directions and 20 minutes of writing. The students are directed to write on the two pages provided in the test folder.

All MAT writing samples were scored by two trained readers at Measurement, Inc.. Before beginning a set, the readers were trained on a set of "anchor papers" to assign a score on an eight-point scale where 1 was the lowest and 8 the highest possible score. If the scores assigned by the readers differed by more than two

points, there was a third reading. This occurs less than 1% of the time. The reliability of the readers ranged from .86 to .95 in the first samples and .92 to .96 on the equating samples. A between forms reliability of .70 to .75 is reported (Hogan, Farr, Prescott, Balow; 1987, p. 39). The validity of the *Writing Test* is demonstrated through its relationship to related measures on the MAT6. The *Writing Test* correlates with the other MAT6 subtests between .40 and .70. The correlations are higher with the language subtests than other parts of the MAT6.

NAEP

The National Assessment of Educational Progress is currently administered by the Educational Testing Service under a grant from the United States Department of Education. The national assessment has been "mandated by Congress to collect data over time on the performance of young Americans in various learning areas." (Norris, 1987)

> Its primary goals are to detect and report the current status of, as well as changes in, the educational attainments of young Americans, and to report long-term trends in those attainments.
>
> In 1984, NAEP conducted an assessment of attitudes and achievement in reading and writing for more than 100,000 students who were 9, 13, or 17 years old or in the corresponding modal grades 4, 8, or 11. For the writing portion of the assessment, nationally representative samples of students completed a variety of writing tasks and answered questions about their attitudes toward and experiences with writing and related activities.
>
> The writing task "Hole in the Box" ... was administered to all three grade/age levels. Of the 8,970 essays collected for this task, a subsample of approximately 500 essays at each age (9, 13, and 17) was randomly sampled, scored, and used to examine students' writing ability to use the conventions of written English. (Norris, p. 1)

This is the sample used in the current study. In selecting the samples to use in the study, "nationally representative probability subsamples were drawn from the total national sample....Results for these papers provide good estimates of national levels of performance." (Applebee, Langer, and Mullis, 1987, p. 10)

Students were given a picture prompt showing a large box that had a hole in it and an eye looking out of one of the openings. The directions to the students were, "Look carefully at the picture. Imagine yourself in the picture. Describe the scene and how you feel about what is going on around you. Try to make your description lively and interesting." (Norris, p. 2) Students had about 16 minutes to write the essay.

The samples were scored by "experienced English teachers, thoroughly trained in scoring students' adherence to the conventions of grammar and usage." (Applebee, Langer, and Mullis, 1987, pg 10) Papers received two overall scores, a "primary trait" score which was a rating of how well the writer accomplished the specific writing task demanded by the prompt, and an "overall fluency" or holistic score. The samples were also coded for every sentence type and error in convention. The samples were read by one reader with a 20 percent reliability check of rereadings. "Second scorers did not see the first scores and discrepancies were resolved by the scoring supervisor." (Beaton, et al, 1987, p. 184) The percentages of exact score point agreement for this prompt varied from .86 to .92 (Beaton, et al, 1987, p. 398).

In the current study, the NAEP data are being used in two ways. First, since it is a very carefully selected national sample, it serves as a comparison for this study's overall national sample. Second, once that comparison has been made, the samples will be added to the Indiana University Spelling List data base. Though the NAEP

samples total just 418 at grade four and 452 at grade eight, they represent between 3.1 and 3.2 million students at each grade level.

Connecticut

This study included 300 samples from the Fall 1986 Connecticut assessment. The writing assessment for the state of Connecticut was designed by the Department of Education in cooperation with a state-wide committee for content and technical concerns. They began with a review of all the existing tests (both direct and indirect), the texts in use in Connecticut, the state curriculum, and the writing goals at the district level. In addition, 3000 educators and non-educators were surveyed on appropriate objectives at each grade level. Following pilot tests statewide a committee reviews the results and makes the final determination of which prompt to use at each grade level. There is a single prompt per grade per year. The prompt for the Connecticut samples used in this study asked students to write, "If you could be any animal for a day."

The protocol allows students 45 minutes to respond to the prompt. They are directed to plan the writing, make notes or a first draft on scratch paper, but also reserve the time they need to write on the answer document. The formal answer document is all that is scored. In the scoring, despite the suggestion to do some prewriting, the samples are treated and scored as first draft writing. This protocol more than doubles the amount of writing time allowed by the *Metropolitan Writing Test* and therefore allows some comparison of such factors as the relative length of samples written with different time limits.

The samples are scored by Connecticut teachers under the supervision of the Psychological Corporation and Measurement, Inc. There are a minimum of two

readings scored on a scale of 1 to 4. There are third readings if scores differ by more than one point. (This third reading is required for approximately 4 to 7 percent of the samples.)

Georgia

Georgia's writing assessments were holistically scored on a scale from 1 (inadequate) to 4 (very good) on each of five "dimensions": Content/Organization, Style, Sentence Formation, Usage, Mechanics. The scoring was conducted by R. & R. Evaluations, Inc.. Two readers score each sample with discrepancies resolved by third readings. The prompts are of personal expressive writing and any kind of response is accepted. The test is conducted in a single sitting. An extra sheet is included for a rough draft. No dictionaries are allowed during the test. Indiana University Spelling List includes 1000 samples from the spring Georgia assessment at grades six and eight.

Rhode Island

The Rhode Island third grade writing assessment was administered in the Spring of 1987. The assessment was designed by the Rhode Island Department of Education, the Center for Evaluation and Research at Rhode Island College, and the Rhode Island Writing Consortium with the assistance of the Educational Testing Service. The purpose was to evaluate writing in a way that promotes the effective teaching of writing in elementary schools. (ETS Scoring Report, p.1) In contrast to the rest of the writing assessments in this study which are first draft writing, the Rhode Island assessment developers designed a

> Two-day writing assignment, one that gives the students some guidance in prewriting, drafting, revising, and editing. The full writing assignment was ...administered by classroom teachers in late March [1987].

> The topic ("something special") and the holistic scoring criteria allowed students a good deal of leeway in developing their own ideas and organizing strategies. On Day One of the administration, the students answered questions that helped them focus on the topic and consider how to shape their responses; then they wrote their first drafts. On Day Two, the students answered a series of revision questions designed to help them improve their first drafts. Finally, after writing their revised compositions, the students used an editing checklist to make final corrections.
>
> Only the final composition was scored, not the draft or the composing process. (ETS Scoring Report, p. 1)

The assessment was administered statewide and then read holistically by 34 teachers and administrators who were trained for two hours. Each sample was read twice. The resulting scores for the entire sample were: Superior, 16%; Good, 37%; Fair, 37%; and Poor, 10%. The readers were very consistent in their scoring; only 3% of the ratings differed by 2 or more points on a 4 point scale. The writing assignment and directions for revision were judged to be valid and reliable.

> It was apparent that the students understood the assignment and that they become fully absorbed in the task. The most skillful writers wrote extraordinarily fine compositions and even the weakest writers had much to say about the item they considered special.
>
> The children wrote about stuffed animals, mothers, toys, stories, places, family heirlooms, and so on. They employed a variety of appropriate techniques, including describing, making comparisons, citing reasons, narrating real or imagined experiences, categorizing, reflecting, and summarizing. Many of the children wrote with humor and wit, impressive word choice, a lively style, and, at times, poignant insight--qualities that made the scoring session all the more enjoyable for the readers. (ETS Scoring Report, p. 4)

The Rhode Island writing procedures were approved by the third grade teachers in Rhode Island, who also developed the scoring criteria, and served as the readers.

The Rhode Island Writing Consortium was involved in design and also in-service assistance to provide a bridge from the assessment to instruction. The reliability of the study was examined by the Center for Evaluation and Research at Rhode Island College (CERRIC).

