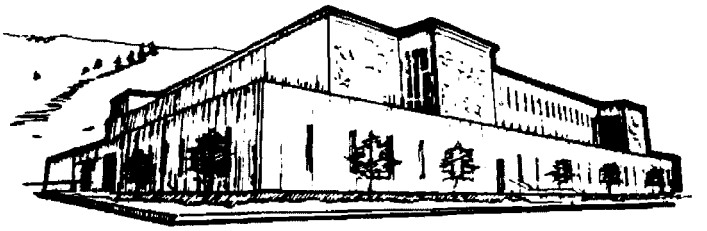
1990

Effects of regional dialect on the word recognition scores of children using the phonetically balanced kindergarten test and the nonsense syllable test

Michael Crews
The University of Montana



THE EFFECTS OF REGIONAL DIALECT ON THE WORD RECOGNITION SCORES OF CHILDREN USING THE PHONETICALLY BALANCED KINDERGARTEN TEST AND THE NONSENSE SYLLABLE TEST

By

Michael Crews

B.A., University of Montana, Missoula, Montana, 1988

Presented in partial fulfillment of the

requirements for the degree of

Masters of Arts

University of Montana

1990

Approved by:

Chairman, Board of Examiners

Dean, Graduate School

June 5, 1990

Date

Crews, Michael, M.A., June 1990 Communication Science Disorders

The Effects of Regional Dialect on the
Word Recognition Scores of Children Using
the Phonetically Balanced Kindergarten Test
and the Nonsense Syllable Test

Directors: Sally J. Johnson, M.A.

The effect of regional dialect on children's word recognition scores was investigated using a Montana and a General Southern dialect speaker presenting the Phonetically Balanced Kindergarten (PBK) word recognition test and the Nonsense Syllable Test (NST). Previous research indicated that dialect affected speech perception and production and that the reliability of word recognition test scores was affected by individual speaker differences in monitored live voice testing. It was hypothesized that children of a Montana dialect would obtain significantly different scores on a taped word recognition test when the speaker had a Montana dialect as opposed to a General Southern dialect. It was further hypothesized that the dialectal effects would be evident on the PBK word recognition scores, but not on the NST word recognition scores.

To test this hypothesis, twenty seven year old subjects with Montana dialects were administered PBK and NST lists presented by both speakers. The subjects' responses were recorded and judged as correct or incorrect by two trained listeners of Montana dialect. Intra- and inter-examiner reliability was measured using a point by point percentage of agreement. The data were analyzed using t-tests performed at the 0.05 level of significance.

Results showed that subjects performed better on the PBK test than on the NST test. Inter-examiner agreement was also better on the PBK test. Both hypotheses were disproved. The differences in scores on the PBK test with the two dialects was not significant statistically. The difference in performance on the NST was verified statistically with better performance when presented by the General Southern dialect speaker.

The implications of these results for word recognition testing were discussed and possible areas for future research were suggested.

ACKNOWLEDGEMENTS

I would like to thank the members of my thesis committee, Sally J. Johnson, M.A., Jesse G. Kennedy III, Ph.D., and Robert B. Hausmann, Ph.D. for their guidance and support. Special appreciation is given to Sally Johnson for her unending patience, encouragement, and constructive criticism on the early drafts of this manuscript.

I also wish to thank Jodie Egosque for her assistance with the reliability measures. I sincerely appreciate the twenty subjects and their parents for taking time out of their busy schedules to participate in this study.

I am also greatly appreciative of my family who gave me constant support and encouragement. I am deeply indebted to Jodi, my sister, who throughout this project as well as all my other endeavors, has been my most supportive, most loyal "fan" and to Jeff, my husband, who supported me and helped me see the lighter side of life.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The following study examined the effect of speaker dialect on the word recognition scores of children. Many studies have examined the effect of dialect on speech perception and production (Levy & Cook, 1973; Ralph, 1967; Tse & Ingram, 1987; and Taylor & Payne, 1983). Although this research has shown evidence to suggest that dialectal differences can negatively affect children's performance in school and on language tests, little research has followed to determine what kind of effects a speaker's dialect would have on word recognition scores. Effect of speaker dialect is of particular concern in word recognition testing with children because monitored live voice (MLV) presentation of test stimuli is often used.

Word recognition testing determines the listener's ability to make correct phonemic judgments on the basis of acoustic information or to recognize the sounds of speech. Although the premise of word recognition testing is simple, there are numerous factors which affect the reliability and validity of the results. Penrod (1985) suggested that the reliability of word recognition testing was influenced by three types of factors: physical factors, linguistic factors, and test administration variables. Physical factors would consist of level of presentation and signal to noise ratio. Linguistic factors would involve the articulation and dialect of the tester or testee and contextual cues (i.e., carrier phrase) provided by the tester. Test

administration variables would include the use of MLV or taped stimulus, response mode and scoring procedure.

As a result of children's developing linguistic, phonological and articulation abilities, special word recognition tests have been developed for use with children and several modifications to the adult word recognition testing procedures have been made. One common modification of particular interest to this study is the use of MLV presentation of word recognition stimuli. While test administration to adults is presented with a standardized tape recording, MLV testing has been routinely used with children because it allows for variations in speed, latency of response and voice characteristics so that children's attention can be maintained.

When testing children using MLV presentation two of the reliability factors suggested by Penrod (1985) must be considered: test administration variables and linguistic factors. Reliability of word recognition scores using MLV as a test administration variable has been questioned by many researchers (Beattie, Svihovec, & Edgerton, 1978; Brandy, 1966; Creston, Gillespie, & Krohn, 1966). Although studies have shown conflicting results concerning the use of MLV presentation, the bulk of the studies have supported the clinical necessity and usefulness of the procedure with some populations as long as caution was observed with regard to the reliability of the results.

Linguistic variables such as speaker characteristics also must be examined with regard to reliability. Hood & Poole (1980) found significant differences in word recognition scores that were attributed to the speaker characteristics. It was not clear if dialect was among

those speaker characteristics which affected word recognition scores. However, any effect speaker dialect had on word recognition scores might be especially evident in children because of the child's developing linguistic system. Elliott, Connors, Kille, Levin, Ball, and Katz (1979) showed that word recognition scores obtained using a test normed for 3 year olds continued to improve during the elementary school years. In their study all subjects were tested to have hearing within normal limits. The ten year old subjects performed significantly better than the 5-8 year old groups. The authors indicated that a portion of the age-related improvement in performance may have reflected an increased ability to use "semantic closure" (Elliott et al., 1979, p. 20) to identify the target stimulus word. This increase in linguistic maturity might contribute to an increased competence in understanding dialects for children with age.

The purpose of this study was to investigate the effect of dialect on word recognition scores of Montana dialect speaking children. The word recognition scores of twenty seven-year old children were obtained via taped presentation of the Phonetically Balanced Kindergarten 50 (PBK) word recognition test and the Nonsense Syllable Test (NST). A Montana dialect and a General Southern dialect speaker presented stimuli on equivalent forms of these word recognition test lists. The study hypothesized that children of a Montana dialect would obtain significantly different scores on a taped word recognition test when the speaker had a Montana dialect as opposed to a General Southern dialect. Furthermore, the study hypothesized that the dialectal effects would be

evident on the PBK word recognition scores, but not on the NST word recognition scores.

CHAPTER 2

LITERATURE REVIEW

Factors Affecting the Reliability of Word Recognition Scores

Test administration variables

Penrod (1985) suggested several test administration variables which could affect the reliability of word recognition test scores: response mode, scoring procedure and test presentation of word lists. The response mode involved how the individual being tested would designate their response. For example, the client could verbalize, write down or point to a picture of what they perceived.

The scoring procedure discussed by Penrod (1985), on the other hand, would involve the method by which the responses were judged as correct or incorrect. The response may be counted correct if it acoustically matched the stimulus presented by the speaker or if it showed a clear understanding of the stimulus meaning (i.e., repeated the word "dog" in a different dialect, but the meaning was still clear).

Test presentation could be via tape or monitored live voice (MLV). Hood and Poole (1980) stated, "little evidence that any of the various parameters . . . in this field in the past are of any great importance to clinical speech audiometry with the possible exception of live-voice presentation" (Hood & Poole, 1980, p. 453).

