8

Spelling and Reading by Rules

JONATHAN BARRON, REBECCA TREIMAN, JENNIFER F. WILF and PHILIP KELLMAN *Department of Psychology, University of Pennsylvania, U.S.A.*

1 Rules in reading and in spelling

In the study to be reported here, we examine individual differences among adults in the ability to use spelling-sound correspondences or rules. We ask where these differences arise and what tasks they affect. Although the study is correlational and therefore does not permit us to draw firm conclusions about cause and effect, our results tell us where to look further for causal antecedents of these individual differences.

The existence of spelling-sound rules in English is best viewed in terms of the history of alphabetic writing (see Gleitman and Rozin, 1977). The earliest writing systems relied heavily on relations between symbols and their meanings. If there were a symbol for 'think' and another symbol for 'study', the combination of the two might be used to indicate 'psychology', even if the phonological form of 'psychology' were unrelated to the phonological forms of the two other words. Such techniques are used in modern Chinese writing. Later writing systems used symbols to stand for syllables, regardless of whether the meaning of a syllable was preserved in different words. Like logographic writing systems, used in Chinese, syllabaries have been invented independently several times. Syllabic systems are used in modern Japanese (along with a logographic system for frequent words) and in indigenous African languages such as Vai (Scribner and Cole, 1978). The use of symbols that stand for units smaller than a syllable apparently began with the transmission of the alphabet from the Phoenicians to the Greeks. All known alphabets based on correspondences between symbols and phonemes (or minimal sound segments) were derived historically from the Greek alphabet, with the possible exception of Semitic writing systems, which indicate vowels with diacritic marks added to what is essentially a syllabary.

Originally, in Greek and Latin, spellings were closely related to pronunciations. This was often true when the Roman or Greek alphabet was adapted for a new language. But English spelling was quite chaotic from the start. 'Caxton, for example, spells the town where he spent the major part of his life before returning to England in at least six different ways: *brugges, bruges, brudgys, Brugis, bruggis, brudgis*' (Venezky, 1976). In the early sixteenth century, the scribes of the English chancery at Westminster were able to promulgate a standardised spelling system (Venezky, 1976), but no effort was made to make the alphabet phonetic. Rather, the spellings of many words were based on spellings of Latin words or other English words to which they were related. In some cases, spellings were modified for the sole purpose of distinguishing homophones. And many idiosyncratic spellings were standardised and made part of what the English schoolchild was expected to learn.

The lack of correspondence between English spellings and pronunciations has been so severe that some have advocated teaching English as if it were Chinese – as if each whole word stood for its meaning, and as if spelling-sound correspondences did not exist at all. It is certainly true that someone who learns the rules must learn several different possible pronunciations of most letters or letter clusters (often depending on context) and several different possible spellings for each sound. Some rules account for so few examples that they are probably not learned at all by most readers, and there are some words – true exceptions – that follow rules of their own, shared with no other words in the language (e.g. ONE, TWO, FOUR). English thus provides an interesting set of stimuli for a psychologist who wants to study the learning of spelling-sound rules in general. (Chomsky, 1970, has argued that many spellings actually map into psychologically real lexical forms that are abstractly related to phonetic forms. However, the psychological reality of these forms is in doubt, especially for children, and Chomsky must admit the existence of exceptions and unproductive rules, especially in frequent words.)

Despite the complexity and irregularity of spelling-sound rules, there is considerable evidence that rules are used from the start in both reading and spelling. Many studies have found high correlations between ability to read nonsense words (which must involve use of rules) and other measures of reading ability, especially in the early grades. Firth (1972), for example, found a correlation of about 0.90 between reading nonsense words and reading words in sentences (aloud) in first- and second-graders. In Firth's study, ability to read nonsense words was hardly correlated with IQ once reading ability was held constant; thus it seems that the correlation between use of rules and reading is not the result of similar general abilities being used in both tasks.