Texas

The writing samples from Texas used in this study came from tryout assessments administered in October 1985 (n = 1870, Grades 3, 5, and 7) and September 1987 (n = 600, Grades 4 and 6). Writing assessment in Texas has been mandated by the state legislature and designed by school district personnel and Texas Education Agency staff as part of the Texas Educational Assessment of Minimum Skills (TEAMS). Its purpose is to promote effective writing instruction in the Texas public schools.

> TEAMS encourages teachers to help their students to improve their ability to think and reason clearly, to advance their thinking from simple to complex ideas, and to express themselves in written form. Many language specialists contend that learning to write is one of the most difficult yet probably one of the most important things a child does. The TEAMS test helps to assure that all students attending the Texas public schools will receive instruction in effective writing skills. (TEAMS Instructional Strategies Guide: Written Composition, pp. 4-5)

The assessments are field tested to assess the potential usefulness of the topics for statewide use. The Texas design tests students' abilities to write for different purposes (persuasive, informative, and expressive) and in different modes (narrative, descriptive, and classificatory). The audience, formal or informal, is specified in the prompt. Elaboration is required; it is considered to be an essential part of good writing and as such a student cannot score a three or four on the exam without sufficient

elaboration. (The scale ranges low to high from 0 to four.)

The writing samples are holistically scored under contract with Measurement, Inc. under the following conditions:

> Approximately 140 qualified readers at each grade level are employed annually to score the 1.1 million TEAMS written compositions. Readers must have at least a bachelor's degree in English, journalism, education, or a related field. In addition, they must have teaching experience or other qualifying experience with children or in the assessment of writing. (TEAMS Instructional Strategies Guide: Written Composition, p. 13)

The assessment is called **focused holistic scoring** "because the criterion for assigning the score is the degree to which the paper organizes ideas for a specified purpose and audience task." (TEAMS Instructional Strategies Guide: Written Composition, p. 13) To ensure uniform scoring, detailed scoring guides are prepared for every set of papers.

> "0" papers, for example, do not address the task at all. "1" papers attempt to address the task, but fail to do so acceptably. "2" papers are minimally successful, "3" papers moderately so,and "4" papers successful. Specific characteristics of papers at each scoring level are drawn up in detail, so that there can be as little doubt as possible what score a given paper should receive. (TEAMS Instructional Strategies Guide: Written Composition, p. 13)

The directions for scoring include specific sections on spelling, writing dialogue, punctuation, run-on sentences, writing stories, and the constraints of the testing situations. Although each of these contributes to the score, the section on spelling is most pertinent to the present study.

> Students progress through various stages as they learn to spell the thousands of words in their oral vocabularies. Some children progress more slowly than others through the stages of learning to spell. Some papers contain numerous errors while others are virtually free of errors. Individuals who have worked with elementary school students generally find it easy to decipher nonstandard spellings, since attempts to spell usually reflect both an underlying knowledge of phonology and knowledge of what

> has been learned up to this point about the orthographic representation of sounds. The students' scores will be affected if numerous errors in basic, familiar words occur. However, if nonstandard spelling of more complex words can be deciphered and the flow of words makes sense to the reader, the score on the written composition will not be affected. (TEAMS Instructional Strategies Guide: Written Composition, p. 17)

Prompts in the TEAMS assessment are picture prompts with directions in the lower grades. In grade seven and nine some prompts do not include pictures, but if the prompt is aimed at descriptive writing, a picture is always included. The pictures often include animals in make-believe situations.

The samples used in this study are prompts used in the tryouts that were not selected for statewide administration. These prompts were not used for statewide testing because of problems such as students having been confused by the wording of the prompt, excessive responses in the wrong mode, confusion about some feature of a picture, the responses not replicating the distribution for the "base" test items, and boredom on the part of readers who encountered very similar responses to a particular prompt. As responses to prompts that were rejected after pilot testing, the samples included in this study may be somewhat shorter and less elaborate than samples written in response to those prompts later used statewide. Since these were tryouts, reliability of the scoring was not calculated. In the actual test, three readers score each paper. They must agree absolutely and they maintain a reliability of .99.

PROCEDURES

Researchers employed a consistent procedure with all of the writing samples, regardless of their source or when they were received. Once the writing samples arrived at the Center for Reading and Language Study, Indiana University, the samples were assigned code numbers representing the origin of the samples and a unique essay identification number. All essays were read and coded for misspellings by three different readers, and one in every ten was read a fourth time for reliability. The first reader circled as many misspellings as the he or she could find, and wrote the conventional spelling near each circled misspelling. The second reader then reread the essays and marked any misspellings the first reader might have missed and checked to make sure the second reader concurred about the misspellings which the first reader marked.

After first and second readings were completed, the essays were given to typists also trained in proofreading for misspellings according to the conventions of this study. Into a computer program produced for this study, the typists entered the complete essays with misspellings and their corrected versions specially coded. Typists proofread the essays as they typed, providing a third reading focused upon letter-by-letter reconstruction in addition to the meaning-focused work of the first and second proofreaders. First, second, and third readers used color-coded pens and initialled the essays they read so the paper copies of the essays reflected all of the decisions made by each coder.

Most samples were read from the original giving the readers the best possible chance of deciphering the subtleties of the students' handwriting. But provision was made for essays which could not be read because of indecipherable handwriting. If neither the first nor the second reader could decode a significant section of an essay, the essay was removed from its packet and from the study, and the essay was placed in a collection which came to be called the "Black Hole." These unreadable samples were surprisingly rare; in fact only 42 samples were removed from the study. If only a few words were indecipherable, the unreadable words were specially coded and the essays were retained in the study.

Occasionally, the 30 lines of type allowed by our computer program were not enough to enter a long essay in its entirety. In this case, typists entered all they could, marked the essays at the point where they stopped, and recorded in a specified log the approximate number of lines of text which had been truncated. (Only 601 samples were truncated, at an average of 5 lines of text per sample.)

Also, in a couple of instances, we received more writing samples from a particular source at particular grade levels than were required for the study. In those cases, samples were selected randomly within the sample set and grade level for inclusion in the study.

After each packet of writing samples had been read three times and typed, researchers reviewed the work to ensure high reliability. Every change made by a third reader/typist was reviewed for accuracy and at least one in every ten essays was proofread for accuracy in typing as well as coding of misspellings. Satisfactorily entered writing samples were combined from diskettes into larger data files,

transformed into a suitable format for entry into the main-frame computer data base and entered into files on the main-frame for analysis.

The vocabulary lists were produced using a combination of programs including the Oxford Concordance Program and programs written specifically for this study. These lists were then proofread by two readers to remove remaining typographical and coding errors. Finally, these "cleaned" lists were run through the final programs to produce the statistics and list in final form for publication.

Personnel

The directors of this study were two doctoral candidates in Language Education at Indiana University. They designed the procedure, trained personnel, closely supervised the work, and managed the data transformations and transfers. Personnel included 14 proofreaders and 11 typists, all carefully trained by the directors to achieve a high degree of consistency in marking. Employment of proofreaders and typists was based on interviews and a proofreading test. During initial training, proofreaders and typists practiced proofreading on training samples aimed at clarifying the study's decisions about handwriting, usage, and misspellings. After training, proofreaders' work was closely monitored. Proofreaders were expected to miss no more than five marking errors in every 20 essays. A few readers who did not meet the study's standards were dismissed early in the study, but the remaining proofreaders and typists were not only highly skilled, but also diligent and conscientious about the quality of their work.

Resources used by the proofreaders and typists included a shelf of dictionaries, a running chalkboard list of words the proofreaders frequently needed to look up and

clarifications on marking, periodic memos from the directors, and collaborative consultation with the directors and eventually with other readers and typists. The team of proofreaders and typists remained relatively stable with very little turnover after the initial training period. Much effort was made to employ especially well-qualified and highly motivated readers and typists, and the effort resulted in an especially high quality group of workers, who not only marked or typed the writing samples with special care, but also made suggestions for analysis from their observations of the features of the student writing samples.