Research on the reliability of MLV vs. taped presentation has had mixed results in the past. Current research articles concur that MLV is

an effective clinical tool necessary for special populations (such as children or mentally handicapped individuals) in spite of the questionable reliability. However, some authors stress the importance of taped presentations for truly reliable results. Brandy (1966) looked at the difference in word recognition scores obtained using taped vs. simulated live voice presentations by the same speaker. Stimuli were three randomizations of one recording of a Central Institute for the Deaf (CID W-22) word list that had been acoustically corrected and tapes of three randomizations of the same word lists spoken by the same speaker on separate days under identical conditions. The second set was not acoustically corrected, so that it mimicked live voice testing. The results for the recorded presentations (i.e., acoustically altered) were significantly better than the results for the live presentations (i.e., stimuli not acoustically equivalent).

Further evidence challenging the reliability of MLV presentation was found in Penrod's (1979) study. This study was conducted using subjects with varying degrees of sensorineural hearing loss. Four different speakers presented word recognition tests at relatively homogeneous intensities. Twenty-six of the 30 subjects exhibited clinically significant variations in their word recognition scores for the different speakers. Penrod used these results to suggest a need for standardized presentation.

Studies have shown MLV does not allow for the control and consistency across testing that taped presentation offers. Authors have stressed the need for standardization not provided by MLV presentation. However, the flexibility of the test procedure is an important clinical

aspect, especially when testing children. Since children often have a short span of attention, a test procedure such as MLV may be necessary to adapt to the child's needs and obtain measures as valid and reliable as possible. Monitored live voice presentation is not the most reliable method of testing, but further research to improve its reliability is necessary due to the great percentage of professionals who utilize the procedure.

Linguistic Factors

Linguistic factors affecting word recognition scores were also addressed by Penrod (1985). The article suggested the most prominent linguistic variables were articulation and dialect, contextual cues, redundancy and the familiarity of the words to the listener. The present study was concerned most about the effects caused by speaker differences, a subject examined frequently in past research.

John Palmer (1955) examined the changes in word recognition scores between men, women and children as the speakers. The commonly accepted premise was that "hard of hearing individuals hear the voices of men 'better' or 'easier' than the voices of women" (Palmer, 1955, p. 192). The subjects were 13 hard of hearing individuals who listened and responded to the three male, three female and three children speakers. The results indicated that there were no significant differences in the subjects' abilities to understand the speaker based on gender. Palmer suggested that other aspects, such as the articulatory characteristics of the speakers be looked at to explain variations in scores obtained by different speakers.

Kreul, et al. (1969) explored changes in item and overall test difficulty of word recognition test scores as a function of carrier phrase, speaker, reutterances by a speaker, and level of accompanying noise. Their results, using General American English speaking subjects with normal hearing, showed that test difficulty changed significantly with regard to speaker differences (and carrier phrase). The authors cautioned that the test standards only be applied to the specific set of conditions and for the population of listeners represented in the standardization of the test.

Hood and Poole (1980) further examined the effects of speaker differences on word recognition testing. The authors looked at changes in the word order difficulty (determined by how many times each item was missed overall) when different speakers presented the test stimuli. The results demonstrated clearly the "dominant role" (Hood & Poole, 1980, p. 451) the individual speaker played in the word order difficulty rating. The difficulty of the words on the list were considerably varied for the three speakers. Frank and Craig (1984) examined the differences in the word recognition scores of normally hearing adults between the Auditec and the Rintelmann recordings of the Northwestern University-6 (NU-6) word recognition test at different intensity levels and with different signal to noise (S/N) ratios. While both the Auditec and Rintelmann recordings were standardized versions of the NU-6 with good interlist equivalency and retest reliability, the tests yielded significantly different results. The authors pointed out that both tests were effective for clinical practice, but that the word recognition scores

were significantly poorer using the Auditec than the Rintelmann recording in multitalker background noise.

Penrod (1979), on the other hand, held that while talker differences were responsible for a small portion of the variability in scores, the talker-listener interaction was a factor of greater importance. This study, using subjects with sensorineural hearing loss, stated that since the variability was spread across all talkers, it was the talker-listener interaction rather than the speaker differences that caused the variability in test scores. They argued that if it were the speaker differences, a particular speaker should yield poorer overall results across subjects compared to the other speakers. As in the majority of the research in this area, the author stressed the need for standardized presentation to overcome the variation reflected in the word recognition scores due to factors such as speaker differences.

Most of the research surveyed agreed that speaker differences could significantly affect word recognition scores. Several of these studies used subjects with sensorineural hearing loss and all of them used adult subjects. However, none of the studies specifically dealt with how these factors affect test results of children or looked at issues of dialect.

Dialect

What is a dialect?

Taylor (1983) stated that there are a number of different factors which constitute dialectal entities (i.e., region, social status, etc.)

included in formal definitions of dialect (Carver, 1987; Gleason, 1989; Reed, 1967 and Taylor, 1986). Gleason defined dialect as a "systematic subvariant of a particular language, spoken by a sizable group" (Gleason, 1989, p. 330). She indicated that dialects can vary across different dimensions (i.e., social status) and share a varying number of features.

Carver defined a dialect as "a variety of language distinguished from other varieties by a set of grammatical, phonetic, and lexical features" (Carver, 1987, p. 1) The term was broken down further into regional dialects, which encompassed certain features distributed geographically over a restricted and relatively uniform area and social dialect, which was a language shared by a particular social grouping. Regional dialects could be as diverse as the difference between a Boston dialect and a West Texas dialect or as close as the difference between a Standard American English dialect and a General American English dialect.

Social dialects would be more difficult to separate. A trained dialectologist would probably be able to differentiate between a middle class and an upper class American. However, the slight variations in dialect would not be noticeable to untrained listeners.

All of the known dialects are made up of "idiolects" (Reed, 1967, p. 3). Reed defined idiolects as the way each person makes use of his language in accordance with his individual history. These idiolects allow us to tell one person from another by voice such as over the telephone. Dialectologists group idiolects together to form the various dialects known within all the language systems.

Effects of dialect on speech perception and production

Bountress (1983) expressed concern about the relationship between dialects (specifically nonstandard English and black English) and educational achievement, intellectual development, racial isolation, and economic impoverishment. He described the Ann Arbor decision (1979), which "ruled that a Michigan school district had to develop a program to ensure that teachers would become sensitive to the linguistic characteristics of black English and to the manner in which dialectal interference affects reading performance, specifically, and academic performance, in general" (Bountress, 1983, p. 72). He indicated that mandates, such as the Ann Arbor decision, magnify the need for further research on regional dialectal characteristics and issues about "dialectal shift" (p. 75).

Bountress (1983) examined the effect of racial composition of the student population on changes in selected dialectal features among speakers of black English. He chose 60 black children (first, second, and third grade level) as subjects. Thirty of the subjects were from an all-black school and the other 30 were from a school with an equal number of black and white students.

Bountress (1983) had the students repeat thirteen sentences from a commonly used language test (Carrow Elicited Language Inventory) and examined the changes in the form of the copula, the omission of /s/ and /z/ inflections, and substitution of /d/ for voiced "th" (common dialectal features of "black English"). The omission of copula seemed to be most resistant to change regardless of setting.

On the other hand, results indicated a statistically significant decrease in the frequency of /s/ and /z/ omission as a function of grade in the integrated school, not in the all-black school. Furthermore, the /s/ and /z/ omissions by the black students in the integrated school were replaced by the common standard English productions. Results showed a statistically significant decrease in the substitution of /d/ for the voiced "th", the common standard English production, as a function of grade for both educational settings.

Sou-Mee Tse and David Ingram (1987) indicated that the influence of dialectal variation on children's language has not been widely studied, but deserved attention. Tse and Ingram studied one child, who lived in a Cantonese speaking household where her father and mother spoke different dialects of Cantonese. The child was observed from age 1 year 7 months to 2 years 8 months to examine her language development. Tse and Ingram hypothesized that there would be a period of confusion before the child sorted out the systematic features of the variation.

The results of Tse and Ingram's (1987) study, focused mainly on the variation between /n/ & /l/ as two distinct phonemes (father's dialect) and /l/ as a single phoneme substituted for /n/ (mother's dialect), showed no evidence that the subject was acquiring either the father's or mother's dialect. Instead, the subject seemed to use /l/ and /n/ as free varying allophones of a single phoneme. The authors indicated that the subject appeared confused and was "actively seeking a solution to the input data" (Tse & Ingram, 1987, p. 291). They further suggested that it is more difficult for a child to resolve dialectal variation than language variation since the monolingual child has the

task of separating two dialectal variations that have many similar characteristics as opposed to a bilingual child faced with two entirely distinct systems.