Not only is use of rules important in the early stages of reading, but it might be more important than use of rote associations between printed words and their respective spoken words. Baron (1979) found that children's ability to read regularly spelled words is correlated more highly with ability to read nonsense words than with ability to read exception words such as SWORD, ONE and HONOR, which violate spelling-sound rules.

The importance of rules extends even into the higher grades. Perfetti and Hogaboam (1975) found that good and poor readers (selected by tests of comprehension) differ more in speed of reading nonsense words than in speed of reading familiar words aloud. Calfee *et al.* (1973) found correlations between knowledge of rules and reading ability even in high school students. Frederiksen (1976) found that good readers in high school are more affected than poor readers by the complexity of spelling-sound correspondences when reading isolated words aloud. And Baron and Strawson (1976) showed that rules are used by adults in reading words aloud; specifically, words violating the rules took longer to read than regular words. While some evidence (Calfee *et al.*, 1969) suggests that other determinants of reading ability become more important once a reasonable level of knowledge of the rules is achieved, the evidence is clear that normal reading requires the learning of spelling-sound rules.

Rules seem to be used in spelling as well as reading. Read (1975) found children who invent their own phonetic spellings according to consistent rules, before receiving formal instruction in spelling. Literate adults seem able to produce credible spellings of new words, such as names. Whether differential knowledge of rules is related to differences in spelling ability is less clear. Boder (1973; see also Camp and Dolcourt, 1977) examined spelling in children diagnosed as dyslexic and found that dyslexics could be distinguished according to their knowledge of spelling-sound rules as indicated by both reading and spelling performance. Dyslexics who did not know the rules made spelling errors that seemed senseless in terms of the rules, while the few dyslexics who did know the rules made errors that were good approximations to phonetic spellings.

Boder's work suggests that there are individual differences in ability to use spelling-sound rules. Baron and Strawson (1976) have demonstrated such differences, in the course of attempting to show that rules are used in reading isolated words aloud. In that study – a precursor of the present study – the terms *Phoenician* and *Chinese* were used to characterise the two ends of the individual-difference continuum of interest. Phoenicians were those who rely heavily on the rules. Chinese were those who rely heavily on word-specific associations between each word and its associated pronunciation. (It is an open question as to whether the Chinese use the meaning of a word to extract the sound; in any case, besides a limited set of rules, they rely on thousands of specific associations between entire strings of letters and entire meanings or sounds.) The idea of the experiment was to look for a correlation between the Phoenician–Chinese dimension as measured by these tests of *use* of rules and specific associations in reading words and the dimension as measured by *knowledge* of rules and specific associations. The tests of use required subjects to read exception words and regular words aloud; use of rules was indicated by slower reading of exception words than of regular words. Baron and Strawson attempted to measure both knowledge of rules and knowledge of word-specific associations. In the test of knowledge of rules, subjects were given a list of nonsense words such as SAIF, CENNEL, MAGOR, WHURM and were asked to say which of these sounded exactly like words when pronounced by the rules (here only SAIF and WHURM). The test of knowledge of word-specific associations was more complex. It was reasoned that a person who knew word-specific associations in reading would use them to check his spelling. Such a person would do better at spelling when he could look at alternative spellings than when he had to make decisions without looking at the word he was writing. Thus, the measure of knowledge of specific associations was the difference in scores between a test in which a subject chose which of two spellings was correct and a test in which he had to spell without looking at what he was writing. The results of the study were as predicted. That is, subjects who knew many rules, as indicated by the nonsense-word test, but few specific associations, as indicated by a small difference score between the two spelling tests, were slow at reading exception words aloud relative to regular words. This was not true, however, of subjects who showed the reverse pattern of results on the nonsense-word and spelling tests. Thus, it was possible to predict the magnitude of the exception vs regular effect from tests that were very different and that concerned the knowledge hypothesised to be involved.