It was clear that the readers and typists had gathered many impressions from the thousands of writing samples that they read over the course of the study. Debriefing sessions allowed discussion of the readers' and typists' observations about children's writing, prompts and the testing process, patterns of spelling errors, and how the project affected their own spelling and writing.

DEFINITIONS: What Is a Spelling Error?

What is a spelling error? On the surface the question seems a simple one. If the appropriate letters are not used, the word is misspelled. However, there are all kinds of words that are not misspelled, but the wrong word is used. Are these spelling errors or errors of usage? Is the incorrect use of apostrophes a spelling error? Is the incorrect formation of a plural or an abbreviation an error? There does not seem to be any simple principle to follow in declaring that a particular word choice is a misspelling. Correct spelling is embedded in a larger question of writing conventions and correctness. It's not easy to draw a definitive line where spelling ends and other conventions begin. For example, the many differences in dialects within the United States pose a challenge in relation to verb forms and inflected endings. It was decided that such verb forms and endings would not constitute spelling errors within this study since it would not be possible to check and see whether what was written simply amounted to the writer's speech written down, or a spelling error.

Initial definitions of what would be considered errors in the study included the following:

1. The misapplication of spelling rules such as RECIEVE for RECEIVE.
2. The addition or deletion of letters that seem to be based on mispronunciation of the word such as CHILADREN for CHILDREN and LIBARY for LIBRARY.
3. The substitution of an alternative letter for the correct letter (probably based on irregularity of spelling of the word or misapplication of a phoneme/grapheme relationship) such as FITE for FIGHT, FOND for FOUND, and INSTED for INSTEAD.
4. Reversals of letters such as WAS for SAW, DEE for BEE, and BECUASE for BECAUSE.

5. Homonyms not used correctly such as DEER for DEAR and HIM for HYMN.

6. Inappropriate substitution of a real word for another real word, as in THE for THEY. Such substitutions will be coded as errors only when proofreaders are certain from context that the writer intended another word.

7. One word written as two words such as ALOT for A LOT and INCHARGE for IN CHARGE. Also two words written as one, as is EVENTHOUGH for EVEN THOUGH.

8. Possessives not formed correctly such as MARYS' for MARY'S and CHILDRENS' for CHILDREN'S. (Note that when the writer does not attempt a possessive form where one belongs, we do not consider the error one of spelling, as in IT WAS MARY DOG.)

9. Unnecessary apostrophes in verbs, as in SIT'S and RUN'S.

10. Omission of colon in time and abbreviations which are not correctly formed such as MISS. or MR.S or ETC are errors, other punctuation errors are not.

11. Contractions formed incorrectly such as DO'NT for DON'T and ITS for IT'S.

12. Plurals not formed correctly such as BABYS for BABIES and SHEEPS for SHEEP.

13. Invented words or spellings of superlatives as in the errors: He is the BESTEST player on the team. Bill is GOODER at swimming than running. Other examples: UPRAGE for uproar or outrage; SPLURTED for splattered and blurted; EXPENDID for expelled and suspended.

14. Word forms that are not actually words (non-words) such as he RUNNED to the store or he hurt HISSELF.

15. A word that cannot be determined at all will be considered a misspelling and the closest approximation of the word will be considered the word attempted.

USAGE ERRORS THAT WILL NOT BE CONSIDERED SPELLING ERRORS INCLUDE:

1. Verb agreements will not count as errors. For example the following underlined words would not be considered errors: He RUN to the store. They RUNS across the street. However, a verb error which produces a word which is not an actual word will be considered a spelling error. For example, He RUNNED to the store.

2. Incorrect usage of comparative and superlative adjectives will not be considered errors. For example in the following sentences, the underlined words will not be considered errors: That book is the BEST of the two I have read. He is the BETTER student in the whole class. However, the formation of non-words will be considered errors. The following underlined words will be considered spelling errors: He is the bestest player on the team. Bill is GOODER at swimming than running.

3. Incorrect pronoun usage will not be considered spelling errors. For example, the following underlined words will not be considered spelling errors: He went with WHO. HER is a good swimmer. However, the formation of non-words will be considered spelling errors. For example, He hurt HISSELF when he fell.

4. Proper nouns for which there is uncertainty about variant spellings, as in personal names, will not be considered errors.

As the study proceeded and the proofreaders were informed by the first thousands of reading, some other issues emerged to be resolved with conventions for use within the study. An early problem centered on the general decision not to consider spellings of proper nouns as errors. We began making decisions about proper nouns by using the standard of categories: if the word in question were in the category of personal noun, we said it would not be recorded as an error. If the word were a place name, we said we would record the error, and so on. Questionable categories which we struggled with were names for baseball teams and product names. Some early examples were PHILLIES=PHILIES, ANGELS=ENGLES, YANKEES=YANKIES, ADIDAS=ADDIDAS, SPRITE=SPRUTE, CANDYLAND=CANDY LAND.

As we discussed these decisions, we considered another standard against which to decide. Instead of using the previous categories, readers could think of all the proper nouns in terms of the reader's degree of certainty that the writer intended a

word which has a definite, conventional spelling. So, in the case of the spelling of somebody's name, the degree of certainty is very low because the reader cannot be aware of possible variations of the spelling of a personal name. In the case of a well-known place name the reader can be very confident because there is a single conventionally accepted spelling. When this certainty standard is applied to the names of baseball teams, a reader could be fairly certain that those major league team names were legitimate misspellings because the conventional spellings of the teams is well established. Invented or local team names, however, might not be counted as errors if readers were relatively uncertain of the conventional spelling. For example, in one sample the writer wrote that the "U" on the shirt of a boy in the prompt "stands for Usners." The reader could be confident that the writer was inventing a team name or she/he is relating a name for which there is either an uncertain convention or no convention at all, so it's easy to decide not to record "Usners" as a spelling error. The same system seems to apply well to product names.

This certainty standard served well in the course of the study. It was not, however, without ambiguities. The local name could easily be a variation on a national team. Although the writer who wrote "Rebells" may have meant "Rebels", s/he may also have been playing with language and making up another new name. (This is from the same essay with the "Usners", a clearly made-up name as the other team.) Product or trade names seem less ambiguous, but they could also stem from the writer's experience with products that the proofreaders are not familiar with and which are deliberately named to be similar to a more famous product. The general decision was to code as errors which the readers were sure of with the convention CORRECT

WORD = INCORRECT WORD. If the reader was only about 50% sure the coding reflected that doubt: CORRECT WORD = = INCORRECT WORD. So we added:

16. Misspelling of proper nouns for which there is a certain conventional spelling, such as YANKIES for YANKEES or SPRUTE for SPRITE or INDEANA for INDIANA.

Another source of ambiguity in spelling had to do with the differences in opinions represented in the various dictionaries. This lack of agreement has been well documented. (Emery, 1973) This meant that deciding what was an error was *not* as easy as looking it up in the dictionary. In the National Assessment of Educational Progress, words which were spelled in different ways in different dictionaries were not marked as errors. In this study, the proofreaders attempted to find a consensus on the spelling convention for a particular word. Researching those conventions sometimes stretched out over a period of time. An example of such a search revolved around what seems to be the last (only?) gender marker in English, the words BLOND AND BLONDE. BLONDE is used to describe females, and BLOND can be males or mixed males and females. In some of the samples it was clear that this convention was being taught, but others did not make the distinction at all. Rather than having a large staff facing these questions over and over, resolutions to these questions were listed on the chalkboards and written up in memos to the entire staff. A part of that list follows:

COMMON PROBLEMS:

NOT AN ERROR:	ERROR:
okay, OK, O.K.	Ckey
all right, alright	allright
okey dokey	
bye (for 'bye)	

The Dictionaries used in the project included *The American Heritage Dictionary*, Second College Edition, 1985; *Webster's New World Dictionary of the American Language*, Second College Edition, 1972; *Webster's New Twentieth Century Dictionary of the English Language*, Second Edition, 1979; *12,000 Words: A Supplement to Webster's Third New International Dictionary*, 1986; *Webster's Third New International Dictionary*, 1986; *Webster's New International Dictionary*, Second Edition, 1957.

