DOCUMENT RESUME

ED 125 004 95 CS 501 410

AUTHOR Rudegeair, Robert E.
TITLE Linguistic Factors in the Assessment of Speech Sound Discrimination. Report No. 52.
INSTITUTION Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO SWRL-TR-52
PUB DATE 25 Apr 75
NOTE 27p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Auditory Discrimination; *Aural Learning; Educational Research; Elementary Education; *Linguistics; *Listening Comprehension; *Listening Tests; *Measurement Instruments; Technical Reports; Test Construction
IDENTIFIERS Southwest Regional Laboratory; SWRL

ABSTRACT

The results of recent studies are presented which show that traditional auditory discrimination tests for children which require the examinee to distinguish one speech sound from another are ill-conceived and consequently of little practical value. Linguistic variables requiring attention in designing useful speech sound discrimination instruments are discussed. (Author)

SWRL

SWRL EDUCATIONAL RESEARCH AND DEVELOPMENT

TECHNICAL REPORT 52

APRIL 1975

LINGUISTIC FACTORS IN THE ASSESSMENT OF SPEECH SOUND DISCRIMINATION

Robert E. Rudegeair

ABSTRACT

The results of recent studies are presented which show that traditional auditory discrimination tests for children which require the examinee to distinguish one speech sound from another are ill-conceived and consequently of little practical value. Linguistic variables requiring attention in designing useful speech sound discrimination instruments are discussed.

CONTENTS

3

LINGUISTIC FACTORS IN THE ASSESSMENT OF SPEECH SOUND DISCRIMINATION

Robert E. Rudegeair

Auditory discrimination tests are designed to assess children's ability to discriminate one speech sound from another. The major applications for such tests in educational and clinical settings are based on correlational studies of performance involving auditory discrimination test scores, measures of articulation, and early reading achievement. While the literature is ambiguous, the notion has nonetheless prevailed that inadequately developed auditory discrimination is a causal factor in articulation disorders as well as reading disabilities. In this paper the results of recent studies are presented to show that traditional auditory discrimination tests, on which the correlations cited are dependent, are basically ill-conceived and consequently of little practical value.

The uncritical acceptance of traditional auditory discrimination assessment procedures by the educational and speech research communities has resulted in a grossly underestimated appraisal of normal auditory perceptual proficiency in school-ready children (see, e.g., Beving & Eblen, 1973). More serious, however, is the gap created in the data base on which models of language acquisition are founded. Sensitive measures of speech sound perception are necessary to adequately diagnose specific language problems in children and to enhance our understanding of the normal course of events in the development of psycholinguistic processes.

TRADITIONAL TESTS OF AUDITORY DISCRIMINATION

The prototype for standard tests of auditory discrimination was devised by Travis and Rasmus (1931). The test is comprised of 331 pairs of contrasting nonsense syllables (e.g., /ta/-/da/) and 35 noncontrasting pairs (e.g., /ta/-/ta/). Subjects are asked to respond "same" or "different" to each pair as it is presented orally. Subsequently the Travis-Rasmus Speech Sound Discrimination Test was used in a study by Hall (1938). Both studies compared speech sound discrimination performance of normal subjects with that of articulatory defective subjects. Both sampled several age levels from kindergarteners to adults. Travis and Rasmus found that experimental group performance on the discrimination measure was significantly poorer at all age levels. Hall found the opposite.

The unwieldy list of almost all possible English speech sound contrasts has not been used in subsequent testing. Since the prototype experiments, speech sound discrimination tests have been confined, almost exclusively, to syllable pairs that represent minimal phonemic differences. Minimal pairs are understood to mean those pairs of speech sounds that contrast with regard to one or sometimes two articulatory features. The transition to the restricted set of contrasts was made gradually, however, as evidenced by the original speech sound discrimination test designed by Templin (1943). Her test consists of 200 items, 68 consonant contrasts in syllable initial position, 68 in syllable final or medial position, and 64 non-contrastive syllable pairs. This lengthy test was compared to a "short test" of 70 items and favorable correlations were demonstrated. Thus, Templin concluded that the short test was an acceptable assessment instrument. The 70-item test, consisting of 51 contrasting and 19 non-contrasting syllables became the first test to focus on minimal pairs. Although Templin's test involves nonsense syllables and relies on same-different judgments, it differs in several respects from the Travis-Rasmus model. While Travis and Rasmus used only CV syllables, where the vowel was /a/, to test consonant discrimination, Templin used CV, VC, and VCV syllables. She also varied the vowel, using three vowels and a diphthong in constructing her syllables. Furthermore, no vowel contrasts are presented in the Templin test. In addition to Templin's own work, the short test discussed here and a shorter 50-item test (Templin, 1957) have been used by several other investigators to study the relation between articulation abilities and phonemic discrimination abilities (Kronvall & Diehl, 1954; Cohen & Diehl, 1963; Aungst & Frick, 1964; Sherman & Geith, 1967).

The most frequently used auditory discrimination test (in correlational studies) has been the Wepman test of Auditory Discrimination (Wepman, 1958). Wepman's test contains 40 real-word pairs, 10 "same" pairs, and 30 "different" pairs. Only errors on "different" pairs count as discrimination errors. "Same" pairs serve to keep a subject honest and thereby constitute a check on the validity of the test (Wepman, 1960). Both members of a contrasting pair are equated for frequency of occurrence according to the Thorndike-Lorge (1944) count. The Wepman Test of Auditory Discrimination is, like Templin's short test, made up of minimal pair contrasts only. But Wepman's test presents a restricted sample of all possible minimal pairs. All consonant contrasts on the test represent contrasts among stops and contrasts among fricatives with regard to place of articulation. In addition to consonant contrasts (in initial and final position), four vowel contrasts are also included. Wepman (1960), in a preliminary report of the findings from studies using his test, concluded that there is a definite relationship between faulty articulation and poor discrimination, but no data were presented.