However, subsequent reflection (and comments of others) revealed that the spelling tests were not very useful, in principle. They measured only a particular kind of word-specific association, which is not necessarily the same kind used in reading. In particular, it was possible to do well in the spelling-recognition test not because of specific associations between printed words and sounds or meanings but rather because of recognition of the spelling as familiar or not. Thus, the measure of specific associations between printed and spoken words might have had little to do with the kinds of specific associations presumably used in reading aloud.

It is in fact hard to find a good measure of knowledge of word-specific associations for adults. In the present study we abandon the effort to find such a measure. Instead, we measure only ability to use rules. We also assume that the ability to use rules is largely uncorrelated with the ability to use word-specific associations. Thus, our Phoenicians and Chinese ought to differ more in rule-using ability than in specific-association ability. In other words, by selecting people who are good at using rules we are also selecting people who are better at using rules than word-specific associations, and by selecting people who are poor at rules we are selecting people who are better at specific associations than at rules. By selecting on only one of the two abilities of interest, we can still argue that we are looking at the same individual-difference dimension as the Baron and Strawson study, *relative* ability at rules vs specific associations.

The Phoenician–Chinese dimension has also been examined in children. Baron (1979) asked children to read lists of exception words (such as PUT, GONE, SWORD), regular words (CUT, BONE, SWEET) and nonsense words (LUT, MONE, SWORP). Chinese, in essence, were those who correctly read more exception words than nonsense words; Phoenicians were better at nonsense words. There was in fact a higher correlation between ability to read nonsense words and ability to read regular words than between ability to read nonsense words and ability to read exception words. Also, ability to read regular words was more highly correlated with ability to read nonsense words than with ability to read exception words. Phoenicians tended to make sound-preserving errors when reading words (e.g. pronouncing the *h* in HONOR or the *w* in SWORD), while the Chinese tended to make meaning-preserving errors (e.g. pronouncing TWELVE as TWENTY or DONE as DID). This study generally supported the conclusions of Boder (1973) about individual differences in use of rules, although this study was restricted to reading, ignoring spelling.

Treiman and Baron (1978) discovered a correlate of the Phoenician-Chinese dimension in children. They asked children (in first grade and kindergarten) and adults to decide which two of three syllables (e.g. /bɪ/, /vɛ/, /bo/ – as in BIT, VET, BOAT) 'went together'. Two of the syllables had been rated as being similar overall but identical in no sound segments (e.g. /bɪ/, /vɛ/). Members of a different pair of syllables were identical in a segment but were rated as being dissimilar overall (e.g. /bɪ/, /bo/). Subjects could thus classify the stimuli by overall similarity or by 'dimensional' identity. (The common segment is analogous to a common value on a dimension, and the choice of this term ties this research to other studies of perceptual development such as that of Smith and Kemler, 1977.) A subject who put together the syllables with a common segment is thus said to make a dimensional classification, based on identity of a part of each syllable, and a subject who put together the similar syllables with no common segment is said to make a similarity classification, based on similarity of whole syllables. The other possible classification (e.g. /vɛ/, /bo/) would be made only if the subject misunderstood the task or if the stimuli were poorly selected by the experimenter; this 'anomalous' classification was chosen on only a small proportion of trials. Treiman and Baron found that adults were more likely than children to make dimensional classifications, and children were more likely to make similarity classifications. This was true even when the children were instructed that the dimensional classification was correct and given feedback on their responses. Thus, adults seem to be more influenced by the presence of identical segments. Further, of the children who could read at all, those children who made a higher percentage of dimensional classifications were able to read more isolated words correctly. And the errors made by these children tended to be those that characterised Phoenicians in the study of Baron (1979), i.e. sound-preserving errors. Those children who made few dimensional classifications tended to make meaning-preserving errors, which had characterised the Chinese in the Baron (1979) study. In sum, it appears that the tendency to classify syllables dimensionally, by common segments, correlates with the Phoenician-Chinese dimension, i.e. with the use of rules, and that this use of rules correlates with reading ability in children. The hypothesis that Phoenicians and Chinese *adults* differ in segmental analysis ability will be tested again in the present study.