Other research revolved around questions which straddle the boundaries of usage and spelling. The appropriate use of hyphens involves both spelling and how the word is being used. John Benbow in the Oxford University stylebook offers the following caution, "If you take hyphens seriously, you will surely go mad." This certainly expresses the difficulties proofreaders found as they tried to make decisions about the appropriate use of hyphens. For instance, we found that TIME-OUT, TIME OUT, AND TIMEOUT were all listed in the dictionaries. These were the hyphen conventions for the study:

17. a. Begin with the dictionaries to determine if a word is a compound or is hyphenated. Dictionaries will show compound nouns and some adjectives.

    b. For the adjectives that are not listed, decide if the sentence will be ambiguous without the hyphen.

    c. With a number, the hyphen is required if the number forms a compound with another modifier like TEN-FOOT pole.

The use of hyphens to show correct splitting of words at the ends of lines was almost totally excluded from the study because that is an issue of syllabification and not of spelling per se. Many of the writers showed, particularly in the earlier grades, that they had not yet mastered that convention. The one sub-category of hyphenation

at the ends of lines that was retained was the splitting of compound words if that compound word was not written as one word elsewhere in the sample. This allowed a count of all the difficulties with compound words without penalizing a writer who knew the word was compound but didn't know to include the hyphen at the end of the line.

Decisions about the use of real words for real words posed a particular challenge to the proofreaders. The proofreaders were very competent users of standard academic English. They wanted to "fix" the instances which didn't "sound right" to their ears. For example, if the writer had SET in a context of SIT, it was to be left unmarked. Since we could not ask the writer for confirmation we had to assume that she/he might have meant SET; thus, it might not be a spelling error. The same applies to the use of LAY and LIE which remained unmarked as errors for this study. In another real word to real word substitution, we found writers often left off what seemed to be letters required for the context. Often a writer would start a sentence with THE in a context that clearly required THEY. Another area of difficulty was AN/A in a context that implies AND. These were marked as an error if the proofreaders were sure from the context that another real word was the writer's intention.

Many of the prompts elicited stories from the writers. Since they were writing stories they often included quotations from characters and attempts to represent the sound of speech and other sound effects. This required further decisions about conventions for this study.

18. Quotations and Speech written down:
    a. Decide if the writer is signalling a quotation from the context; don't depend on proper use of punctuation for quotations.

b. When the writer is clearly trying to represent speech like WAAAAAA TA GOOOOOOO! Do not mark as an error.

c. Other words within the speech though should be corrected, so if the student has written "Whammm, WICH way did it go?, correct the WICH.

Verb forms provided many questions during the proofreading. As in the case of words like LAY and LIE, the proofreaders were tempted to mark such words as: WOULD HAVE, WOULD OF, SHOULD HAVE, SHOULD OF, SUPPOSED TO, SUPPOSE TO. Following the earlier decisions, and not being able to check the intention of the writers, these words were not marked as errors. With non-word verb forms, the decision was made to correct to the word with the closest real word for that meaning.

19. non-word verb forms:

When the writer has used a non-word, replace it with the closest equivalent chosen on the basis of meaning rather than grapho-phonic (letters and sounds) similarity. For example, "Emily BRANG her favorite bat to practice." The preferred correction here would be BROUGHT (closer in meaning) rather than BRING (closer in letters to BRANG).

In general we chose to mark words, that we could see them for the analysis rather than leave them unmarked. Regardless of all the decisions which have been made, all the data remains in the data base, and it could be analyzed using different criteria.

LIMITATIONS

During the course of the study some limitations have emerged. Some are related to the writing process itself. Others relate to the specific kinds of writing that are the basis of the study and the conditions under which the writing took place.

Writing Samples

The writing samples were gathered from children across the country and were analyzed as products. Consequently, though there were many occasions when the readers wished they could ask the writer what was meant in a passage, they were unable to directly question the writer about his/her intentions. For example, readers did not know whether the unusual names (ANGLEA) were meant to be familiar, more conventionally spelled names, were made up by the writer and were "correct" as we read them, or were, though unusual, actual names.

Another complication of not being able to talk to the writers stems from the many ambiguities posed by handwriting. While some of the writing is clear and fairly easy to interpret, other samples pose substantial difficulty in judging whether the writer actually misspelled a word or whether we have simply misinterpreted the letters based on poor formation. For example, TEAM written TEUM so the "a" looks potentially like a "u." If this is marked as an error then we will be looking at a convention that "eu" meant long "e" to this writer, which may or may not have been the case at all. Initially, we marked handwriting-related errors more frequently, but in the later part of the study chose to err on the side of under marking to avoid analyzing products of handwriting as if they were intended spellings.

The proofreaders noted that reading the children's handwriting brought back memories of strategies they had used themselves to protect themselves from making errors: making a word illegible if uncertain of the correct spelling, using ambiguously spaced apostrophes in contractions and possessives, marking faint hyphens and apostrophes when uncertain if they are needed, and, when unclear whether a compound should be one word or hyphenated, placing it at the end of the line so that it can safely be hyphenated. All of these strategies have resulted in decisions which may or may not reflect what the writer knew about spelling and usage.

Finally, we faced a few limitations in accurately representing the writing samples in our data format. Because of excessive length, we had to truncate 601 samples at an average of 5 lines per essay. Also, we had no way to represent the flexibility of handwriting as we enter the essays into the computer. We did not have a way to show the mirror image B's or the quadruple hump M's. We could not show the emphasis created by words written four times as large as the rest of the text. Although we were only infrequently hampered by the constraints of the keyboard, we cannot say our data represent every orthographic intention in every writing sample.

Assessments

The samples in this study were all written in response to carefully designed prompts and gathered under test conditions. This kind of writing adds strengths to the study which were discussed earlier, including: inclusion of students who were very good, average and poor writers; samples collected from students under known conditions and scored by a system for which there was known reliability and validity; student writers motivated by the test conditions to do their best writing within the time

limitations; samples from students across the country in a nationally representative sample. However, as assessment writing, it does not include informal writing such as the "friendly letter" of earlier studies, or other writing which takes place outside of school.

The writing tasks for most writing assessments in the elementary schools are either descriptive or narrative. One limitation of using only narrative and descriptive writing is that it may limit the vocabulary that the students use in their writing. Nevertheless, narrative and descriptive writing is the type most often assigned in the schools and, therefore, represents the vocabulary most often used by students in the schools. According to Jongsma and Dean, " Some researchers... have found that writing performance varies significantly across modes of discourse." (1988, p. 10) Most of the writing included in this study is either descriptive or narrative, so there may be some limitation of vocabulary.

Additionally, the words that the students chose were prompted by the task. These prompt-specific words might well not appear in the frequency that they do here if those prompts had not been the basis of the writing. For example, a picture prompt which shows children in a classroom requires the use of such words as TEACHER, PRINCIPAL, PRINCIPAL'S, DETENTION, CLASSROOM. Similarly, a baseball game prompt elicited especially high frequencies of baseball-specific language such as BASEBALL, PITCHER, HOME RUN, BASE. These words would not be unusual for children to write, but may be more heavily represented because the writing was in response to that prompt. This limits the ability to generalize the frequencies of the vocabulary most elicited by the writing task. Because of the topic constraint, children

also may have been limited in the use of other words which might have occurred had they been writing on subjects of their own choosing. These concerns are partially compensated for, however, by our interest in percentages of misspellings of particular words. We are less interested, for example, in the frequency of the word BASEBALL than in the proportion of instances students misspelled the word.