The Wepman Test has been used in attempts to define relationships between auditory discrimination abilities and articulation proficiency (Prins, 1963), beginning reading achievement (Christine & Christine, 1964; Sivaroli & Wheelock, 1966; Blank, 1968), and differences in dialect (Deutsch, 1964; Coller, Coleman, & Schwartz, 1968; Deutsch, 1972; Elenbogen & Thompson, 1972). It is not intended in this paper to discuss the merits of these studies or the interdependencies or correlations among various linguistic processes such as those cited. The literature is referenced only to illustrate that traditional auditory discrimination tests, in addition to their widespread clinical use, are heavily relied upon in educational and linguistic research. It appears, however, that the overwhelming concern with correlations between discrimination ability and various other language processes has distracted the research community from addressing fundamental questions about speech sound discrimination per se. The long-standing assumption has been that since speech sound contrasts are presented, speech sound discrimination is being tested. The results of several recent studies call this assumption into question and emphasize the need to reevaluate prevailing notions of speech sound discrimination.

COGNITIVE FACTORS IN AUDITORY DISCRIMINATION ASSESSMENT

The same-different technique typically used in assessing speech sound discrimination has been critized on several grounds. Vellutino, DeSetto, and Steger (1972) hypothesized that discrimination tasks with unequal response alternatives have a built-in response bias because of an intrinsic tendency to employ response categories (e.g., "same" and "different") with equal frequency. These investigators tested this hypothesis using the Wepman Auditory Discrimination Test which is comprised of 30 "different" and 10 "same" items. Half of their sample of fourth, fifth, and sixth graders were administered the standard Wepman and the other half were administered a revised version of the test in which same and different items were equalized. The group presented with equal numbers of response categories made significantly fewer errors, leading Vellutino et al. to conclude that "the Wepman and instruments with a similar format may be characterized by a substantial degree of variability unrelated to individual differences in auditory discrimination..." (p. 255). In an earlier study, Brière (1967) also reported finding a bias for responding "same" when the stimuli were different. Findings such as these raise serious questions about the interpretations given to error data gathered in the context of the traditional testing format.

Further criticism has been leveled against the use of the same-different paradigm because it demands an operational understanding of the concepts "same" and "different." Beving and Eblen (1973) tested 30 children 4- to 7- years-old to evaluate the influence of the concepts "same" and "different" on speech sound discrimination performance at various age levels. Three groups, 4-, 6-, and 8-year-olds, were given a 25-item discrimination test under the same-different format. Three days later, the same children were given the same items and asked to repeat the syllable pairs in lieu of responding "same" or "different." From the responses in this second task, same-different judgments were inferred by recording whether a child reported two words that were the same or two words that were different. No regard was paid to repetition accuracy per se. Results showed that while all three groups exhibited significantly different error rates on the judgment test, performance on the repetition test was equivalent across the three groups (error rate = approximately 8%). But only for the 4-year-old group was a significant difference found between the judgment scores and the repetition scores, judgment errors being significantly more numerous. It was concluded that while 4-year-olds were able to discriminate among speech sounds on a par with 6- and 8-year-olds, as shown by their performance on the repetition task, they appear to be unable to work with the concepts "same" and "different."

In an earlier study, Blank (1968) sought to determine whether differences in auditory discrimination scores between good and poor readers were a function of failure to discriminate or a failure related to the complex cognitive processes demanded by the same-different task. Her study consisted of three experiments. In the first experiment, two groups of 7-year-olds, classified as good and poor readers, exhibited differential scores on a same-different auditory discrimination task. Nine children in the original group of poor readers had to be eliminated because of a failure to understand the task. With regard to this point, Blank suggested that "the need to make a cognitive judgment of 'same-different' posed a problem for the retarded reader which went beyond the perceptual demands of the task." Thus, the second experiment was designed to eliminate the intervening cognitive judgment of "same-different" by having the child report directly the word pairs he heard, as in the Beving and Eblen study. When "same-different" judgments were inferred on the basis of what the child reported, the good readers still exhibited significantly lower error rates. This in in accord with Beving and Eblen's finding that 6- and 8-year-olds showed no improvement when repetition errors were compared to judgment errors.

While it appeared that the cognitive judgment of same or different did not in itself cause a problem for the poor readers, an analysis of the kinds of errors made by Blank's two groups still indicated that task strategy factors rather than perceptual abilities accounted for differential performance between the groups. Poor readers, in the repetition data, showed an overwhelming tendency to perseverate pairs (i.e., match the second member of the pair to the first). Good readers did not exhibit a strong tendency in this regard and therefore made fewer errors. The high frequency of perseverations suggested a lack of attention across pair members and led Blank to her third experiment. This experiment was designed to remove the conditions for perseveration bias and still test accuracy of perception. Thus, the children in Experiment 3 were required to repeat single words. The first members of pairs employed in Experiments 1 and 2 were presented, one at a time, followed by the words that were originally second members. Each group exhibited a mean rate of correct repetition of 85%, suggesting that the groups are not differentiated as a function of speech sound perception abilities. Cognitive demands of the task, extraneous to speech sound perception, appear to account for differential performance on a measure of speech sound discrimination.

Other studies have also questioned the value of standard speech sound discrimination tests as measures of perceptual processes. These tests are typically administered in a single test session and are presumed to be so straightforward and simple that training procedures are unnecessary. Yet, in the case of the Wepman Test of Auditory Discrimination, it has been shown that significantly better performance occurs when 6-year-olds repeat the test (Rudegeair & Kamil, 1970). Berlin and Dill (1967) also reported improved performance when feedback and positive reinforcement were provided during a second administration of the Wepman test. In spite of overwhelming evidence that traditional test procedures preclude sensitive measures of perceptual abilities, these procedures continue to be used to diagnose auditory perceptual deficits. The notion that nothing can be simpler than asking the child if two spoken words are the same or different prevails. But, as the Blank study demonstrates, a simple imitative articulation task provides a purer measure of perceptual accuracy than does the test requiring same-different judgments. The imitation task, we can assume, is not trustworthy because misarticulations and misperceptions are confounded. Crucial to the very definition of an auditory discrimination task is that it bypasses articulatory responses which in themselves cannot be used to substantiate perceptual inadequacy.

LINGUISTIC FACTORS IN AUDITORY DISCRIMINATION ASSESSMENT

EXPERIMENT 1: MINIMALLY AND MAXIMALLY DIFFERENT CONTRASTS

While it is not immediately necessary or even possible to isolate each and every process or skill that plays a role in the traditional paradigm of auditory discrimination assessment, it is important to know to what extent perceptual confusability of speech sounds is actually being measured. In an effort to establish a means for making such an appraisal, an experiment in speech sound discrimination was conducted using, as test items, all possible English consonant contrasts.