Levy and Cook (1973) administered a taped expressive dialect proficiency task and a taped auditory comprehension task to 32 black second graders. The examiner was a black male, who spoke to all of the children in standard English dialect. The dialect proficiency task consisted of a tape recorded version of 20 sentences, ten sentences in standard English and the same ten sentences in "black nonstandard English" (Levy & Cook, 1973, p. 642). The children's oral speech was considered "bidialectal" (p. 647) with features of both standard English and "black nonstandard English" (p. 642), even though they were "generally more proficient" (p. 647) in repeating sentences presented in standard English.

After Levy and Cook's (1973) dialect proficiency task was completed, the auditory comprehension task was administered. This task consisted of taped sets of four stories with seven questions following each story. One set of stories was presented in standard English and the other in "black nonstandard English." Half of the subjects received the tape of the auditory comprehension task in "black nonstandard English" and the other half in standard English.

Results of Levy and Cook's auditory comprehension task showed that subjects who listened to stories in standard English scored higher on auditory questions than when they listened to the same stories in "black nonstandard English" (p. 642). Levy and Cook (1973) described "demand characteristics" (p. 648) possibly contributing to the results. They

stated that even though the experimenter was black like the children, the experimental situation may have been too formal and test like to permit the children to respond freely to black dialect material. Furthermore, the experimenter's use of standard English dialect may have set the stage and may have been indicated as more appropriate to the children for that situation.

Based on Levy and Cook's (1973) results, they posed two pertinent questions with regard to bidialectal children: "What are the social and situational cues influencing auditory and reading comprehension of the two dialects (i. e., standard English and 'black nonstandard English') or degrees of dialect? What are the influences of age, sex, socioeconomic class, geography, child-rearing practices, housing patterns, and types of schooling on dual dialect learning and proficiency?" (Levy and Cook, 1973, p. 648).

Research has shown that dialect can cause significant effects on children's educational achievement, intellectual development, and speech perception and production. If not remediated, the difficulties may compound or intensify. Adler (1973) questioned whether or not a test could be totally culture free, but stressed that "attempts to design such a test should be fostered . . . without question" (Alder, 1973, p. 31). Dialect is only one of many factors associated with "culture", however, in light of previous results, designing tests free of dialectal factors would be an appropriate direction for further research.

Dialect and testing

Taylor and Payne (1983) addressed the demands of Public Law (PL) 94-142 (1975) that "all test materials and procedures used for the evaluation . . . be administered in such a manner that they are not racially or culturally discriminatory" (Taylor & Payne, 1983, p. 8). They discussed several aspects concerning nondiscriminatory testing including results of discriminatory testing, types of bias in speech and language assessments, and ideas for minimizing cultural bias in assessment procedures for various populations.

Among their ideas, Taylor and Payne (1983) suggested that a tester conduct an "item analysis" (p. 16) of the instrument being used to determine whether or not it was biased against the individual being tested because of their language or dialect. For the item analysis the phonological, semantic, syntactic, and pragmatic "assumptions of normalcy would be compared with the linguistic assumptions of the client's home community" (p. 16-17).

Once the particular biases had been identified, one option of remediation would be to establish new test norms for the targeted population by obtaining typical response profiles and scores from random samples of normal persons in the targeted population. Taylor and Payne (1983) called these strategies "proactive because they propose constructive solutions to a real clinical problem" (p. 19).

Taylor and Payne (1983) stressed that the issues of discriminatory testing pertain to any type of clinical setting that could lead to faulty management. The authors indicated that the use of discriminatory tests in school settings frequently resulted in inappropriate placement

of many children from culturally and linguistically diverse populations into special education classes where they did not belong. Even though clinicians might be reluctant to alter standardized test procedures, there are professionally ethical techniques that should be employed to modify testing procedures and minimize bias. On the other hand, the authors indicated that bias could occur when an examiner "thinks that the assessment procedure has to be altered to take into account the presumed dialect of the client" (p. 13), not taking into account the fact that some individuals can communicate according to the rules of other dialect groups as well as their own.

Miller-Jones (1989) discussed various aspects of importance regarding culture and testing. She suggested that tests of generalized cognitive functioning will provide a less than accurate portrayal of individuals' capacities and that appropriate assessment requires an understanding of the constraints that regulate a person's knowledge, conceptualization abilities, and reasoning processes. Items of a standardized test must not be biased and should not "favor a particular sociocultural experience over any other" (Miller-Jones, 1989, p. 360-361).

Miller-Jones (1989) recommended that in order to improve success in assessing the competencies of children from "diverse ethnic backgrounds" (p. 364), test procedures should use multiple tasks with a variety of materials, use tasks which sample the domain of the culture in question, and probe for the reasoning behind a child's response to determine the consequences of cultural differences. These suggestions for improved success of test procedures could possibly be applied to

overcome dialect effects as well, even though Miller-Jones did not address dialect specifically.

The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985) published Standards for Educational and Psychological Testing which indicated, "for a non-native English speaker and for a speaker of some dialects of English, every test given in English becomes, in part, a language or literacy test" (American Educational Research Association et al., 1985, p. 73). Standard 13.1 stated that for speakers of some dialects "testing should be designed to minimize threats to test reliability and validity that may arise from language differences" (p. 74). The standards suggested that specially trained personnel conduct the test administration for particular populations.

Standard 13.4 held that "when testing is translated from one language or dialect to another, its reliability and validity for the uses intended in the linguistic groups to be tested should be established" (p. 75). Further standardization with members of particular groups could be performed to eliminate disadvantages caused by language differences, including dialect.

The American Speech-Language-Hearing Association (ASHA) adopted a position paper "Clinical Management of Communicatively Handicapped Minority Language Population" (ASHA, 1985). ASHA reported 1980 Census data that indicated 34.6 million or 15% of the U.S. population was composed of native speakers of various minority languages (i.e. non-English speaking or English as a second language). An estimated 3.5

million of these speakers have speech, language, or hearing disorders unrelated to their minority language.

While some assessment and remediation tasks are not affected by a client's use of a minority language, ASHA (1985) indicated several aspects of speech, language, and hearing assessment and remediation that was highly complicated by the client's use of a minority language (or language different from the tester). Among the list was auditory discrimination, which required responses and understanding from both the tester and the testee. The article indicated that even if an examiner was familiar with the language of the subject, dialect differences within that language might have been a confounding variable in assessment. Therefore, speech-language pathologists and audiologists "must provide services with consideration of such cultural variables, in addition to consideration of language differences" (ASHA, 1985, p. 30).

Future directions included promotion of "continued advancement of knowledge" (p. 31) to increase the number of speech language pathologists and audiologists competent to serve minority language populations and independent study of the growing literature on minority language populations. ASHA (1985) acknowledged the need for further research regarding minority language populations to allow audiologists and speech-language pathologists to more competently handle clinical situations (ASHA, 1985).

Although dialect was not mentioned specifically within ASHA's (1985) position statement, previous research had indicated that some dialects could significantly affect individual's performance on various tasks. The task of word recognition testing may be one possible

situation in which dialectal effects would be evident. Not only must professionals be prepared and competent to deal with problems as diverse as a minority language, but they must also be aware and competent to handle any affects of different dialects within the English language.

Effective in 1993, ASHA will require students in speech language pathology and audiology to have coursework that addresses "issues pertaining to normal and abnormal human development and behavior across the life span to culturally diverse populations." Again, dialect is one factor within the many variables affecting a "culturally diverse population." The report urged students to prepare for future practice by questioning or fact finding issues related to "multicultural groups," writing research papers on various related topics, and attending conferences on "cultural diversity." ASHA suggested that with the demographic make-up of our country the chances of working with individuals of culturally diverse populations was increasing. As a result, current training of students was imperative for adequate management of professional situations.

Several of the articles discussed are primarily concerned with differences and biases as a result of various "cultural differences" or problems from having English as a second language. Dialect, a part of the "cultural differences," has not been studied as thoroughly as more broad topics such as minority languages. However, effects of tester dialect on test results such as word recognition test scores also warrant further attention and research to allow professionals to test all individuals in a valid and reliable manner. Bountress (1983) stressed that there is a need for more data-based information on cross-

cultural examiner effects in therapy settings for the development of "culture-fair" (Bountress, 1983, p. 75) evaluative tools.

Testing Children with Speech Audiometry

Diefendorf (1983) suggested that speech audiometry with pediatric patients "must be viewed as essential for developing a complete profile of auditory function and hearing ability" (Diefendorf, 1983, p. 241). Pure tone thresholds do not provide the audiologist with a precise evaluation of the child's ability to receive and respond to a speech message. He described the information from speech audiometry "fundamental" (p. 241) in habilitation and educational strategies for young hearing-impaired children. Furthermore, speech audiometry tests provide a validation check of the pure tone data.