We should comment on the possible relations between use of rules in reading and use of rules in spelling. There are two reasons why use of rules in one of these tasks ought to be strongly related to use in the other. One possible relation between reading and spelling is based on the 'Principle of Associative Symmetry' (Asch and Ebenholtz, 1962). By this principle, associations formed in one direction should be usable in the opposite direction. Thus, if a person learns to respond with a certain sound segment or group of segments to a letter or groups of letters, this learning should be equally accessible whether the letters or the segments are given as the stimuli when the learning is tested. Learning to respond with sound segments to letters as stimuli is a possible description of learning to use rules in reading, while learning to respond with letters to segments as stimuli is a possible description of learning to use rules in spelling. The Principle of Associative Symmetry thus implies that there ought to be complete transfer of learning from one use of rules to the other. One problem with this account is that the Principle of Associative Symmetry sometimes fails. Asch and Ebenholtz in fact found that when subjects are taught response B to stimulus A in a list of paired associates, recall of B given A is usually more likely than recall of A given B. However, they argued that such asymmetry could be attributed to differential learning of the responses themselves, as opposed to the stimulus-response associations. When they equated stimuli and responses by making sure that subjects had learned the lists of possible responses equally well before any associative learning, associative symmetry was found. By this argument, associative symmetry should apply in the case of sound segments and letters, since children have presumably learned both the letters used and the segments used before any associative learning occurs.

A second reason why reading and spelling should be related is that it is possible to use one skill to check the other. When we spell, we frequently read what we have spelled to make sure it is correct. We might also check possible readings of a word by using our knowledge of spelling. In principle, we might be able to learn to spell by trial and error, based on our knowledge of reading, even without any backward associations from segments to letters.

There are also reasons not to expect relations between reading and spelling in the knowledge of rules. For one thing, the Principle of Associative Symmetry might not apply. Waugh (1970), for example, has shown that the *speed* of recalling a practised association is faster than the corresponding backward association, even when response availability is controlled. This result suggests that after practice, associations are stronger in the direction in which they are learned. The Principle of Associative Symmetry might apply only to unpractised associations. If so, the Principle would probably not apply to reading and spelling. Further, the asymmetry found by Waugh may manifest itself in other ways than speed. Stronger associations (those originally learned) might be less readily forgotten, for example. It is also possible that associative symmetry is not the result of a basic mechanism of learning, but rather is due to subjects' use of special strategies for retrieving responses from memory when they are faced with the task of producing backward associations (as suggested by the results of Spyropoulos and Ceraso, 1977). For example, subjects might search (serially or in parallel) through the list of stimuli to find the one associated with the response (see Baron, 1978). When faced with a difficult task such as spelling, in which use of such a strategy might require more mental resources than are left over from other components of the task, this strategy might break down.

Another argument against strong relations between reading and spelling is the possibility that children, at least, do not check as often as they should. Holt (1964), for example, reports an incident in which a child spelled MICROSCOPIC as MINCONPERT, and later laughed at her own mistake – thinking it was someone else's – when asked if what she had written spelled MICROSCOPIC. The student evidently failed to check.

It is to be noted that the arguments against a strong relation between reading and spelling apply largely to children. Adults presumably will use strategies such as checking and searching their memories for stimuli corresponding to a given response. Therefore, we would expect use of rules in reading and use of rules in spelling to be largely indistinguishable in adults.

Although we may be unable to separate use of rules in reading and use of rules in spelling in adult subjects, we chose to use adults rather than children for several reasons. First, it is easier to use adults. This allows more tests to be given and greater risks to be taken in testing unlikely hypotheses. Second, there is no reason to think that the determinants of individual differences in rule using ability will change with age. Third, there is a sense in which adults' learning of rules is more 'natural' than that of children. The rules we test in our selection test are for the most part not taught in school. These rules concern complex vowel clusters and relations between subsequent consonants and the pronunciation of vowels, for example. Thus, our experiments are likely to tell us about the determinants of spontaneous acquisition of spelling-sound rules rather than about the ability to profit from explicit instruction.