While it is usual to test only minimally distinct consonant contrasts, it was hypothesized that maximally different contrasts, i.e., those that differ by four, five, and six features, can serve as an appropriate control condition to support the claim that perceptual confusability has been measured. Since consonant confusability is, in general, a function of the number of feature differences involved in a contrast (Tikofsky & McInish, 1968; Graham & House, 1971), any test sensitive to perceptual confusion should yield more errors on minimal consonant contrasts than on maximal consonant contrasts. If subjects' error rates are equivalent across minimal and maximal categories and at the same time beyond the range of chance, it would seem safe to conclude that such error rates are due to factors other than inadequate speech sound discrimination ability. To maintain the traditional procedure of presenting two stimulus syllables for comparison and, at the same time, avoid the problem of same-different judgments, a forced-choice, matching-to-sample (A-B-X) procedure was employed.

Method

Participants. Twenty first-grade pupils, randomly selected from three classrooms in an elementary school, and eight pupils from a 4-year-old class at a preschool served as Ss. All Ss were normal-hearing, monolingual speakers of English.

Procedure. The test sessions were held with individuals in a soundproof experimental trailer set up on the school grounds. Test items, consisting of two contrasting syllables, were presented in a forced-choice, matching-to-sample (A-B-X) procedure. Participants were seated midway between the two speakers of a stereo tape recorder. A warning signal (1000-cycle tone) followed by the first member of the contrast pair (syllable A) was heard over the left speaker. One second later the second member of the contrast pair (syllable B) was heard over the right speaker. One second later "who said X?" (where X is either A or B) was heard stereophonically. Three seconds elapsed before the warning signal initiated the next trial. Subjects responded by pointing to one of the speakers. The experimenter, who was seated behind S at all times, recorded all responses immediately on prepared data sheets.

Stimuli. All consonants of English, i.e., phonemes whose feature specification includes [+ consonantal] and [- vocalic], were paired with one another, each in combination with the vowel /a/, for a total of 171 contrasting syllables. For reference, all contrasts categorized by distinguishing features are presented in Appendix A.

All contrasts were in the syllable-initial position with the exception of those involving /ž/ or /ŋ/. In these cases the contrast occurred in syllable-final position (e.g., /až/ vs /aŋ/). Nine dummy items were constructed from liquid and glide combinations to make 180 pairs. These were randomly divided into six lists, 30 items per list, which were tape recorded in a sound studio by the experimenter. Because a pupil only responded once to each item and because it is necessary to counterbalance the occurrence of a given item between A-B-A and A-B-B instances, two sets of tapes were made. Any item appearing in A-B-A form in one set of tapes appeared in A-B-B form in the other. Half the pupils in an age group got one set of tapes and the other half the other set. Pupils were tested over 7 successive school days, 1 day of training followed by 6 days of testing. Training was achieved by selecting one of the test tapes at random and leading the child through the actual test conditions with appropriate instructions and feedback. On test days, the experimenter merely said "good" after each pointing response by the child.

Results

For the analysis of the error data, the test pairs (171 consonant contrasts) were categorized according to the number of features distinguishing the contrasting consonants. Since consonants can contrast on one to six features (see Appendix A), 6 categories resulted. Figure 1 displays the mean error rates for each of the age groups as a function of the number of contrasting features.

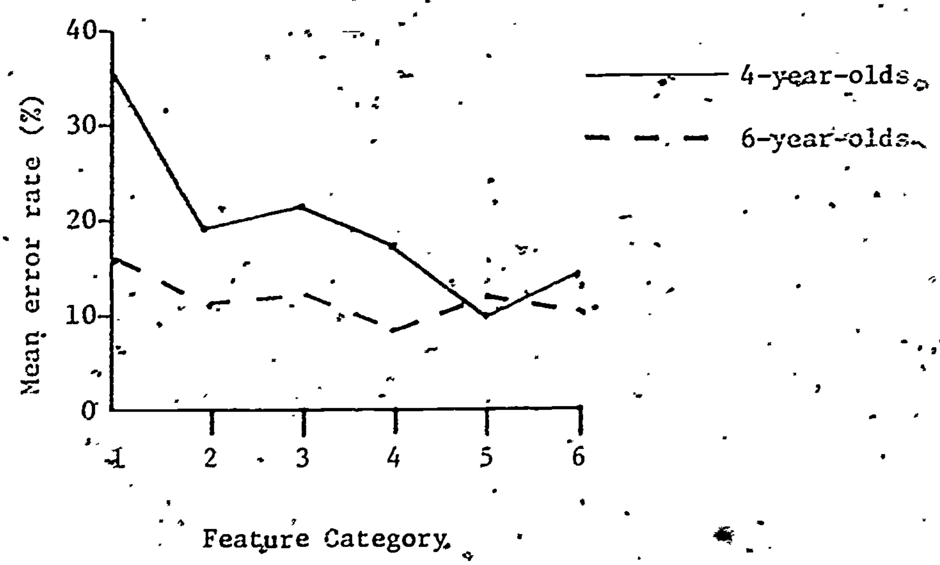

Fig. 1. Mean error rates plotted for 4- and 6-year-old groups according to the number of features distinguishing the consonants in a contrast

For the 6-year-olds, the overall error rate (171 items) was 12.1%, with a range among feature category means from 9% to 16%, and a range among pupil means from 3% to 28%. For the 4-year-olds, the overall error rate was 19% with a range among feature category means from 9% to 35%, and a range among pupil means from 12% to 27%. The source table obtained from a two-way analysis of variance (subjects within group [4- or 6-year-olds] by feature category [1 through 6]) is presented in Table 1.