Phonetically Balanced Kindergarten-50 Test (PBK)

Martin and Gravel (1989) surveyed 500 randomly selected certified audiologists in the United States to assess the current status of pediatric audiometry. Two-hundred and fifty audiologists responded to the survey. Ninety-seven percent of the responding audiologists used some type of speech recognition test for children aged 3-6 years old. Respondents emphasized the need to use the most sophisticated test which lies within the child's linguistic and cognitive abilities. The test most frequently used by responding audiologists was the Phonetically Balanced Kindergarten-50 (PBK) test. Eighty percent of the responding audiologists reported using the PBK for word recognition testing.

Northern and Downs (1984) concurred that the PBK currently was "probably the most widely used speech discrimination test for children" (Northern & Downs, 1984, p. 153). Haskins' (1949) PBK word recognition test has 3 lists, each with 50 phonetically balanced monosyllabic words selected from spoken vocabularies of kindergartners. The test was developed using normally hearing adults with the goal of providing a test to measure the discrimination abilities of hearing-impaired children and adults with limited language ability. The open-set test is usually presented via MLV and is most effective for children above 4 and one-half years of age because of the kindergarten level of words (Northern and Downs, 1984).

Sanderson-Leepa and Rintlemann (1976) compared different aged children's word recognition scores on two children's word recognition tests, the Word Intelligibility by Picture Identification (WIPI) and the PBK, and one adult test, the Northwestern University Auditory Test Number Six (NU-6). The selection of the three tests was based on the authors' judgement of their clinical and research utility. The goal of the investigation was to provide data to assist clinicians in selecting test procedures appropriate for particular age groups.

Subjects in Sanderson-Leepa and Rintlemann's (1976) study were 60 normal hearing children. Twelve children from each of the age-groups 3-1/2, 5-1/2, 7-1/2, 9-1/2, and 11-1/2 were tested using the three word recognition tests. The 3-1/2 year olds scored better on the WIPI than on the PBK or NU-6. Since normal hearing children are expected to perform at a high level on a word recognition test that is within their

linguistic level, the authors indicated that the WIPI was the most appropriate test for children 3-1/2 years old. The five and one-half year old and older children performed as well on the PBK as they did on the WIPI, but only the 11-1/2 year old children scored equally well on the NU-6.

Sanderson-Leepa and Rintlemann (1976) suggested that the PBK would be the most appropriate test for the 7-1/2 and 9-1/2 year old children since the NU-6 was too difficult and the WIPI was below their linguistic functioning level. Children aged 11-1/2 could be tested with any of the three tests. However the NU-6 would be the test of choice since it would be the most sophisticated test that they could successfully complete.

Nonsense Syllable Test (NST)

While most audiologists use standardized word lists to test word recognition, Resnick (1984) reported that nonsense syllable materials appeared to offer several advantages for assessing phoneme identification errors in children. Danhauer, Lewis, and Edgerton (1985) supported the use of a nonsense syllable test with children because it would permit an evaluation of phoneme recognition without the influence of semantic content. Edgerton and Danhauer (1979) developed the Nonsense Syllable Test (NST) as a 25 item CVCV (consonant-vowel-consonant-vowel) open-set test of phoneme identification. The list was developed from stimulus items that were directly tested on subjects with sensorineural hearing loss and was found to be sensitive to "the phoneme

recognition abilities of both normally hearing and hearing-impaired adult listeners" (Edgerton and Danhauer, 1979).

Danhauer, Lewis, and Edgerton (1985) examined normal hearing, school-aged children's and adults' performance on the NST to provide normative data for future clinical use. The NST was administered to three children's age groups and one adult age group at four different sensation levels. The youngest group (6:0-7:11) had some difficulty at the softest presentation level of 25 dB SL. Their mean score was 81.2% correct compared to 86% in the two other children's groups and 89.5% correct for the adult group. At all higher presentation levels the children's scores were not significantly different from the adults' scores.

According to Danhauer et al. (1985), the consistency across groups suggested the results were not influenced by the children's receptive language abilities. Since the NST appeared to tax even adult normal listeners (i.e., few scored 100% correct), the authors suggested that the NST in combination with a monosyllabic word test could provide a better idea of pediatric listener's speech discrimination abilities than with meaningful stimuli alone. Furthermore, a combination of test procedures could eliminate the influence of other confounding factors such as receptive language level.

Issues in speech audiometry testing

Jerger (1984) indicated that children's performance on speech audiometry tasks was influenced by their nonauditory cognitive skills, receptive language abilities, and chronological age. She suggested the

influence of children's receptive language abilities could be minimized in three ways. Test materials could be limited to: monosyllabic words documented to be in the recognition vocabulary of normal children within the age range of the subjects being tested; materials which represented the actual responses to picture stimulus cards by normal children, within or below the age range for which the test would be used; and materials based on speech samples elicited from hearing-impaired children above the age for which the test would be used.

Elliott et al. (1979) tested six different subject groups (5, 6, 7, 8, 10 years, and adult) using a closed set word recognition test standardized on inner city three year old children. The subjects were normally progressing school children and adults, some with learning problems, and some with "developmental articulation problems" (Elliott et al., 1979, p. 16) (school's diagnostic team anticipated self-correction without therapy). Tapes of the word recognition test were made using two General American English speaking testers and administered under four test conditions: quiet, open-set; quiet, closed-set; babble, closed-set; and filtered noise, closed-set.

Elliott et al. (1979) hypothesized that the stimulus words would be highly familiar to children five years old or older who had normal intelligence since the test was developed to be within the receptive language skills of three year old inner city children. A Newman-Keuls analysis of the data for the quiet, closed-set and quiet, open-set conditions indicated that the normal ten year old group performed significantly better than all subjects in the younger groups. The five year old subjects performed significantly more poorly than the eight and

ten year old children for the quiet, closed-set condition. Therefore, even though the factor of linguistic ability was supposedly removed, the results continued to show significant differences for various age groups. Elliott, et al. suggested the possibility that a portion of the age-related improvement in performance may have reflected an increased ability of the older subjects to use "minimal amounts of acoustic information regarding the vowels and to achieve 'semantic closure'" (p. 20).

Another controversial issue with speech audiometry testing has been the use of MLV presentation. The reliability issues involved with MLV presentation (discussed in a previous section) were important to pediatric speech audiometry because of the widespread use of MLV presentation with this population. Olsen and Matkin (979) cited a Martin and Pennington (1971) survey which indicated that 65% of clinical respondents used MLV methods for speech discrimination testing. While Martin and Pennington did not survey the use of MLV directly with word recognition testing, their survey showed that 98% of responding audiologists use MLV when performing speech detection thresholds with children. Diefendorf (1983) stated that "the drawbacks of monitored live voice when testing speech discrimination must be recognized, as they are with adults, particularly when testing children over time with different speakers" (Diefendorf, 1983, p. 247).

Summary

Several factors had been described which affected the reliability of word recognition test scores. Among the more significant were the

speaker differences observed during MLV presentation of word recognition tests. Dialect, one aspect of speaker difference, was examined with regard to its effect on children's speech production and perception and performance on tests. However, dialectal effects observed during speech recognition testing had not been studied.

Dialectal effects were of considerable importance in word recognition testing of children since MLV presentation was frequently used. Issues regarding children's receptive language abilities and maturation must be considered during pediatric speech audiometry. Elliott, et al. (1979) showed that young children appear to require more acoustic information than do older children in order to identify familiar words.

Furthermore, Graham and House (1981) suggested that differences in speech sound identification between children and adults may be attributed to children "responding to small but perceptible differences which would have been ignored had their phonological systems been further developed" (Graham & House, 1971). Any dialectal effects on speech recognition testing might have been more apparent when testing children whose linguistic skills were still developing.

The PBK and NST were both shown to reduce the effects of receptive language abilities for children aged seven years old. Therefore, any dialectal effects with this population would be observed in a setting where the research design had controlled for other factors such as test administration, receptive language abilities, and maturation level.

CHAPTER 3

METHODS

Subjects

Twenty subjects between the ages of 7 years, 0 months, and 7 years, 11 months, were used for this study. To participate, each child was required to have normal hearing sensitivity, age appropriate phonological development, attendance in an age appropriate classroom with no special services and a Montana dialect with no exposure to other dialects within the immediate family.