A limitation which has emerged as a result of our initial analysis with the word lists is the function of presenting words individually, except when a phrase is coded as a misspelling. The result is to separate compound words into two separate words represented in separate places on the word list. Our list does not indicate, for instance, how often our words BALL and GAME might represent the compound BALL GAME. While our list tells us how frequently we found the misspelling ALOT, it does not indicate how often A LOT was spelled correctly, for the A will be included with the thousands of examples of A as an article. This is a limitation of the list approach alone, for since we preserved the essays in their entireties, further computer manipulation will allow us to search for and analyze phrase constructions. Although obvious names have been marked on the list, there are several words which can be other nouns (RAY, BILL, PEARL). These have been left unmarked on the list, but they could be categorized by searching the data base for the context of each use.

Although this study is distinguished for its use of samples which have all been scored holistically, there are differences in the scoring protocols among some of the various assessments. This has been partially accounted for by assigning standard score equivalents for the assessments. But there remain differences in weighting in the scoring. For instance and most germane to this study, in the Texas TEAMS

assessment, a greater emphasis was placed on spelling accuracy than in the MAT. Thus, samples with poorer spelling received lower scores on the TEAMS than they might have on the MAT. However, all scoring procedures indicated an expectation of reasonably conventional spelling as a trait of good writing.

The Texas teams samples were all from try outs of potential prompts for use in the statewide assessments. The ones included in this study are all "failed" prompts, ones which were not chosen for use with the whole population. The reasons for not using them vary from the prompt not eliciting the kind of writing that was expected, generating too predictable responses which would be boring for the readers in large numbers of a statewide assessment, and finding distributions of scores which did not match the target distributions. These samples may be shorter and less elaborated than students might have written for one of the accepted prompts.

The writers had different amounts of time to write the samples, ranging from twenty minutes for the MAT (73% of total samples) to writing that spanned two days in the Rhode Island assessment (4.5% of total samples). Rhode Island students wrote a rough draft the first day and a final draft the second. This means that the samples are not directly comparable on the basis of the amount of time the students had to write. While this represents some limitation in generalization, it also will allow us to compare subsets of the data to determine the effect of writing time and the opportunity for revision on spelling.

Error Count

Each occurrence of a misspelling is counted as an error. So a writer whose only error was BASE BALL for BASEBALL eleven times in the essay would be counted as

having made eleven errors instead of one. Also the totals include these multiple occurrences without distinguishing whether eleven writers made the error once or one writer eleven times. The complete nature of our data base will make later analysis of this kind possible.

Time Span

The writing samples were written from the fall of 1984 to the fall of 1987, representing a three-year span of time. The sample represents that span of time rather than an single administration of an assessment.

Independent Samples

More than half of the samples in the study represent one sample from each writer. But among the MAT set there are five thousand writers who each wrote two samples. This allows for a comparison between two samples from one writer, but it also poses a limitation in terms of the independence of the samples. The total number of samples is 21,800. Of these, 17,300 were single efforts, and 4,500 were second samples from writers already present in the data base.

Socio-economic Status

The socio-economic categories of the samples in this study are based on designations for entire school districts. Since there is variation in socio-economic status within most school districts, this is a limitation of the precision of these categories. For example, New York City is all one code. These categories should be interpreted with this limitation in mind.

RESULTS

There are several products of the study. The first product is the set of seven word lists at grades two through eight.

Second Grade Word List
Third Grade Word List--Volume 1
Third Grade Word List--Volume 2
Fourth Grade Word List
Fifth Grade Word List
Sixth Grade Word List--Volume 1
Sixth Grade Word List--Volume 2
Seventh Grade Word List
Eighth Grade Word List

The lists contain the correct (or "root") words, the variations of spelling for those words in descending order of frequency, the percentage of the time that the words were spelled correctly and misspelled, what proportion that target word is of the total words used by students at that grade, the percentage each spelling variation represents of the total variations for that word, and a ranking that is a measure of the relative frequency of the words in the list. The ranking was derived from the number of times that the most frequent word was written (THE) divided into ten parts. THE has a rank of 1. Words which are written much less frequently have a frequency of 10. Very few words approach the frequency of the very frequent words such as THE (1), AND (4), TO (5). In comparison to these A falls to a rank of 7 and IT to a rank of 8. Most words have a ranking of 10.

A second product of the study is the *Spelling List Concordance: Words Taught in Seven Major Spelling Programs*. The words are divided into grades from 1 to 8 by the mean grade at which the words are introduced. Other data includes:

Grade Place Range: the range of grade placement

No. Agree: an agreement index (highest grade of use minus lowest grade of use); a lower number indicates greater agreement on when to teach a word while a higher number indicates less agreement

No. Using Word: the number of programs which teach the word

The list contains about 10,400 words. Of those words, one hundred twenty-four are found in all seven series.

Preliminary statistical description of the data in this study follows:

TABLE SEVEN: MEAN NUMBER OF WORDS BY GRADE LEVEL

Grade	Assessment	Number	S.D.	Range
Two	Metropolitan	88.4	49.0	322
Three	Metropolitan1	124.37	61.15	357
	Metropolitan2	117.14	57.22	392
	Texas	131.77	68.16	375
	Rhode Island	118.70	64.11	360
Four	Metropolitan	144.19	61.64	350
	Texas	132.78	67.81	364
	Connecticut	186.60	82.76	345
Five	Metropolitan	156.82	61.26	316
	Texas	155.53	63.31	330
Six	Metropolitan1	168.01	62.43	352
	Metropolitan2	161.96	63.79	353
	Georgia	218.00	77.38	382
Seven	Metropolitan1	167.32	65.44	321
	Texas	154.21	57.69	322
Eight	Metropolitan1	179.16	70.73	341
	Georgia	253.78	70.70	363

TABLE EIGHT: PERCENT OF WORDS SPELLED CORRECTLY BY GRADE LEVEL

Grade	Assessment	Mean	SD	Range
Two	Metropolitan1	87.53	9.19	75
Three	Metropolitan1	89.93	7.49	46
	Metropolitan2	90.98	6.91	67
	Texas	84.70	9.58	53
	Rhode Island	90.66	7.25	50
Four	Metropolitan1	92.22	5.79	43
	Texas	88.60	7.85	47
	Connecticut	92.84	5.57	30
Five	Metropolitan1	93.23	5.59	33
	Texas	91.69	4.96	46
Six	Metropolitan1	94.35	4.96	40
	Metropolitan2	94.80	4.16	41
	Georgia	95.15	3.88	32
Seven	Metropolitan1	94.67	4.24	35
	Texas	93.86	4.44	36
Eight	Metropolitan1	95.58	4.07	37
	Georgia	96.13	3.57	29

TABLE NINE: PERCENT CORRECT BY PROFICIENCY LEVELS

	LOW			LOWMID			HIGHMID			HIGH		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
GRADE 2												
MAT1	81.5	11.8	75	88.4	7.7	40	94.0	4.0	20	96.6	2.1	7
GRADE 3												
MAT 1	80.1	12.2	46	88.7	7.2	44	93.6	4.0	32	95.0	2.0	10
MAT 2	5.0	11.8	55	89.9	7.3	41	92.5	4.9	48	95.1	3.1	18
TEXAS	87.0	11.3	23	83.8	9.8	52	89.	75.6	32	93.1	5.4	22
R.I.	85.5	9.6	50	91.3	5.[illegible]	45	93.1	4.6	23	95.0	3.9	20
GRADE 4												
MAT 1	84.1	9.4	37	90.2	6.3	43	94.1	3.4	20	96.0	2.5	15
TEXAS	85.3	9.8	44	89.3	6.9	38	90.2	6.9	33	95.8	2.5	11
CONN.	89.8	7.6	28	92.3	5.4	30	93.8	5.3	21	94.6	3.8	20
GRADE 5												
MAT 1	86.4	8.7	33	91.8	6.0	3[illegible]	95.1	3.3	20	96.7	2.1	9
TEXAS	94.0			1.3	5.3	46	93.1	3.1	14	94.3	3.0	7
GRADE 6												
MAT 1	84.8	10.1	40	93.0	5.3	29	95.5	2.8	18	96.8	2.0	13
MAT 2	89.5	8.1	35	93.9	4.5	33	95.6	2.9	30	96.7	2.1	17
GEORGIA	----------NOT AVAILABLE----------											
GRADE 7												
MAT 1	89.4	8.8	35	93.7	4.6	24	95.1	3.2	19	96.67	2.2	13
TEXAS	92.0	3.1	9	93.1	4.7	36	95.6	3.1	19	96.7	3.1	12
GRADE 8												
MAT 1	87.0	9.3	37	94.1	4.7	25	96.1	2.9	15	97.3	1.8	11
GEORGIA	----------NOT AVAILABLE----------											