Four-year-olds made significantly more errors than 6-year-olds. Performance on discriminations at different levels of feature differences was not equal, and this inequality interacted significantly with age group. These three results, as shown by subsequent paired comparisons, can all be accounted for by the same datum, viz., that the 4-year-old group exhibited significantly more errors on contrasts distinguished by only one feature. Paired comparisons were made among

Table 1

Source Table for 2-Way ANOVA: Ss within 2 Groups x 6 Feature Categories

Source of Variation	df	F
Between Ss		
Groups	1	6.29*
Ss/Groups	26	--
Within Ss		
Feature Category	5	5.86**
Groups x Feature Category	5	2.68*
Ss/Groups x Feature Category	130	--

*p<.05.
**p<.01.

cell means according to procedures for unequal group sizes presented in Winer (1971, p. 603). These tests show no reliable differences among feature category means for the older group. The only reliable difference found for the younger group was between the first feature category (where consonants contrast on only one feature) and all others. Furthermore, the only category in which the younger group made significantly more errors than the older group is the first. In other words, both main effects and the interaction effect revealed by the two-way ANOVA are a function of the 35% error rate for 4-year-olds on minimal-pair contrasts (see Figure 1).

Discussion

The finding that only one-feature contrasts cause significant discrimination problems is reinforced by the results of several previous studies. Tikofsky and McInish (1968), in an experiment where all possible contrasts among 15 consonants were tested on 7-year-olds, concluded "that as the number of feature differences increases, error rate will drop. In fact, any increase beyond the one feature difference assures almost perfect discriminability in terms of this experiment" (p. 62). Graham and House (1971), who tested children ranging in age from 3 to 4.5 on all possible contrasts among 16 consonants, elicited error patterns as depicted in Figure 2. These authors report that an analysis of variance and Newman-Keuls procedures showed that the error rates for one- and two-feature differences were reliably different while two- through six-feature contrasts did not differ. Graham and House concluded "that discrimination improved with the addition of a second feature, but did not improve significantly with further additions of features" (p. 563).

In the present study, two distinct age groups were subjected to the same test format and testing conditions. Previous studies either tested age-group differences and ignored featural analysis (Templin, 1957; Deutsch, 1964) or vice versa (Tikofsky & McInish, 1968; Graham & House, 1971). The significant interaction between age group and feature category reflects the fact that the 6-year-olds, as a group,

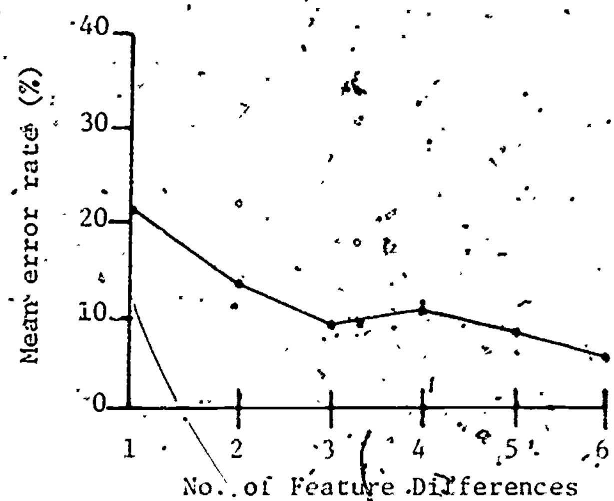

Fig 2. Mean error rates from Graham and House (1971) Ss according to the number of features distinguishing the consonants in contrast.

have no more difficulty discriminating minimal contrasts than contrasts distinguished by four, five, or six features. In other words, these data suggest that they have no discrimination difficulty at all. This conclusion is based on the supposition that perceptual confusion errors will be distributed unequally across feature difference categories (the lower the number, the higher the error rate), while error rates due to task constraints (e.g., attention problems or storage and retrieval problems) will be distributed equally among feature difference categories.

At two, three, and four-features different, the performance of 4-year-olds was not significantly different from the performance of 6-year-olds, but it did exhibit a predictable pattern--error rates in these categories were higher for 4-year-olds. At five-features different, however, there was clearly no difference in performance between the groups. Contrasts that are six-features different are not worthy of discussion since there are not enough of them to justify any comparisons. Five-feature differences appear to represent a clear indication of baseline performance. The data suggest that error rates at this level are more likely attributable to task constraints than perceptual confusion. Since no reliable meaning can be attributed to scores achieved on currently existing auditory discrimination tests, the needs of those who rely on such tests can be served by suggesting meaningful ways to improve test design. With little apparent trouble, control items from maximal contrast categories can be inserted into auditory discrimination tests. Assumptions about perceptual confusion problems can be tested by appropriate comparisons between error rates on minimal and maximal pairs.

The present approach suggests a different view of "auditory discrimination deficit" than is found, for example, in the Wepman Manual of Directions (1958). There it is asserted that a 5-year-old who makes more than six errors (on "different" items) is revealing "inadequate development." Six-year olds who make more than five errors are placed in the same category. Children 8 years of age or older are allowed three errors before being judged deficient in auditory discrimination skill. Based on the data presented in Table 2, it is more likely that any inadequacy lies in the criteria established for the Wepman test. In Table 2, error data from several studies in which the Wepman test was used are presented. The error data contained in the table reflect scores achieved by control groups only since non-normal behavior is not at issue.

Table 2

Error Data for Control Groups from Various Studies in which the Wepman Test of Auditory Discrimination Was Administered

Study	No. of Ss	Age Range	Mean No. of Errors	Wepman Criterion
Elenbogen (1972)	30	5-6	5.2	6
Sivaroli & Wheelock (1966)	120	5-6	9.3	6
[illegible]ller et al. (1968)	128	6-7	9.8	5
Prins (1963)	19	6-7	5.6	5
Rudegeair & Kamil (1970)	12	6-7	5.4	5
Deutsch (1964)	8	6-7	7.8	5
Berlin & Dill (1967)	11	8-9	5.9	3
Deutsch (1964)	8	8-9	5.5	3
Duetsch (1964)	8	10-11	6.9	3
Vellutino et al. (1972)	60	10-13	3.03	3

Age of Ss, number of Ss and mean number of errors (on "different" pairs) provide enough information to conclude that the average child in each group fails to meet the test's own standard for adequate development. With the exception of the Elenbogen study, the mean number of errors for Ss in each of these studies exceeds the value designated as a cut-off point for the age group in question.

The Wepman test is typical of currently used tests in that all rely on a global discrimination score. These tests are apparently deemed useful because they are easy to administer and score. A compromise between an ambiguous global score and complicated subsets of scores appears attainable with a list that combines minimal and maximal contrasts. Discrimination difficulty or "deficit" in this context is represented by the difference in performance between the two item types. Not only is such a test relatively easy to score, but it offers some guarantee that the measure obtained is a function of perceptual confusion.