In the experiments to be reported, we selected subjects who were good (Phoenicians) or poor (Chinese) at spelling and reading by rules. We tried to test hypotheses about the origin of these individual differences by giving other tests to these subjects. In general we consider two classes of hypotheses. One type assumes that the difference between the two groups arises from different kinds of reading experience. In particular, if Phoenicians have more experience extracting speech codes from printed words, they might have more opportunity to learn spelling sound rules through simple repetition of the associations. By this account, we might expect Phoenicians to use speech mediation in reading silently for meaning. Such mediation would give the Phoenicians extra experience at associating sounds and letters. Conversely, Chinese might use associations between printed words and meaning even when reading aloud; they may mediate extraction of sound with meanings rather than extraction of meaning with sounds. For these reasons, our battery of tests included tests of the use of speech codes in reading for meaning and of the use of semantic codes in reading aloud. We hypothesise that Phoenicians should rely more on speech codes in extracting meaning and less on semantic codes in extracting sound. (However, we shall see that the hypothesised results have other interpretations.)

The second type of hypothesis concerns general abilities or tendencies. One tendency of interest is simply the tendency to learn rules. There are many learning situations in which people are faced with the choice of learning rules or memorising examples (e.g. English inflections, see Berko, 1958). Perhaps Phoenicians are more likely to learn spelling-sound rules simply because they are more likely to learn rules of any sort.

For this reason, our battery includes a test of propensity to learn rules. We use a paired-associate learning task in which there are rules that can be used to derive elements of the responses from elements of the stimuli. This task is as close as we could come to actually asking the subject to learn spelling-sound rules, except that the responses are letters rather than sounds.

A second general ability we consider is the ability to perceive speech segments as identical attributes of different words; we call this ability *segmental analysis*. In order to learn that the letter *b* is associated with the sound /b/, however this is learned, it is necessary to recognise that the same segment /b/ can occur in different words. For example, this segment occurs in the beginning of both BOY and BANANA. If the learner is not prone to perceive the initial sounds of these words as identical, he may have difficulty learning that this common sound segment is associated with a letter. The same goes for groups of segments and groups of letters. Of all the hypotheses we test, this is the only one with prior empirical support, as described earlier (Treiman and Baron, in press).

2 Rationale for the tests and description of methods

We gave our subjects quite a number of tests. (Most tests are shown in the Appendix.) Some tests were developed after some subjects had become unavailable. Other subjects, especially Chinese, repeatedly failed to show up for experimental sessions. And some tests were abandoned because they seemed unpromising or because nobody wanted to run them. Thus, there are different numbers of subjects run in different tests. Also, with one exception to be explained, tests were given in different orders to different subjects. Nonetheless, we shall argue that the results present a consistent picture of the difference between Phoenicians and Chinese. Most subjects were run in most tests.

3 Tests used for selection of subjects and validation of the selection

3.1 *Selection test*

In this test, subjects were asked to re-spell a number of words so that someone else would give the correct pronunciation of the word they had written as the only legal pronunciation. Subjects then did a multiple-choice version of the same test; they had to indicate which of several re-spellings of each word were correct. The total score consisted of the mean percent correct on the two parts. This test was supposed to measure knowledge of spelling-sound rules and ability to use them. Note that in the first part of the test, the production part, the number of responses given was not necessarily correlated with the percent of correct responses. Intuitively, it would seem that the number of responses given is an interesting measure in itself, perhaps comparable to the number of responses given in response to Rorschach cards (see Baron, 1978, for discussion). However, the number was not our concern here.