TABLE TEN: MEAN STANDARD SCORES BY GRADE LEVEL

Grade	Assessment	Mean	SD	Range
TWO	METROPOLITAN1	35.39	14.5	92
THREE	METROPOLITAN1	47.47	15.68	98
	METROPOLITAN2	50.4	16.12	98
	TEXAS	40.0	18.27	98
	RHODE ISLAND	42.5	23.53	98
FOUR	METROPOLITAN1	55.13	17.16	98
	TEXAS	31.73	21.93	98
	CONNECTICUT	54.27	25.26	98
FIVE	METROPOLITAN1	52.16	18.40	98
	TEXAS	48.50	17.75	85
SIX	METROPOLITAN1	57.60	19.87	98
	METROPOLITAN2	54.86	18.08	98
	GEORGIA	NOT AVAILABLE		
SEVEN	METROPOLITAN1	57.39	20.08	98
	TEXAS	52.42	16.26	98
EIGHT	METROPOLITAN1	62.36	20.87	98
	GEORGIA	NOT AVAILABLE		

TABLE ELEVEN: MEAN SCORES BY PROFICIENCY LEVELS AND GRADES

Grade	Proficiency Index			
	Low (0-25	Low-Medium (26-50)	Medium-High (51-75)	High (76-99)
TWO				
METRO1	20.7	69.7	8.8	0.8
THREE				
METRO1	5.2	67.3	27.4	4.7
METRO2	3.4	54.5	35.2	6.7
TEXAS	0.6	85.8	10.9	2.6
R.I.	23.2	49.1	16.6	11.1
FOUR				
METRO1	3.24	6.6	36.8	13.3
TEXAS	20.6	59.7	8.2	3.1
CONN.	13.1	41.2	20.9	24.8
FIVE				
METRO1	5.2	47.9	35.7	11.1
TEXAS	0.3	79.1	16.5	4.0
SIX				
METRO1	4.0	39.1	36.4	20.5
METRO2	3.4	44.1	38.8	13.7
GEORGIA		Not available		
SEVEN				
METRO1	3.5	38.8	37.0	20.6
TEXAS	0.9	70.7	25.7	2.7
EIGHT				
METRO1	2.6	31.5	36.8	29.2
GEORGIA		Not available		

SPELLING AVOIDANCE PILOT STUDY

Rationale

Extensive lists of words have been developed for teaching spelling, some based on reading vocabularies, others on the writings of both adults and children. There is an interest in current research and in spelling series to base programs on actual student writing. However, many adults and students report that when they write they sometimes substitute words when they are uncertain about spelling in order to avoid embarrassment. For this reason, when studies of spelling are derived from writing samples, we must examine not only what has been written, but also what has been avoided.

To investigate spelling avoidance, not in the context of the weekly spelling exam but in the context of composing stories, we initiated a study to explore the extent to which students avoid writing words because they are unsure of how to spell them. The degree to which students avoid using words they are not sure how to spell is the degree to which student writing samples which serve as the basis for spelling programs may be misleading.

Method

In order to determine whether students avoid using words because of spelling, and under what conditions they avoid words, students were asked both to tell and to write the same story in response to a picture prompt. Student participants in the study were 27 students from grades 1 - 5 in rural schools near a major university town in the Midwest. We selected a prompt from a national writing assessment program which

allowed a comparison of results to another body of research. We expected the prompt to be comparatively open and generative, and we used the same prompt for all students for comparison of subject matter and vocabulary choice across students and grade levels.

We employed a counter-balanced design with one-half of the students telling their story first and the other half writing first to control the effects of order on the stories. Students took a break of 15 to 45 minutes between telling and writing tasks. Research sessions were conducted by college students enrolled in a Reading Methods course. Researchers were responsible for conducting and reporting their experience with one of the elementary school students, usually a child with whom the researcher was already acquainted. All researchers followed the same protocol. (See Appendix A for the complete protocol.) The instructions given to students were:

> "I am interested in the differences between the language you use when you tell a story and when you write a story. So I am going to ask you to write/tell a story for me now. I want you to use this picture for writing your story. Please tell/write your story now." [Later] "I really liked your story that you wrote/told earlier. Now I would like you to write/tell that same story."

If the participants asked for help, they were told to do the best they could. When the participants finished writing, they were asked for self reports about their writing and spelling. Participants were asked to underline any words which they weren't sure about, or were not satisfied with, or had questions about. They were also asked whether there was anything else they wanted to write but did not, and whether they had thought about how to spell some words. Other interview data included:

-- both participant and teacher estimates of the participants' reading and writing proficiencies and general academic performance,

-- the type and amount of writing and spelling instruction and practice,

-- the participants' self-reported spelling competence and strategies,

-- the teachers' suggested spelling strategies,
a description of the research setting and participants' attitudes and behavior,

-- the number and kinds of words used or avoided in writing and/or telling.

These constitute the context of the writing and telling, as well as the basis of results.

Results

Some of the questions with which we undertook the spelling avoidance study include:

-- Do students avoid writing words because they are unsure of spelling? If so, to what degree do students avoid writing words because of spelling?

-- Are there discernable patterns in the words students avoid?

-- Do students with strict spelling teachers avoid words more than students whose teachers do not emphasize spelling correctness?

-- Are there relationships between spelling instruction and student spelling strategies, and do some kinds of spelling instruction promote spelling avoidance more than others?

-- Does the amount of time spent on spelling lessons and on writing, especially continuous writing, have an impact on spelling avoidance?

Unfortunately, our initial study of spelling avoidance does not provide us with the information we hoped for. We did find that students occasionally substituted words when changing from one form of expression to another. A few examples look potentially convincing:

Used in Telling	Used in Writing
Doberman	dog
44 Colt Magnum	gun
forest	woods

But any degree of clarity in the students' substitution is rare. More typically the substitutions look like this:

Telling	Writing
tired	hungry
gold	treasure
boy	son
walked	went
mother	mom
kingdom	village
shouted	asked
replied	said
wanted	decided

Most of the time, however, there was no evidence of direct word or phrase substitution at all. Students used many more words in telling [2,842; average 109] than in writing [1,827; average 70]. Only four students in 27 wrote more words than they told. But there are many plausible reasons besides spelling avoidance which might explain the difference between telling and writing word frequencies. Several students reported disliking to write. They reported writing to be hard because there were too many things to think about, including neatness, the difficulty of handwriting, punctuation, capitalization, and other technical aspects of writing in addition to spelling. Although only one-half of the student participants [13] told their story first, all but four students provided more elaborate and detailed versions of their stories in the telling mode, apparently preferring to condense their ideas into as little writing as possible.

Several of the students left out whole sections of their stories when writing. Most deletions and changes appeared either to be the result of reluctance to write or else the result of the student forgetting the details of his/her earlier version.