Normal hearing sensitivity was established by passing a pure tone screening at 15 dB HL for the octave frequencies from 250 Hz through 8000 Hz. Phonological development was informally assessed by asking the subjects various questions and by having the subjects name 10 pictures (see Appendix A). Subjects exhibiting articulation errors outside of the 7 year old age-equivalent norms were eliminated from this study. Information regarding appropriate classroom attendance and dialect was obtained by posing questions to the parents (see Questionnaire, Appendix B). Since minors were used for this study, a release/consent form was signed by the parents prior to testing (see Appendix C).

Instrumentation

Test materials and development

The first 25 monosyllabic words of list one of the PBK-50 word recognition test (Haskins, 1949) and the 25 items of List A of the

Nonsense Syllable Test (Edgerton & Danhauer, 1979) were selected as test stimuli. Both lists of words were presented in two forms distinguished by item order. The two PBK word lists were labeled as PBK 1 and PBK 2 and the two NST lists were labeled as NST 1 and NST 2. PBK 1 was the first 25 words of PBK-50 List 1 in the original order. PBK 2 consisted of the same 25 words in a randomly assigned order. NST 1 was the Nonsense Syllable Test List A in the original order. Those same words were in a randomly assigned order to create NST 2 (see Appendices D and E for the four forms).

Two females, ages 24 and 27, were selected to represent the Montana and General Southern dialects. Their dialects were certified by a University of Montana Linguistics professor with a specialty in dialectology. Tape recordings of the two speakers reading PBK 1 & 2 and NST 1 & 2 were prepared as described in the test preparation section. The nonsense syllables were transcribed by the author from the Phonetic Alphabet to a standard English version. Both speakers repeated the words prior to testing to verify understanding of the transcriptions.

Test preparation

The PBK and NST word recognition tests used for this study were recorded in a sound treated room (Industrial Acoustics Company, Inc.) through a Pioneer Stereo Amplifier A-5 connected to a Nakamichi BX-100 tape recorder onto Maxell XL II 90 cassette tapes. The TOA Electric Co., LTD wireless microphone system was attached to the speaker's clothes approximately 4 inches from her mouth.

The speaker presented the items from the four lists at four second intervals using the carrier phrase "You will say _____." Each speaker monitored her presentation level by having the carrier phrase peak at 0 dB on the VU meter of the Nakamichi tape recorder. The microphone sensitivity was adjusted such that the speaker's presentation at a comfortable level would peak at 0 dB on the VU meter of the tape recorder, standard procedure for word recognition testing.

A 1000 Hz tone was superimposed onto the beginning of each list for calibration purposes. The tone was presented via sound field with the microphone sensitivity adjusted to the speaker's "comfortable setting." The intensity of the tone was adjusted to peak at 0 dB on the audiometer's VU meter during test administration.

Subject testing

Pure tone screening and word recognition testing were conducted in a sound treated room (Industrial Acoustics Company, Inc.) with a Grason-Stadler 16 audiometer. The PBK and the NST word recognition lists were presented to the subjects on a Fisher cassette tape recorder and fed through the Grason-Stadler 16 audiometer and TDH-50 earphones coupled with TDH 50P MX4/AR cushions to the client's right ear. A TOA Electric Co., LTD wireless microphone system was attached approximately four inches down from the client's mouth and the verbal responses were recorded through a Pioneer Stereo Amplifier A-5 to a Nakamichi BX-100 tape recorder onto Maxell XL II 90 cassette tapes.

Instructions

Each subject was verbally instructed face-to-face as follows: You will hear a woman's voice on a tape telling you to say some words. I want you to repeat what you hear her say. For example she might say, "You will say 'cat'." What do you say? Yes, "cat" (or reinstruct if incorrect). You will also hear the woman say some non-sense items that don't make sense to you. Please repeat back what you hear her say. So if she said, "You will say zoofu," what do you say? Yes, that is correct (or reinstruct). I want you to guess if you are not sure what the woman said. Please speak nice and clearly for me so I can understand you. Do you have any questions?

Test Conditions

Word recognition scores were obtained with the test stimuli presented at 55 dB HL (a normal conversational level). The level was set by having the calibration tone of each list peak at 0 dB on the audiometer's VU meter. Each subject was given each of the four tests. The test conditions were randomly assigned to the subjects with regard to the four test forms, the two speakers and presentation order.

The subjects' responses were recorded and judged as correct or incorrect by trained listeners of Montana dialect. The examiners were provided with taped versions of the PBK and NST item lists used in this study and taped versions of the subjects' responses. All of the word recognition tests scored by the primary examiner were played back on a Nakamichi BX-100 tape recorder system. The word recognition tests for a

randomly selected ten percent of the subjects was scored by the secondary examiner on the same system for inter-examiner reliability measures. Both examiners were graduate students in speech language pathology and audiology with a minimum of 100 clinical hours in audiometric testing. The clinical hours ensured that the examiners were trained listeners.

The observers were instructed to score each response for the NST tests as correct if it acoustically matched the stimulus. Responses for the PBK tests were scored on two criteria: if they acoustically matched the stimuli and/or if the response was judged to be the same word as the stimulus despite acoustic differences. Errors were transcribed and examined for patterns.

Data Analysis

The dependent variable in this study was the word recognition score. The independent variable was the dialect used for presentation of the two tests. Inter-observer reliability was determined by calculating a point by point percentage of agreement for a randomly selected 10% of the subjects. The data were analyzed using t-tests performed at the 0.05 level of significance.

CHAPTER 4

RESULTS

Three hundred and fifty consent forms requesting participation in the study were sent out with a yield of 33 responses. Of these, 20 children met the subject criteria and kept their appointments. Subjects generally performed better on the PBK test (84%-100% correct) than on the NST (44%-96% correct). Inter-observer reliability was also better on the PBK list than on the NST list.

Phonetically Balanced Kindergarten Test (PBK)

Individual subjects' scores on the PBK test presented with the Montana dialect showed scores between 88% and 100% correct with a mean of 96.2%. For the General Southern dialect presentation of the PBK test the range was 84% to 100% correct. The mean score was 94.6%. Figure 4.1 shows the distribution of scores across subjects for both presentations.

Individual subjects had similar scores on the PBK lists presented in the two different dialects. All but two subjects had scores within 8% (2 words) of each other. Subjects 4 and 9 had differences of 16% and 12%, respectively, between the two dialect presentations; both had better scores with the Montana dialect presentation.

Table 1 shows the frequency with which words on the PBK were missed when presented in the two dialects. Most of the errors (61%) occurred on four words: "fold," "sled," "bad," and "mouth." "Bad" and "sled" were missed almost exclusively when presented by the General