EXPERIMENT 2: MEANINGFUL AND MEANINGLESS STIMULUS SYLLABLES

While it has been firmly established that extra-linguistic factors play a part in speech sound discrimination test performance, little attention has been paid to the competing linguistic factors inherent in discrimination tests. In addition to the speech sound comparisons required in same-different and A-B-X tests, the meaningfulness of the syllables used to present the contrasts appears to have an important influence on test results. In the view of some investigators, auditory discrimination can be measured by presenting, in succession, two familiar words that differ with regard to one of their constituent phonemes and asking the child to judge them "same" or "different." In an opposing view it is held that meaningful words are not appropriate for such a test because those children in the sample familiar with the words have an advantage over those to whom the words are less familiar. Thus, it is argued, a more sensitive measure of speech sound discrimination can be achieved by using nonsense syllables because they are a means of controlling for familiarity. Implicit in this strategy, however, is the notion that children are capable of processing speech sounds out of context--or out of any meaningful context. This suggests that speech sounds are, in some sense, functional perceptual units which can be processed in and of themselves. In this experiment an attempt was made to test this assumption because of its significance for the design of auditory discrimination tests. In addition, whether children can or cannot process phonological data as such has a bearing on the linguistic nature of the discrimination task involving nonsense syllables. Previous studies comparing children's ability to discriminate sounds in meaningful versus meaningless contexts have been inconclusive.

Sapir (1972) hypothesized that "children's performance in auditory discrimination is significantly better when words rather than nonsense syllables are used in the task." Yet no notion of why this might be true was provided. Her data from 150 children ranging in age from 5 to 6.7 showed no reliable differences between the two stimulus types. Sapir suggested, by way of explaining this result, that children may not be able to process meaning when the word is separate from its context, implying that all items on the test were, in effect, meaningless. This suggests that if real words were found easier than nonsense syllables, the "meaning" factor would account for the results.

Perozzi and Kunze (1971) intended to demonstrate that word and nonsense tests were equivalent in their measurement properties so that the use of syllable tests, though apparently artificial, could be justified. The advantages of the nonsense-syllable tests, according to Perozzi and Kunze, lie in the potential they offer for constructing a test with controlled degrees of difficulty. Apparently, this potential is a function of the fact that nonsense syllables allow testing of a wide range of contrasts, whereas real words allow testing of only those minimal sound contrasts that actually occur in the lexical inventory of children.

In discussing why the two stimulus types may elicit different discrimination performance, Perozzi and Kunze cite Weiner's (1967) argument that paired-word tests are less abstract than paired-syllable tests because the paired words signal a difference in meaning as well as a difference in sound. Neither Weiner nor Perozzi and Kunze discuss the implications of this two-level signalling for a psycholinguistic model of sound discrimination behavior. Perozzi and Kunze's 30 kindergarten pupils, who ranged in age from 5.3 to 6.5, showed a mean of 71.2 correct responses on the Syllable Test, compared with 73.6 on the Word Test. No test of significance was reported. The Pearson-Product correlation between the two tests was 0.873 ($p < .01$) and the investigators concluded that the performance on one test would be highly predictable from performance on the other.

While both the Sapir, and Perozzi and Kunze studies show no reliable differences between words and nonsense syllables, studies by several investigators confirmed such a difference. Two of these, Blank (1968) and Katz and Deutsch (1967), tested Hebrew versus English stimuli and thus represent more than the simple word versus nonsense experiment. The introduction of foreign sounds represents a confounding factor. Elenbogen and Thompson (1972), however, used the standard Wepman Test of Auditory Discrimination (1958) and a test they called the "Distorted Wepman" in which the real words in the original Wepman were distorted into nonsense syllables while the same contrasts were maintained. In a repeated measures design involving a group of suburban, Anglo children of middle socioeconomic status and a group of urban Black children from a low socioeconomic range, these investigators found that error rates were no different for Anglos and Blacks on the distorted Wepman and Blacks on the standard Wepman. But error rates for Anglos on the standard Wepman were significantly lower than the other scores. The investigators concluded that middle class children have difficulty discriminating minimal pairs when meaning is removed from the carrier syllables. They further suggest that the "vocabulary factor" is an aid to performance on the Wepman Test of Auditory Discrimination.

Elenbogen and Thompson expand on this theme to make the point that real word pairs "would provide clues to children with extensive language backgrounds. This would detract from accurate assessment of auditory discrimination and would affect norms when children with varied language experience are evaluated with the same norms" (p. 211). This claim that real word pairs lead to an inaccurate profile of sound discrimination ability was echoed by Perozzi and Kunze when they suggested "that some youngsters may have an unforeseen advantage over others when taking paired-word tests" (p. 384).

It is ironic that the argument that meaning enhances discriminability has been made most forcefully in the reports of two studies in which no difference between real and nonsense conditions was observed. If the argument itself has substance, the obvious conclusion is that it does not apply in the case of the children tested by Sapir, and Perozzi and Kunze, perhaps because, at 5 and 6 years of age, these children were capable of successfully executing any auditory discrimination task.

Experiment 2A: Four-Year-Olds

The present study was intended to investigate relative discriminability of contrasts in meaningful and nonsense carrier syllables for children younger than age 6.

Method

Participants. Ten pupils from a 4-year old class at a public preschool were administered the test. All children were monolingual Anglos with no known hearing loss.

Procedure. The procedure followed in this experiment was identical to that used in Experiment 1, i.e., all syllable pairs were presented according to the A-B-X format.

Stimuli. Test items were 16 consonant contrasts all of which represented a difference in only one articulatory/acoustic distinctive feature. Test items are shown in Table 3. Since each list contained nonsense-syllable and real-word representatives of each contrast, there was a total of 32 test items per list. Real and nonsense items were randomly mixed in the lists. For purposes of counterbalancing the order of presentation of the two members of a contrast pair as well as the position of the queried member, four lists were required. Four tapes, each representing a different counterbalancing version of the master list, were recorded by the experimenter and administered to each participant over four successive school days. Thus, each child responded to each real-word contrast four times and to each nonsense-syllable contrast four times. Two scores were tabulated for each subject--percent error in response to real-word contrasts and percent error in response to nonsense-syllable contrasts.