Limitations

Chief among the limitations of the spelling avoidance study is that we cannot determine what words, if any, that students avoided. All we can say is that certain words in one mode or another were not used. Only one student reported avoiding words at any time because of spelling, and in this study, no students reported avoiding words because of spelling. When students were asked to underline parts of their writing they were uncertain about, few paid attention to spelling.

Spelling was a concern to some students, for they asked researchers how to spell certain words, but no student appeared especially concerned with spelling in particular. One student responded that when the researcher said in request for help with spelling "Do your best," it "took the pressure off" of spelling concern and may have changed the student's spelling avoidance behavior. In fact, two of our researchers followed up with their students by testing them on words the students had misspelled in their writing samples. One student was able to correctly spell half of the words he/she had misspelled, and the other student missed only one of ten words he/she had misspelled in the writing, suggesting that these students at least were not attending much to spelling while writing and were therefore unlikely to have been avoiding the use of words because of spelling.

Another limitation to the study arose when a researcher thought she observed a student *limit* the number of times some words were used because the student had

been struggling with them all week. Although the student used the words in her story, the same words appeared more frequently in the oral version, suggesting some written avoidance even though the words appear on our list. The list approach does not accommodate this difficulty in establishing spelling avoidance, and it demonstrates a growing difficulty in defining avoidance.

The study may also be limited by the artificial circumstance in which we asked students to write. Many students demonstrated extreme difficulty in knowing what to do with our standardized picture prompt whether they started in the telling or writing mode. Many squirmed and halted and needed coaxing to continue. Curiously, responses from our students were generally much shorter than the responses to the same prompt in nationwide testing.

Many students were also reluctant to tell or write the same story over again. Although two students said they were glad to write the same story they had told because that would make the writing easier, several others were apparently bored with the task or else they demonstrated concern that they would not be able to remember exactly what they had told or written before. One student indicated that although she liked writing stories, she didn't like writing for the researcher rather than for herself.

Although we entered this pilot study with some confidence in the design, the results have made clear its shortcomings. There appears to be less avoidance of words because of spelling that we suspected, but our study only hints this is so. The conclusion of this study leaves us with many unanswered questions about spelling avoidance, and with a lesson about the limitations of our approach. The complexity of

the data leave us with two intriguing questions. What exactly *is* spelling avoidance, and how can we accurately observe it?

APPENDIX A: SPELLING AVOIDANCE STUDY PROTOCOL

BEFORE YOU GO:

1. Read everything in this packet, especially the things to say to the student during the writing and telling experiences.

2. Test your tape recorder before going to the school, check it again at school, and if your recorder uses batteries, install new ones so there is no chance you will miss recording part of your interview because of mechanical failure. Also consider surrounding noise when selecting a location for your interview because too much background noise may make your tape very hard to transcribe.

3. Write a statement describing what you expect will happen during the encounter with the student, how you think the student will probably respond to the task, and what information or ideas you think you might gain from this experience.

PROCEDURE FOR GETTING READY AT THE SCHOOL:

Clear with the teacher your choice of respondent. Secure the agreement of the student.

Establish with the teacher when you will do the study. Arrange a time when she will be able to answer a few questions about how your student is doing in school, as a reader and writer. Be sure that the student can tell a story, take a break of some kind from the study, and then write the story. Make sure the break between telling and writing is no shorter than 15 minutes and no longer than 1 hour.

Establish with the teacher where you can do your interview. Try to use a place which will be relatively quiet and free of interruption. This might be a part of a classroom,

another room, the library, or even a quiet hallway. You will need a desk or table for writing.

Meet with the student. You may complete the instruction history from the teacher's point of view either before or after the telling/writing. You will be gathering this information twice, once from the teacher's point of view and once from the student's. It's important that you do not ask the student these questions until after the story telling/writing sessions to avoid focusing the student's attention on spelling.

Special Note: As you collect information, please remember that several people need to read the data. Please write as carefully as possible given the pressures of the situation and go over the sheets soon after the interviews to add missing words, write out abbreviations, make generally them legible.

TEACHER INFORMATION FORM

BACKGROUND INFORMATION ABOUT THE STUDENT

Teacher's Name:_____

Interviewer's Name:_____

Student's Name:_____

School:_____

Student's Grade:_____ Observation Date:_____ Sex: M F

General Academic Performance: high medium low (circle one)

_____

_____

Reading Performance: high medium low (circle one)

_____

_____

Writing Performance: high medium low (circle one)

_____

_____

_____

How did you arrive at these estimates?

_____

_____

Average number of classroom hours per week spent on writing. Writing includes all forms of writing from stories to workbook pages, to practicing spelling words:

_____

How much time per week is spent on continuous writing such as stories, reports, letters, journals, etc.?

_____

_____

_____

What approaches does the teacher suggest when students ask for help with their spelling while they're writing? When and how much does spelling count in students' grades, evaluation? When do you (if ever) mark misspelled words?

_____

_____

_____

_____

What is the teacher's approach to teaching spelling? Is it integrated in writing instruction, does she use a series (if so, which one)? What does she think is most effective in helping students to learn to spell?

_____

_____

_____

_____

Ask if she knows what kind (series, approach) spelling instruction this student experienced in earlier grades:

_____

_____

OBSERVATIONS DURING DATA COLLECTION

Description of the environmental setting where the experience took place: (consider the type and size of room, any pertinent wall decorations or objects in the room, noise and distractions, social interaction, what others are doing, etc.)

Describe the student's apparent attitude toward you and about participating in the study:

Behavior noted during Story Telling (For instance is the rhythm of the telling stilted, word by word, painful or engaging, excited? Does the student squirm, look around as if for escape, distracted by other events?):

Behavior noted during Story Writing (Does writing seem painful? Does the student hold the pencil so tightly his/her knuckles are white? Does the student think, then write, does it just flow?):

_____

_____

BACKGROUND INFORMATION FROM THE STUDENT'S POINT OF VIEW

(student interview)

General Academic Performance: high medium low (circle one)

(Ask: What do you think of yourself as a student and how you do in school?)

_____

_____

_____

Reading Performance: high medium low (circle one)

(Ask: What do you think of yourself as a reader?)

_____

_____

_____

Writing Performance: high medium low (circle one)

(Ask: What do you think of yourself as a writer?)

_____

_____

_____

Ask: How did you decide about how well you're doing as a reader and writer and in school in general?

_____

_____

_____

Ask: Are you a good speller? (why or why not?)

_____

_____

_____

Average number of classroom hours per week spent on writing: (Ask: **How much time do you think you spend writing in class each day, each week? When I say writing I mean all kinds of writing from stories to filling in the blanks in your workbooks and practicing spelling words.**)

_____

_____

_____

What approaches does the teacher suggest for spelling? When and how much does spelling count? (Ask: **What does your teacher tell you to do when you're not sure how to spell a word? Is spelling very important to you? Is it very important to your teacher? Explain.**)

_____

_____

_____

Past spelling instruction, if known:

(Tell me about how you've learned to spell.)

PROCEDURE FOR GETTING THE SAMPLES (Writing first, Telling second):

Introduction for Student: "I am interested in the differences between the language you use when you tell a story and when you write a story. So I am going to ask you to write a story for me now."

"I want you to use this picture for writing your story. Please tell me your story now." When the student is finished, ask the student to read over what he/she has written and if there are any places he/she isn't sure about, or isn't quite satisfied with, or has questions about, ask them to underline those places. When this is done, ask the student what it is about those places in the story that made him/her underline them. If the student has not already answered the following questions, ask him/her: "Were there things you wanted to write but didn't? Why didn't you?" "While you were writing this story, did you ever think about how to spell some words?" Thank the student when you are finished. Be sure to record this.

It's possible that the student will say "I don't have any ideas." **(Take another look at the picture, you might imagine you're in the picture or what might happen next.)** or "Is this okay? **(Nod, I'm looking forward to reading your story.)** What do you want? **(I want you to write a story starting from this picture. How you tell it is up to you.)** Or the student may only write a little and then stop. When these prompts were originally administered, students had 20 minutes to write. Don't limit the time, but do encourage your student to speak/write fully.