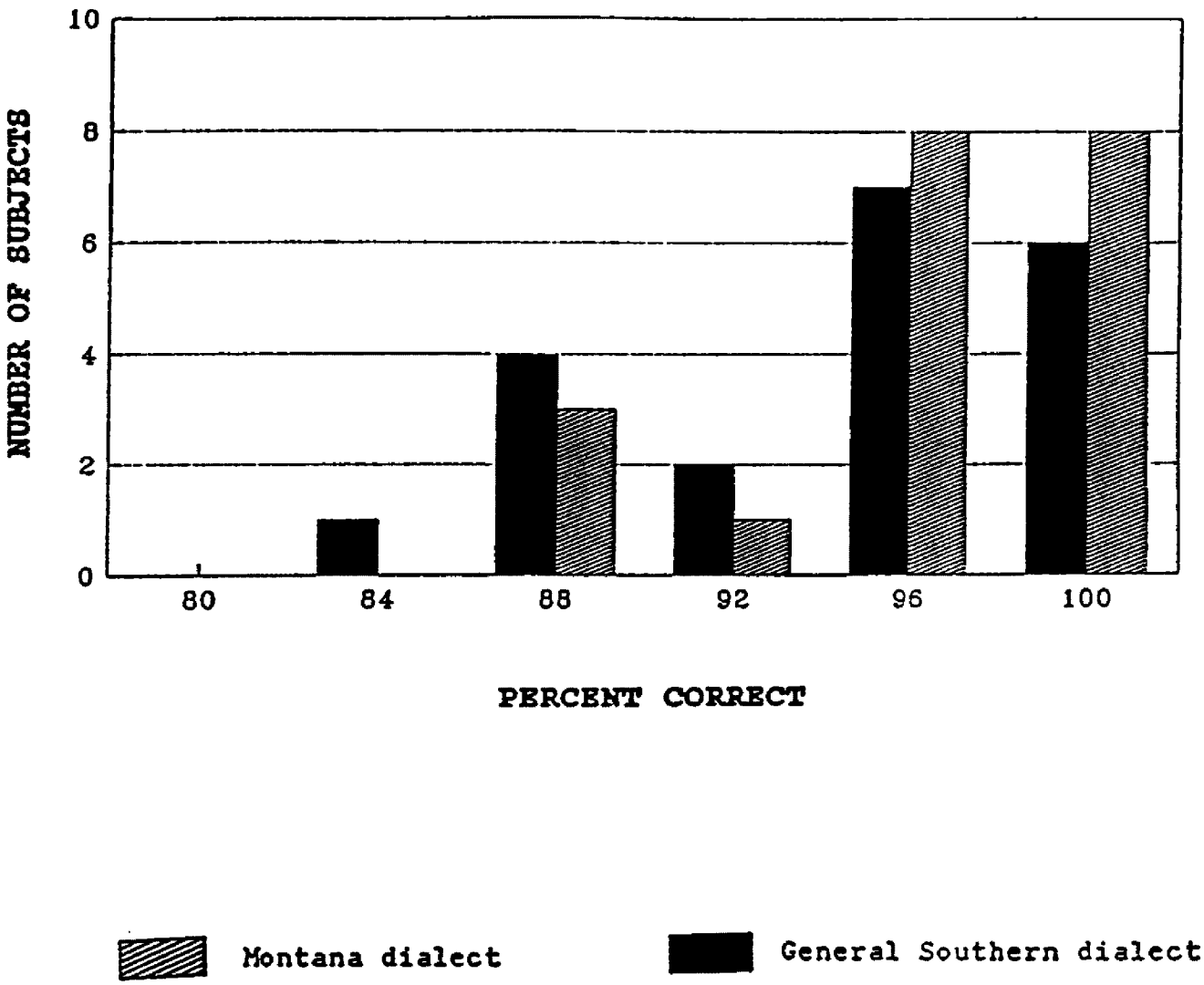


FIGURE 4.1 Subject's percent correct scores on the PBK Test as presented by a Montana and General Southern dialect speaker

TABLE 1

TOTAL NUMBER OF ERRORS ON EACH PBK WORD FOR TWENTY SUBJECTS WHEN PRESENTED BY A MONTANA AND GENERAL SOUTHERN DIALECT SPEAKER

PBK WORD	NUMBER OF ERRORS WITH MONTANA DIALECT PRESENTATION	NUMBER OF ERRORS WITH GENERAL SOUTHERN DIALECT PRESENTATION
fold	7	5
sled	0	6
mouth	5	1
bad	1	4
rag	2	1
box	2	1
bus	2	1
five	0	2
no	0	2
need	0	2
slice	1	0
please	0	1
such	0	1
great	0	0
pants	0	0
rat	0	0
pinch	0	0
ways	0	0
put	0	0
fed	0	0
hunt	0	0
are	0	0
teach	0	0
is	0	0
tree	0	0

Southern dialect speaker, "mouth" was typically missed when presented by the Montana dialect speaker, and "fold" was missed frequently during both dialect presentations. The remaining 39% of the errors were accounted for by nine other words for which no more than two errors were made per dialect presentation.

The results of the t-test for performance on the PBK with Montana and General Southern dialects indicated that the difference was not statistically significant at the .05 level of confidence (t (38) = .988 $p > .05$). Thus the hypothesis that speaker dialect would affect children's word recognition scores could not be accepted. The similarity of scores with both the Montana dialect and General Southern dialect indicated that the speaker dialect did not significantly affect the subject's performance on the PBK word recognition test.

Nonsense Syllable Test (NST)

The second hypothesis was that subjects' performance on the NST would not be affected by the dialect of the speaker presenting the list. The scores with the Montana dialect presentation ranged from 44% to 96% correct with a mean of 66.8%. The scores for the NST with the General Southern dialect presentation ranged from 68% to 96% correct. The mean score was 84% correct. Figure 4.2 shows subject's scores on the NST presented with both dialects.

Subject's performance on the NST appeared to be related to speaker. The vast majority of subjects (17 of 20) scored better when the NST was presented by the General Southern speaker than by the Montanan speaker. There were two subjects who scored better on the NST with the Montana dialect presentation (with a difference of one error) and one subject who scored equally well regardless of speaker dialect.

The apparent difference in performance on the NST was verified statistically. A t-test showed statistically poorer performance on the NST when presented by the Montana dialect speaker than when presented by the General Southern dialect speaker (t (38) = -13.03 $p < .05$)

Table 2 shows the frequency with which specific items on the NST were missed for each dialect presentation. There were three items that had no errors for the General Southern dialect presentation (/hɪθʊ/, /haðo/, and /ɡoθɹ/). All of the items had at least one error for the Montana dialect presentation.

Six of the NST items had ten or more errors with the Montana dialect presentation, while only one item presented by the General Southern dialect speaker had as many as ten errors. The majority of the items presented by the General Southern dialect speaker had less than five errors. Only ten of the NST items had five or fewer total errors when the test was presented by the Montana dialect speaker.

Reliability

The measures of reliability used were inter-examiner and intra-examiner point by point percentage of agreement. An 88% criteria was designated as acceptable. Agreement ranged from 88% to 100% as shown in Table 3. The inter-examiner agreement on the PBK was 96% to 100%, while

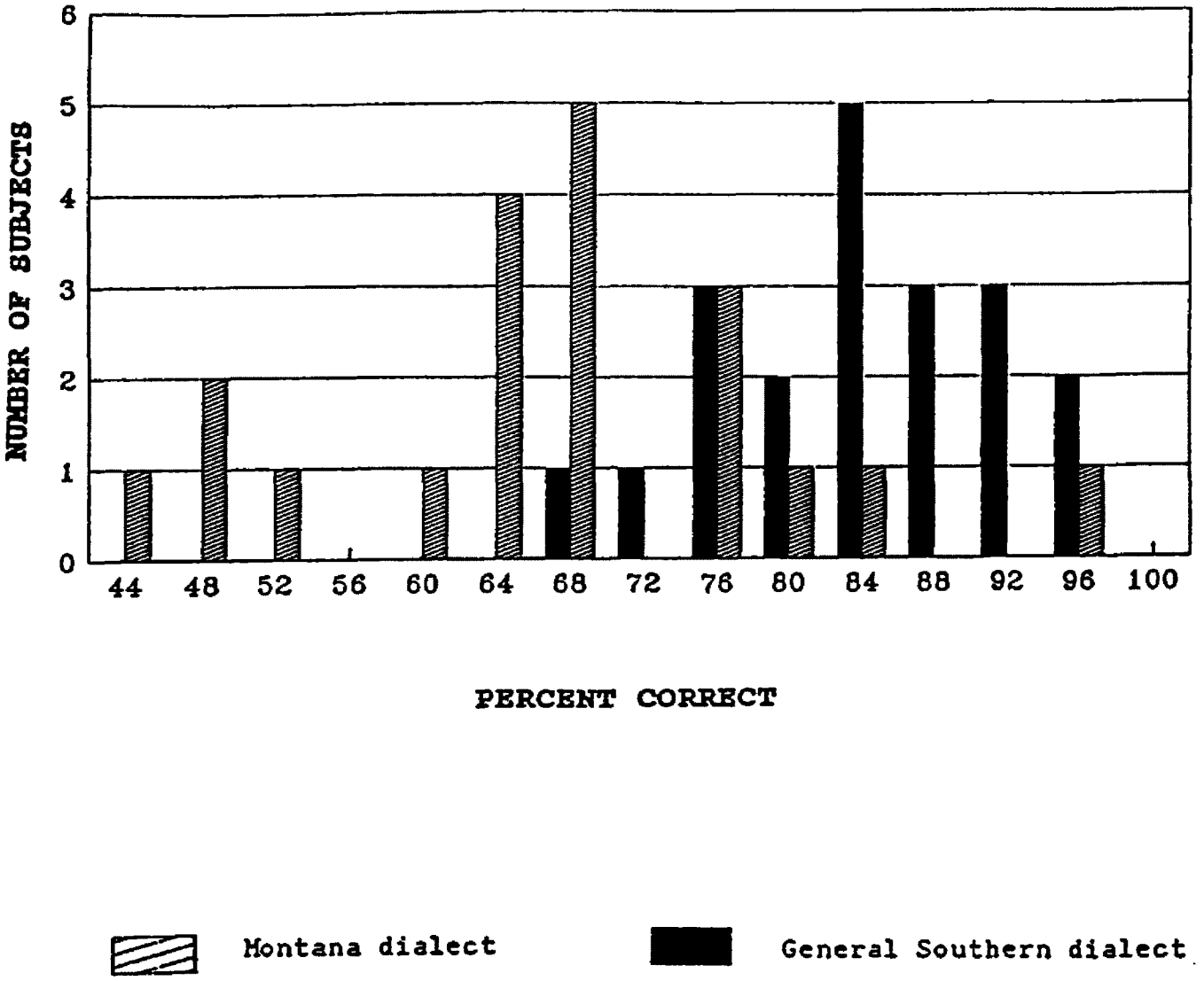


FIGURE 4.2 Subject's percent correct scores on the NST as presented by a Montana and General Southern dialect speaker