Table 3

Sixteen Minimal Consonant Contrasts Employed in the Present Study in Real-word and Nonsense-syllable Forms

Contrast	Nonsense Form		Real-word Form	
pt	pa	ta	pie	tie
fθ	fa	θa	fin	thin
mn	ma	na	meat	neat
pk	pa	ka	pan	can
sš	sa	ša	see	she
mŋ	am	aŋ	rim	ring
pb	pa	ba	pig	big
td	ta	da	two	do
kg	ka	ga	came	game
fv	fa	va	fairy	very
θð	θa	ða	thigh	thy
sz	sa	za	sue	zoo
čǰ	ča	ǰa	choke	joke
sθ	sa	θa	sing	thing
zð	za	ða	zees	these
tk	ta	ka	take	cake

Results

The 10 children who responded to the experimental tapes showed a mean error rate of 34.5% for real word items and 47% for nonsense syllable items. Table 4 is a list of error rates for each child as a function of the two stimulus types. No child showed a reversal of the trend to do better in response to real-word items, although S2 and S10 showed nearly equivalent performance in response to both item types. A two-way analysis of variance (Item Type (2) x Subjects

Table 4

Mean Error Rates for Each Child as a Function of Two Stimulus Types

Subject	Real-word Items*	Nonsense Items*
S1	28	44
S2	37.5	39
S3	25	37.5
S4	37.5	48.5
S5	36	51.5
S6	39	53
S7	31	51.5
S8	42	58
S9	34.5	47
S10	34.5	37.5
Mean	34.5	47

*In percent

(10)) showed a significant difference between responses to the two item types (see Table 5). It should be noted that overall performance on nonsense-syllable pairs was at a level probably equivalent to chance (mean = 47% error rate).

Table 5

Source Table for 2-Way ANOVA (Item Type X Subjects)

Source of Variation	df	Mean Squares	F-RATIO
Item Type	1	288.7	43.08*
Subjects	9	23.3	---
Residual	9	6.7	---

*$p < .01$.

Experiment 2B: Five-Year-Olds

In most respects this experiment was identical to Experiment 2A. The differences involve minor modifications in list construction, number of test days, and the age of the children tested.

Method

Participants. Twenty pupils from a 5-year-old class at a public school were administered the test. All children were monolingual Anglos with no known hearing loss. The experimental procedure was the same as in the first experiment.

Stimuli. Test items were the same 16 consonant contrasts as in Experiment 2A. Two separate lists were constructed. One list consisted of only the real-word contrasts and the other consisted of the nonsense-syllable contrasts. Each list contained 32 items since each item appeared in both A-B-A and A-B-B form. The position of the queried item was not counterbalanced in this experiment. There were two days of testing. Ten participants (Group A) received the real-word list and 10 (Group B) received the nonsense-syllable lists on the first day. On the second day each group received the opposite list.

Results

The children tested exhibited a mean error rate of 21.2% in response to real-word contrasts and 48.7% in response to nonsense-syllable contrasts. As shown in Table 6, this differential was influenced by the reversed order of presentation between the two

Table 6

Mean Error Rates for Each of Two Groups as a Function of Days

	Day 1	Day 2
Group A	21.2% (real words)	45.6% (nonsense syllables)
Group B	51.8% (nonsense syllables)	21.2% (real words)

groups. For the purposes of the analysis of variance, the data were tabulated in terms of two factors, Groups (A and B) and Days (1 and 2). Thus, differential error rates for real-word and nonsense-syllable contrasts will appear as a significant group x days interaction. The two-way analysis of variance confirmed a significant Group x Days interaction ($F(1,18) = 17.80$, $p<.01$). As in Experiment 1, it should be noted that performance in response to nonsense-syllable contrasts was equivalent to chance (45.6 and 51.8 percent mean error rates) for both groups.

Discussion

The results of Experiments 2A and 2B clearly demonstrate differential error potentials between nonsense-syllable and real-word contrasts when preschool children are tested. For all practical purposes, the children tested in these experiments were unable to discriminate between the members of nonsense-syllable pairs. Yet this failure cannot be accounted for on the basis of the particular minimal sound contrasts presented because the same contrasts comprised the real-word tests. Nor is it reasonable to argue that the nonsense stimuli require sound discrimination while the real-word stimuli require only a discrimination of meanings. Success on either type of stimulus pairs requires that the sounds be discriminated. While meaning is a factor in one task and not in the other, different meanings are a function of different sounds and accurate sound perception is prerequisite to accurate recognition of the words themselves.

The experiments comparing discriminability of speech sounds in the context of real-word pairs and discriminability of the same speech sounds in nonsense-syllable pairs bear on the question of the form of stored linguistic data in the cognitive system of the preschool child. Two modes of stored data can be considered. Either the child has access to speech sounds as a function of stored word entries, or he has access to speech sounds per se. Such speech sound concepts can be conceived of as cognitive representations of speech sounds which underlie the organization of phonological data in memory. Word concepts can be assumed to be prior in origin to speech sound concepts.

Consider the auditory discrimination task in which contrasts are presented in the form of real-word pairs. Psychological processes already established, i.e., normal word-recognition processes, allow the child to successfully differentiate such pairs. Word concepts serve to mediate the required discrimination. Nonsense-syllable pairs, on the other hand, may or may not relate to previously developed processing skills. If the nonsense syllables are perceived as words, the same processes as discussed for real-word stimuli come into play. If meaningless syllables are perceived as such, the system either breaks down or is equipped to deal with such syllables on the basis of speech sound concepts. Successful discrimination in the case of nonsense-syllable stimuli turns on the system's capacity to recognize the onset of one syllable as "sound x" and the onset of the other syllable as "sound y" (or, in some tasks, non-x). Like word concepts, speech sound concepts mediate discrimination of contrasting pairs when stimuli are input in the form of meaningless speech sounds. This view of two independent inventories of functional perceptual units is not necessary but does represent a sufficient theoretical framework to account for the behavior observed in the experiments at issue.

The development of speech sound concepts can account for other linguistic behaviors observed in research on the acquisition of phonological competence. Phonemic segmentation, sounding-out, sound blending, and other word analysis tasks can all be seen to relate to the acquisition of the speech sound as a functional perceptual unit. While speech sound concepts are not a necessary aspect of normal phonological development, they appear to play a role, for example, in the acquisition of phonic reading skills. Speech sound concepts may, in fact, account for correlations reported between poor performance in early reading and poor performance on auditory discrimination or other word analysis tasks (Durrell & Murphy, 1953; Wepman, 1960; Dykstra, 1966; McNeil & Coleman, 1967).