AFTER A REST OF 15 TO 60 MINUTES:

Use the **same** prompt you used to get the first sample time.

"I really liked your story that you wrote earlier. Now I would like you to me that same story." Put the prompt on the table in front of the student again. Tape record his/her story, and take occasional field notes of behavioral observations which would not be recorded on tape. Try to handle the tape recording and notes in an unobtrusive manner. Responses are permissible if they are not judgmental or guiding. Take an interest in the story as the student tells his/her own story his/her own way. Nodding or other non-verbal accepting/encouraging may help. Thank the student when you are finished.

PROCEDURE FOR GETTING THE SAMPLES (Telling first/ writing second):

"I would like you to make up a story about this picture and tell it to me." Tape record his/her story, and take occasional field notes of behavioral observations which would not be recorded on tape. Try to handle the tape recording and notes in an unobtrusive manner. Responses are permissible if they are not judgmental or guiding. Take an interest in the story as the student tells his/her own story his/her own way. Nodding or other non-verbal accepting/encouraging may help. Thank the student when you are finished.

It's possible that the student will say "I don't have any ideas." (**Take another look at the picture, you might imagine you're in the pict re or what might happen next.**) or "Is this okay? (**Nod, I'm looking forward to reading your story.**) What do you want? (**I want you to write a story starting from this picture. How you tell it is up to you.**) Or the student may only write a little and then stop. When these prompts were originally administered, students had 20 minutes to write. Don't limit the time, but do encourage your student to speak/write fully.

AFTER A REST OF 15 TO 60 MINUTES:

Use the **same** prompt you used to get the first sample time.

"We'll use this same picture for writing your story. Please write your story now." When the student is finished, ask the student to read over what he/she has written and if there are any places he/she isn't sure about, or isn't quite satisfied with, or has questions about, ask them to underline those places. When this is done, ask the student what it is about those places in the story that made him/her underline them. If the student has not already answered the following questions, ask him/her: "Were there things you wanted to write but didn't? Why didn't you?" "While you were writing this story, did you ever think about how to spell some words?" Thank the student when you are finished.

REFERENCES

Applebee, Arthur; Langer, Judith; & Mullis, Ina. (1987) *Grammar,Punctuation, and Spelling: Controlling the Conventions of Written English at Ages 9, 13, and 17.* New Jersey: Educational Testing Service.

Bauer, N. (1916) *The New Orleans public school spelling list.* New Orleans: Honsell and Brothers.

Beaton, et al. (1987) *Implementing the New Design: The NAEP 1983-84 Technical Report.* Princeton: NAEP.

Carroll, B. Davies P., Richman, H. (1971) *The American Heritage Word Frequency Book.* Boston: Houghton Mifflin.

Dale, E. (1931) *A comparison of two word lists.* Educational Research Bulletin. 18, 484-488.

Dale, E. & Chall, J. (1948) *A formula for predicting readability: Instruction.* Educational Research Bulletin, 27, 37-56.

Dale, E. (1936) A basic sight vocabulary. *Elementary School Journal*, 36, 456-460.

Emery, Donald W. (1973) *Variant Spellings in Modern Dictionaries* Revised Edition. Urbana: National Council of Teachers of English.

Fitzgerald, James A. (1951) *A Basic Life Spelling Vocabulary.* Milwaukee: The Bruce Publishing Company.

Fowles, Mary & Young, Saundra. (April 1987) *ETS Report of the 1987 Third Grade Rhode Island Writing Assessment.*

Fry, E. (1957) Developing a word list for remedial reading. *Elementary English*, 34, 456-458.

Gates, A. I. (1935) *A reading vocabulary for the primary grades.* New York: Teachers College.

Gentry, J.R. (1988) *Spelling Connections.* Columbus, Ohio: Zaner-Bloser.

Georgia Grade 6 Writing Test: Scoring Manual. Georgia Department of Education.

Harris, A.J., & Jacobson, M. Basic vocabulary for beginning reading. *The Reading Teacher*, 1973, 26, 392-395.

Hagerty, Patricia. (1981) *Comparative Analysis of Selected High Frequency Words Found in Commercial Spelling Series and Misspelled in Students' Writing to a Standard Measure of Word Frequency.* ED 218 608

Hillerich, R. L. *Analysis of words used in creating writing, grades1-6.* Unpublished study, Glenview, Illinois, 1966.

Hogan, Farr, Prescott, & Balow. (1987) *Writing Test: Norms and Technical Information Booklet.* The Psychological Corporation.

Horn, Ernest. *A Basic Writing Vocabulary: 10,000 most common words used in writing.* Iowa City: College of Education, University of Iowa.

Jones, F. W., *Concrete investigation of the materials of English spelling.* Unpublished doctoral dissertation, University of South Dakota, 1913.

Jones, L. V. & Wepman, J. M., *A spoken word count: Adults.* Chicago: Language Research Associates, 1966.

Jongsma, Eugene & Dean, Patricia. (1988) *Comparison of Holistic and Analytic Scores of the Scores of the MAT6 Writing Test.* Paper presented at the annual meeting of the National Council on Measurement in Education, April 6-8, 1988, New Orleans, Louisiana.

Kucera, H., & Francis, W.N. (1967) *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Moe, A. J. & Hopkins, C. J., *The speaking vocabularies of kindergarten, first-grade and second-grade children.* Paper presented at the Annual Meeting of the National Conference on Research in English, March, 1975. (ERIC Document Reproduction Service No. ED 105 465.)

Murphy, H., et al. *The spontaneous speaking vocabulary of children in the primary grades.* Boston University Journal of Education, 1957, 140, 1-105.

Norris, Norma. (1988) *National Assessment of Educational Progress 1984 Writing Mechanics Data Tapes User Guide.* New Jersey: Educational Testing Service.

Rinsland, H. D., *A basic vocabulary of elementary children.* New York: The Macmillan Company, 1945.

Sakiey, E. H. *Syllables: A weighted grapheme inventory.* Unpublished doctoral dissertation, Rutgers: The State University of New Jersey, 1977.

Smith, C.B. & Ingersoll, G.M. (1984) *Written Vocabulary of Elementary School Pupils, Ages 6-14.* Monographs in Teaching and Learning 6. Bloomington, Indiana: School of Education, Indiana University.

Sopko, Cindy. (1988) *An Investigation of Direct and Indirect Measures of Writing. Paper presented at the annual meeting of the National Council on Measurement in Education*, April 6-8, 1988, New Orleans, Louisiana.

"TEAMS: Instructional Strategies Guide: Written Composition, Grades 3, 5, 7, & 9." (August, 1986) Texas Education Agency, Austin.

Thorndike, E. L. *Teacher's word book.* New York: Teachers College, Columbia University, 1921.

Thorndike, E. L. *A teacher's word book of the 20,000 words found most frequently and widely in general reading for children and young people.* New York: Teachers College, Columbia University, 1931.

Thorndike, E.L. & Lorge, I. (1944) *The Teachers Word List of Book of 30,000 Words.* New York: Bureau of Publications, Teachers College.

Tidyman, W. F., *Survey of the writing vocabularies of public school children in Connecticut*, Teachers Leaflet No. 15, Washington, D.C.: U. S. Bureau of Education, 1921.

Wepman, J. M., *A spoken word count (children--ages 5, 6, and 7)*. Chicago: Language Research Associates, 1969.

Wiser, Berton. (1988) *Building a Developmental Scale for a Nationally Normed Writing Test.* Paper presented at the annual meeting of the National Council on Measurement in Education, April 6-8, 1988, New Orleans, Louisiana.

Writing Test Manual for Interpreting. (1987) The Psychological Corporation, HBJ, Inc.:San Antonio, Texas.