TABLE 2

TOTAL NUMBER OF ERRORS ON THE NST ITEMS FOR TWENTY SUBJECTS WHEN PRESENTED BY A MONTANA AND GENERAL SOUTHERN DIALECT SPEAKER

NST ITEM	NUMBER OF ERRORS WITH MONTANA DIALECT PRESENTATION	NUMBER OF ERRORS WITH GENERAL SOUTHERN DIALECT PRESENTATION
daθa	13	5
ðɪlɚ	13	4
bʌðɪ	11	5
ðɚwo	11	2
fʊθɪ	10	5
vʊvɪ	10	4
ðɪjo	7	10
bɚðɚ	9	3
ʤʌðʊ	8	4
ʃiʧa	8	2
ʃɪθɚ	8	2
pʌðʊ	8	1
ʃeθa	6	5
giðo	6	3
θʌfɪ	5	5
haðo	6	0
ðuma	5	3
sɛfɛ	3	5
ʧoðɚ	5	0
nɪθʊ	5	0
fisɚ	4	2
ðana	2	3
ðʊmɚ	2	2
piɾɪ	2	1
ʤɚfe	1	2

TABLE 3

INTER-EXAMINER AND INTRA-EXAMINER RELIABILITY MEASURES FOR TWO EXAMINERS AS A POINT BY POINT PERCENTAGE OF AGREEMENT FOR TEN PERCENT OF THE TESTS

INTER-EXAMINER AGREEMENT

Subject Number	PBK 1	PBK 2	NST 1	NST 2
12	96%	100%	88%	92%
3	100%	96%	92%	96%

INTRA-EXAMINER AGREEMENT

Examiner	Percent of Agreement on NST Lists
Primary Examiner	94%
Secondary Examiner	92%

on the NST agreement was 88% to 96%. Intra-examiner agreement for four lists of the NST was 94% for the primary examiner and 92% for the secondary examiner.

The subjects used for inter-observer agreement were randomly chosen. The two examiners could not meet the 88% agreement criteria on subject number four despite repeated attempts and examiner training (72% to 84%) Inter-tester measures for subject number twelve were acceptable (88% to 92%). A third subject was randomly chosen (subject 3) and acceptable inter-observer agreement (92%-96%) was obtained on the first trial. The inter- and intra-examiner agreement data reported above are from subjects three and twelve.

CHAPTER 5

DISCUSSION

The purpose of this study was to determine if dialect had an effect on children's word recognition scores by testing Montana dialect speaking children's word recognition abilities on the PBK and NST. The tests were presented by a Montana dialect and a General Southern dialect speaker. There was only one significant difference apparent in the results of this study. Results of the NST test showed that the Montana children did significantly poorer when the test was presented by the Montana speaker, who used the more familiar dialect. These results suggest that differences in word recognition scores cannot necessarily be attributed to the dialect, but may be a result of individual factors inherent in the Montana speaker's pronunciation.

Possibly some factors inherent in the tester's speech caused them to be more or less intelligible. The literature, as cited earlier in this paper, supports the possible presence of individual speaker effects. Penrod (1985) indicated that one of the most prominent linguistic variables affecting speech discrimination scores was individual speaker articulation. When John Palmer (1955) looked at speaker gender as a possible factor affecting word recognition scores, he postulated that articulatory characteristics of the speaker may have contributed to differences in performance with different speakers.

In addition to possible speaker factors affecting the word recognition scores, subject factors, such as level of attention, may

have contributed to the differences observed between dialects. The Montana dialect and General Southern dialect were phonemically similar. However, the General Southern dialect was presumably a more novel stimulus to the subjects, therefore, they may have been more attentive during the General Southern dialect presentation of the NST.

NST scores obtained in this study were compared to the normative scores expected for seven year old children on the NST (Danhauer et al., 1985). In Danhauer, Lewis, and Edgerton's (1985) study, the seven year old subjects scored between 88% and 99% at 55 dB SL, comparable to the 55 dB HL level used on normal hearing subjects in this study. Danhauer et al. used the phoneme method of scoring (1 point for each phoneme correct). Tapes of the responses from this study were scored using the phoneme method (as opposed to original computations based on percent of words correct) so they could be compared to the NST norms. The subjects' scores with the Montana speaker ranged from 75% to 99% correct. The General Southern speaker yielded scores between 87% and 99% correct with the phoneme method.

Several of the scores obtained with presentation by the Montana speaker were below the normal range of scores for seven year old children, while with the General Southern presentation, only one subject scored slightly (1%) outside the normal range. The differences in scores suggest that some factor(s) in the Montana speaker's presentation made that version more difficult. Since content, presentation order, and intensity were controlled for, and since individual articulation characteristics can affect word recognition scores, individual articulation is a likely contributing factor.

Due to possible contamination by individual speaker characteristics, it cannot be determined whether speaker dialect affected the word recognition scores in this study. One way to control for the speaker differences would be to use several speakers with equal training and experience from each of the two dialects. If differences in scores are a result of individual articulation characteristics, then the results should show a variation of scores for materials presented by specific speakers. On the other hand, if differences in word recognition performance are a result of dialect, there should be a significant pattern of errors for the speakers of one dialect.

While the proposed study would help to separate the effects of Montana versus General Southern dialect from individual speaker characteristics in normal hearing children, other questions remain about the effect of dialect on word recognition scores. It is still not known if dialect would have an effect on different populations, for example, a hearing impaired or learning disabled population. Normally hearing subjects make use of semantic closure cues (Elliott et al., 1979), which may help them compensate for dialectal effects. Hearing impaired and learning disabled subjects may not be able to use those compensation skills (Thornton and Raffin, 1978). Therefore, smaller individual differences in speakers or dialects may cause significant effects in these children's word recognition scores. One option for continued research would be to determine the effects of dialect on the word recognition scores of hearing impaired, learning disabled, or other populations with poor semantic closure capabilities.

Additionally, the two speakers chosen for this study did not significantly differ in their pronunciation of most of the stimulus items. Research with more contrasting dialects or with speakers who have English as a second language may reveal differences caused by dialect or accent.

Several questions remain with regard to the effects of dialect on children's word recognition scores. Studies are necessary to probe the effects of dialect on hearing impaired, learning disabled, and other populations. In addition, studies using speakers with more distinctly different dialects or speakers with English as a second language are suggested.

APPENDIX A

INFORMAL ASSESSMENT OF PHONOLOGICAL DEVELOPMENT

Questions

1.) What is your name?

2.) What did you do this/last weekend?

3.) Tell me about some of the things you get to do in school.

4.) What kinds of things do you like to do when you aren't in school?

Pictures

grapes

watch

clock

sheep

smoke

chain

feather

lock

three

slide

APPENDIX B

PARENT QUESTIONNAIRE

Dialect exposure

How many years has your child lived in Montana? _____.

Is your child exposed to any non-Montana dialects from members of the immediate family? _____. If so, please explain. _____

_____.

Classroom Attendance

What grade does your child currently attend at school? _____.

Does your child receive any special services outside of his/her regular classroom (either remedial or gifted)? _____.

APPENDIX C

RELEASE/CONSENT FORM

Dear Parent:

The University of Montana Speech, Hearing, and Language Clinic is doing a study that involves the hearing abilities of young children. The study will be examining any effects on children's hearing test scores when the tests are presented by speakers of different regional dialects. The study requires children who have hearing within normal limits, normal health and development, attendance in a regular classroom, a Montana dialect with no influence from other dialects within the immediate family, and who are between the age 7 years 0 months to 7 years 11 months.