Claims were made in the reports surveyed earlier that illegitimate measures of auditory discrimination would be obtained through the use of paired-word tests due to an "unforeseen advantage" for those children familiar with the words in the test. It can be countered that the ability to differentiate minimal phonemic differences can legitimately be related to paired-word tests or paired-syllable tests depending on what notion of auditory discrimination ability one adopts. If an investigator merely intends to assess the child's ability to differentiate the functional units of speech, why should it matter that meaningful word entries serve to mediate the distinctions which perhaps only have reality as a function of the lexicon? If, on the other hand, an investigator intends to differentiate among children on the basis of cognitive functions that result from linguistic experience, he may choose to verify the formation of speech sound concepts by testing the child's ability to process speech sound data as such, i.e., in the form of meaningless syllables.

To what extent traditional auditory discrimination tests involve linguistic processes is unknown and, unhappily, unquestioned. The experiment demonstrating differential discrimination potentials between real-word pairs and nonsense pairs suggests that for a test to reflect linguistic skill, it must respect the cognitive framework in which linguistic data is organized. While nonsense syllables posing as words have been used extensively with apparent success in psycholinguistic research, little is known about the young child's ability to deal with nonsense syllables per se.

While it is not disputed that traditional auditory discrimination procedures involve several linguistic levels and a network of psychological processes, it has apparently been assumed that these factors cannot adversely affect test scores. The studies discussed in this paper demonstrate that traditional auditory discrimination tests do not represent straightforward or reliable indices of speech sound discriminability. Low scores on the tests cannot be equated with inadequately developed speech sound discrimination. Alternatives to established procedures for assessing speech sound perception are clearly required. There appears to be no reason why tasks that reflect only the linguistic skills at issue are precluded. For example, presently used tasks involve a match or mismatch between successively presented syllables. Yet misperceptions of the sort that are likely to affect speech comprehension involve a mismatch between utterance and cognitive representation. A test modelled on matching-to-memory is not only more relevant but also avoids the short-term memory problems inherent in same-different and even A-B-X procedures.

APPENDIX A

ALL POSSIBLE CONSONANT OPPOSITIONS CATEGORIZED BY THE DISTINCTIVE FEATURES ON WHICH THEY CONTRAST

[illegible] feature differences

[illegible]	Anterior	Voice	Strident	Continuant	Nasal
	[illegible]	[illegible]	[illegible]	f-[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	v-[illegible]	[illegible]
	[illegible]	[illegible]		[illegible]	ŋ-[illegible]
	[illegible]	ǰ-č		[illegible]	
	[illegible]	[illegible]		š-č	
		[illegible]		ž-ǰ	
		[illegible]			
		[illegible]			

[illegible] feature differences

[illegible] Voice	Ant Voice	Strid Voice	Cont Voice	Nasal Voice
[illegible]	[illegible]	[illegible]	p-[illegible]	p-m
[illegible]	[illegible]	[illegible]-s	b-f	t-ŋ
[illegible]	[illegible]		t-[illegible]	k-n
[illegible]	[illegible]		d-[illegible]	
			č-ž	
			ǰ-š	

Cor Ant	Cor Strid	Cor Cont	Cor Nasal	Ant Strid	Ant Cont	Ant Nasal	Strid Cont	Cont Nasal
[illegible]	[illegible]	[illegible]	t-n	t-č	k-f	b-ŋ	t-s	v-[illegible]
[illegible]	[illegible]	[illegible]	d-m	d-ǰ	g-v	g-m	d-z	[illegible]
[illegible]	[illegible]	[illegible]-f		[illegible]-š	č-s			
	[illegible]	d-v		[illegible]-ž	ǰ-z			

Three Feature Differences

Cor Ant Voice	Cor Strid Voice	Cor Cont Voice	Cor Nasal Voice	Ant Strid Voice	Ant Cont Voice	Ant Nasal Voice	Strid Cont Voice	Cont Nasal Voice
[illegible]	[illegible]	p-[illegible]	p-n	t-ǰ	k-v	p-ŋ	t-z	f-m
[illegible]	g-š	b-[illegible]	t-m	d-č	g-f	k-m	d-s	[illegible]
	f-z	d-f		θ-ž	č-z			
	v-s	t-v		[illegible]-š	ǰ-s			

Cor Ant Strid	Cor Ant Cont	Cor Ant Nasal	Cor Strid Cont	Cor Strid Nasal	Cor Cont Nasal	Ant Strid Cont	Ant Strid Nasal	Ant Cont Nasal	Strid Cont Nasal
[illegible]	k-[illegible]	d-ŋ	p-s	ǰ-n	v-n	t-š	ǰ-ŋ	v-ŋ	z-n
[illegible]-ǰ	g-[illegible]	g-n	b-z		[illegible]-m	d-ž			
[illegible]-š			k-š			č-ŋ			
[illegible]-ž			g-ž			ǰ-ð			

Four Feature differences

Cor	Cor	Cor	Cor	Cor	Cor	Ant	Ant	Ant	Strid
Ant	Ant	Ant	Cont	Strid	Nasal	Strid	Strid	Cont	Cont
Strid	Cont	Nasal	Strid	Nasal	Cont	Cont	Nasal	Nasal	Nasal
Voice	Voice	Voice	Voice	Voice	Voice	Voice	Voice	Voice	Voice
[illegible]-ǰ	k-[illegible]	t-ŋ	p-z	č-ŋ	f-n	t-ž	č-n	f-ŋ	[illegible]
[illegible]-ž	g-[illegible]	k-n	b-s		[illegible]-m	d-š			
[illegible]-ž			k-ž			č-[illegible]			
[illegible]			g-š			ǰ-[illegible]			

Cor	Cor	Cor	Cor	Ant
Ant	Ant	Ant	Strid	Strid
Strid	Strid	Cont	Cont	Cont
Cont	Nasal	Nasal	Nasal	Nasal
p-š	m-ǰ	ð-ŋ	z-m	ž-n
b-[illegible]			ž-ŋ	
[illegible]				
g-z				
v-ǰ				
f-č				

Five Feature differences

Cor	Cor	Cor	Cor	Ant
Ant	Ant	Ant	Strid	Strid
Strid	Strid	Cont	Cont	Cont
Cont	Nasal	Nasal	Nasal	Nasal
Voice	Voice	Voice	Voice	Voice
p-ž	m-č	θ-ŋ	s-m	š-n
b-š			š-ŋ	
k-z				
g-s				
v-č				
f-ǰ				

Six Feature differences

Cor	Cor
Ant	Ant
Nasal	Nasal
Cont	Cont
Strid	Strid
z-ŋ	Voice
ž-m	s-ŋ
	š-m

REFERENCES

Aungst, L.F. and Frick, J.V. Auditory Discrimination ability and consistency of articulation of /r/. Journal of Speech and Hearing Disorders, 1964, 29, 76-85.