Each child will receive a brief phonological development screening (i.e.; to determine if they can correctly produce all of the sounds expected for their age) and a hearing screening. The children will listen to words through earphones and then repeat them back to the examiner. The total test time will be about 30 minutes and will be done in the University of Montana Speech and Hearing facilities. All procedures used are standard clinical procedures and do not pose any risk to your child. Testing will be done at a convenient time for you. Permission to use your child in this study is greatly appreciated. Please fill in the information below and have your child return it to his/her teacher. You or your child have the option to discontinue your child's participation at any time during the procedure if you so choose. Please do not hesitate to contact me or Sally Johnson at 243-4131 or 251-3586 if you have any questions.

Sincerely;

Michael A. Crews, B.A.

Yes, I give permission for my child to take part in the study.

Child's Name _____

Child's age _____

D.O.B. _____

School _____

Telephone number _____

Signed _____ (Parent)

Date _____

Appendix D

Phonetically Balanced Kindergarten 50 (PBK) test lists

PBK 1		PBK 2	
1.	please	1.	pants
2.	great	2.	pinch
3.	sled	3.	bus
4.	pants	4.	bad
5.	rat	5.	five
6.	bad	6.	teach
7.	pinch	7.	fold
8.	such	8.	hunt
9.	bus	9.	great
10.	need	10.	no
11.	ways	11.	rag
12.	five	12.	fed
13.	mouth	13.	please
14.	rag	14.	rat
15.	put	15.	box
16.	fed	16.	is
17.	fold	17.	mouth
18.	hunt	18.	put
19.	no	19.	slice
20.	box	20.	are
21.	are	21.	ways
22.	teach	22.	such
23.	slice	23.	sled
24.	is	24.	tree
25.	tree	25.	need

Appendix E

Nonsense Syllable Test (NST)

NST 1		NST 2	
1.	ðɪjo	1.	ʤʌðʊ
2.	gɾðo	2.	haðo
3.	ðɹwo	3.	ðɪjo
4.	fʊθɪ	4.	ʃeθa
5.	ʃɪθɹ	5.	pitɪ
6.	fisɹ	6.	ʃɪθɹ
7.	vʊvɪ	7.	sɛfɛ
8.	ʤɹfe	8.	ðɪlɹ
9.	ðuma	9.	ʤɹfe
10.	sɛfɛ	10.	ʃiʧa
11.	ʃeθa	11.	pʌðʊ
12.	ðɪlɹ	12.	nɪθʊ
13.	bʌθɪ	13.	ðana
14.	ðana	14.	ʧoθɹ
15.	bɹðɹ	15.	bɹðɹ
16.	ðʊmɹ	16.	ðɹwo
17.	ʃiʧa	17.	fisɹ
18.	pʌðu	18.	giðo
19.	ʧoθɹ	19.	θʌfɪ
20.	pitɪ	20.	ðʊmɹ
21.	ʤʌðʊ	21.	daθa
22.	θʌfɪ	22.	bʌðɪ
23.	haðo	23.	ðuma
24.	daθa	24.	fʊθɪ
25.	nɪθʊ	25.	vʊvɪ

REFERENCES

Adler, S. (1973). Data gathering: The reliability and validity of test data from culturally different children. Journal of Learning Disabilities, 6(7), 30-35.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985). Standards for Educational and Psychological Testing.

Beattie, R.C., Svihovec, D.A., and Edgerton, B.J. (1978). Comparison of speech detection and spondee thresholds and half-versus full-list intelligibility scores with MLV and taped presentations of NU-6. Journal of the American Audiology Society, 3(6), 267-272.

Bountress, N.G. (1983). Effect of segregated and integrated educational settings upon selected dialectal features. Perceptual and Motor Skills, 57, 71-78.

Brandy, W.T. (1966). Reliability of voice tests of speech discrimination. Journal of Speech and Hearing Research, 9, 461-465.

Carhart, R., and Evanston, H.J. (1965). Problems in the measurement of speech discrimination. Archives of Otolaryngology, 82, 253-260.

Carver, C.M. (1987). American Regional Dialects. Ann Arbor: University of Michigan Press.

Committee on the Status of Racial Minorities (1985). Clinical management of communicatively handicapped minority language populations. American Speech-Language-Hearing Association, 27(6), 29-32.

Creston, J.E., Gillespie, M.A., and Krohn, C. (1966). Speech audiometry: Taped vs live voice. Archives of Otolaryngology, 83, 14-17.

Danhauer, J.L., Lewis, A., and Edgerton, B. (1985). Normally hearing children's responses to a nonsense syllable test (NST: Edgerton and Danhauer, 1979). Journal of Speech and Hearing Disorders, 50, 100-109.

Diefendorf, A.O. (1983). Speech audiometry with infants and children. Seminars in Hearing, 4(3), 241-252.

Edgerton, B.J., and Danhauer, J.L. (1979). Clinical Implications of Speech Discrimination Testing Using Nonsense Stimuli. Baltimore: University Park Press.

Elliott, L.L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. Journal Acoustical Society of America, 66(1), 12-21.

Elliott, L.L., Hammer, M.A., Scholl, M.E., and Wasowicz, J.M. (1989). Age differences in discrimination of simulated single-formant frequency transitions. Perception and Psychophysics, 46(2), 181-186.

Frank, T., and Craig, C.H. (1984). Comparison of the Auditec and Rintelmann recordings of the NU-6. Journal of Speech and Hearing Disorders, 49, 267-271.

Hood, J.D., and Poole, J.P. (1980). Influence of the speaker and other factors affecting speech intelligibility. Audiology, 19, 434-455.

Jerger, S. (1984). Speech audiometry. In J. Jerger (ed.), Pediatric Audiology. 71-93, San Diego, College Hill Press.

Jerger, S., Jerger, J., and Abrams, S. (1983). Speech audiometry in the young child. Ear and Hearing, 4(1), 56-66.

Jerger, S., Lewis, S., Hawkins, J., and Jerger, J. (1980). Pediatric speech intelligibility test. I. Generation of test materials. International Journal of Pediatric Otorhinolaryngology, 2, 217-230.

Kreul, J.E., Bell, D.W., and Nixon, J.C. (1969). Factors affecting speech discrimination test difficulty. Journal of Speech and Hearing Research, 12, 281-287.

Levy, B.B., and Cook, H. (1973). Dialect proficiency and auditory comprehension in standard and black nonstandard english. Journal of Speech and Hearing Research, 16, 642-649.

Martin, F.N., and Gravel, K.L., (1989). Pediatric audiologic practices in the united states. The Hearing Journal, 42(8), 33-48.

Miller-Jones, D. (1989). Culture and testing. American Psychologist, February, 360-366.

Northern, J.L., and Downs, M.P. (1984). Hearing in Children. Baltimore: Williams & Wilkins.

Olsen, W.O., and Matkin N.D., (1979). Speech audiometry. In W.F. Rintelmann (ed.), Hearing Assessment, 172-196. Baltimore, University Park Press.

Palmer, J.M. (1955). The effect of speaker differences on the intelligibility of phonetically balanced word lists. Journal of Speech and Hearing Disorders, 20, 192-195.

Penrod, J.M. (1985). Speech discrimination testing. In J. Katz (ed), Handbook of Clinical Audiology. 235-255, Baltimore, Williams and Wilkins.

Penrod, J.P. (1979). Talker effects on word-discrimination scores of adults with sensorineural hearing impairment. Journal of Speech and Hearing Disorders, 44, 340-349.

Ralph, J.B. (1967). Language and speech deficits in culturally disadvantaged children: Implications for the speech clinician. Journal of Speech and Hearing Disorders, 32, 203-214.

Reed, C.E. (1967). Dialects of American English. Cleveland: The World Publishing Company.

Resnick, S.B. (1984). Speech recognition testing: Nonconventional testing techniques and applications to children. ASHA Reports, 14, 65-69.

Sanderson-Leepa, M.E., and Rintelmann, W.F. (1976). Articulation functions and test-retest performance of normal-hearing children on three speech discrimination tests: WIPI, PBK-50, and NU Auditory Test No. 6. Journal of Speech and Hearing Disorders, 41, 503-519.

Tse, S., and Ingram, D. (1987). The influence of dialectal variation on phonological acquisition: a case study on the acquisition of Cantonese. Journal of Child Language, 14, 281-294.

Taylor, O.L., and Payne, K.T. (1983). Culturally valid testing: A proactive approach. Topics in Language Disorders, 3(3), 8-20.

Thornton, A.R., and Raffin, M.J.M. (1978). Speech discrimination scores modeled as a binomial variable. Journal of Speech and Hearing Research. 21, 507-518.