Berlin, C.L. and Dill, A.C. Effects of feedback and positive reinforcement on the Wepman auditory discrimination test scores of lower-class Negro and white children. Journal of Speech and Hearing Research, 1967, 10, 384-389.

Beving, B. and Eblen, R.E. "Same" and "different" concepts and children's performance on speech sound discrimination. Journal of Speech and Hearing Research, 1973, 16(3), 513-517.

Blank, M. Cognitive processes in auditory discrimination in normal and retarded readers. Child Development, 1968, 39, 1901-2101.

Briere, E.J. Phonological testing reconsidered. Language Learning, 1967, 17(3 & 4), 163-171.

Chistine, D. and Christine, C. The relationships of auditory discrimination to artic defects and reading retardation. Elementary School Journal, 1964, 67, 67-100.

Cohen, J.G. and Diehl, C.F. Relation of speech-sound discrimination ability to articulation-type speech defects. Journal of Speech and Hearing Disorders, 1963, 28(2), 187-190.

Coller, A.R., Coleman, R., and Schwartz, S. Initial and final consonant relationships in speech-sound tests: a discrimination or response set problem. ERIC Doc. No. ED 032 131, New York Institute for Developmental Studies, New York University, 1968.

Deutsch, C.P. Auditory discrimination and learning: social factors. Merrill Palmer Quarterly, 1964, 10, 277-296.

Deutsch, C.P. The development of auditory discrimination: relationship to reading proficiency and to social class. Bureau of Research Report No. BR-6-3034, June, 1972, Office of Education, Washington, D.C.

Durrell, D. and Murphy, H. The auditory discrimination factor in reading readiness and reading disability. Education, 1953, 73, 556-560.

Dykstra, R. Auditory discrimination abilities and beginning reading achievement. Reading Research Quarterly, 1966, 1, 5-33.

Elenbogen, E.M. and Thompson, G.R. A comparison of social class effects in two tests of auditory discrimination. Journal of Learning Disabilities, 1972, 5(4), 209-212.

Graham, L.W. and House, A.S. Phonological oppositions in children: a perceptual study. Journal of the Acoustic Society of America, 1971, 49(2), 559-566.

Hall, M.E. Auditory factors in functional articulatory speech defects. Journal of Experimental Education, 1938, 7, 110-132.

Katz, P. and Deutsch, M. The relationship of auditory and visual functioning to reading achievement in disadvantaged children. Paper presented at Society for Research in Child Development, New York, March, 1967.

Koenigsknecht, R.A. An investigation of the discrimination of certain spectral and temporal acoustic cues for speech sounds in three-year-old children, six-year-old children, and adults. Unpublished doctoral dissertation, Northwestern University, Ann Arbor, Michigan, 1968.

Kronvall, E.L. and Diehl, C.F. The relation of auditory discrimination to articulatory defects of children with no known organic impairment. Journal of Speech and Hearing Disorders, 1954, 19, 335-338.

Locke, J.L. Phoneme perception in two- and three-year-old children. Perceptual and Motor Skills, 1971, 32, 215-217.

McNeil, J. and Coleman, J. Auditory discrimination training and the development of word analysis skills. Final Report, Project No. 5-0503, University of California at Los Angeles, Los Angeles, California, July, 1967.

Perozzi, J.A. and Kunze, L.H. Relationship between speech sound discrimination skills and language abilities of kindergarten children. Journal of Speech and Hearing Research, 1971, 14, 382-390.

Prins, D. Relations among specific articulatory deviations and responses to a clinical measure of sound discrimination ability. Journal of Speech and Hearing Disorders, 1963, 28, 382-388.

Rudegeair, R.E. and Kamil, M.L. Assessment of phonological discrimination in children. Technical Report No. 118, 1970, Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, Madison, Wisconsin.

Sapir, S. Auditory discrimination with words and nonsense syllables. Academic Therapy, 1972, VII, 307-313.

Sherman, D. and Geith, A. Speech sound discrimination and articulation skill. Journal of Speech and Hearing Research, 1967, 10, 277-280.

Silvaroli, N.J. and Wheelock, W.H. An investigation of auditory discrimination training for beginning readers. The Reading Teacher, 1966, 20, 247-251.

Templin, M. A study of sound discrimination ability of elementary school pupils. Journal of Speech Disorders, 1943, 8, 127-132.

Templin, M. Certain language skills in children. Minneapolis: University of Minnesota Press, 1957.

Thorndike, E.L. The teacher's word book of 30,000 words. By E.L. Thorndike and I. Lorge, New York, Teacher's College, Columbia University, 1944.

Tikofsky, R.S. and McInish, J.R. Consonant discrimination by seven-year-olds: a pilot study. Psychonomic Science, 1968, 10, 61-62.

Travis, L.E. and Rasmus, B. The speech sound discrimination ability of cases with functional disorders of articulation. Quarterly Journal of Speech, 1931, 17, 217-226.

Vellutino, F.R., DeSettc, L., and Steger, J.A. Categorical judgement and the Wepman test of auditory discrimination. Journal of Speech and Hearing Disorders, 1972, 37(2), 252-257.

Weiner, P.S. Auditory discrimination and articulation. Journal of Speech and Hearing Disorders, 1967, 32, 19-28.

Wepman, J.M. Auditory Discrimination Test (Manual of Directions). Chicago: Language Research Associates, 1958.

Wepman, J.M. Auditory discrimination, speech and reading. Elementary School Journal, 1960, 60, 325-333.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.