In the test, it was possible to give correct answers on the basis of analogies. For example, the response HAMN as a re-spelling of HAM was counted as correct, even though it is supported by only a single analogy, DAMN. The rationale for accepting this response comes from our instructions to the subjects. Another person, presented with HAMN, would pronounce it correctly if he did what most people do when they follow spelling-sound rules in pronunciation. Much evidence (Baron, 1977a, 1979) indicates that analogies are commonly used when reading nonsense words. Further, the only alternative pronunciation in this case would involve pronouncing both the *m* and the *n*, which would yield a sequence of sounds that is illegal in English. Otherwise, scoring was done as follows: Analogies to each response were sought by breaking the response between the initial consonant cluster and the first vowel cluster and trying to think of words containing these parts in these positions. If there were only a few exceptions to the intended pronunciation, the response was counted as correct. Also, the pronunciation of some letters depends on the succeeding vowels. Thus, SCULE was a not a correct re-spelling of SCHOOL because of MOLECULE (MOLEKYULE), and SKULE was not counted. Responses that were familiar proper nouns were pronounced as in the name. JAYNE was not counted as a correct or incorrect re-spelling of JANE. Nor were otherwise correct responses that were impossible as spellings (e.g. ROOOLL, SSOLE) counted as correct or incorrect.

The selection test was given to 94 subjects under a variety of conditions. In some cases, the test was handed out in classes; in other cases, it was given to subjects in other experiments after the experiments were completed. Subjects were given as much time as they wanted to complete both parts. It was hoped that the increased variability resulting from this opportunistic method of selecting subjects would be compensated by the use of subjects with extreme scores on the selection test in the other tests; this hope was fulfilled, it seems. The convenience of the test is also in its favor; it usually took less than ten minutes (although one graduate student apparently spent quite a while on it, producing over a hundred responses in the first part). Subjects did not see the second part (the multiple-choice part) until they had completed the first part. The subjects were regular undergraduate students at the University of Pennsylvania, graduate students and summer school students.

Phoenicians were originally defined as those who scored 85% or higher correct on the average of the two test scores. Chinese were those with less than 70%. Midway through the experiment, these cutoffs were changed to 90% and 65%, respectively (when it appeared that the reliability of the selection test was low – a matter we shall discuss later). However, in all the analyses reported, the Phoenician-Chinese dimension will be treated as a dichotomous variable. Data from 18 Phoenicians and 16 Chinese were actually used, although the numbers of subjects who met the criteria were somewhat larger. (Data from one subject were dropped because that subject knew English as a second language. Another subject met the criterion for being Chinese in part as a result of giving only three responses on the first part of the selection test, one of which was an error; this subject was dropped. A third subject was a research assistant who had run the free classification test using other stimuli; her data on this test were dropped.) The number of subjects run in each of the different tests ranged from 16 in the speech interference test to 29 in the segment comparison test.

3.2 *Word spelling*

Subjects were given a spelling test consisting of a number of commonly mis-spelled words. The words were both exception words and regular words. Errors were scored according to whether or not these errors followed the rules. If our selection of subjects has been successful, a simple prediction would be that Phoenicians would make more phonetic errors than Chinese.

The spelling test consisted of commonly mis-spelled English words. There were 42 words – 29 regular words and 13 exception words. The experimenter pronounced each word, giving a short phrase, if necessary, to specify the meaning of the word (e.g. a military colonel). The subject was asked to use the word in a sentence to indicate that he knew its meaning, and then to spell it. All subjects could use all the words correctly.

Errors were scored as phonetic or nonphonetic. Phonetic mis-spellings are those that when pronounced according to rules and/or analogies give the correct pronunciation. Nonphonetic mis-spellings are those that lead to an incorrect pronunciation.

3.3 *Illegal spelling*

If Chinese really have difficulty associating a sound segment with its corresponding letter, we might expect them to have trouble spelling spoken 'words' that contain segment sequences that are illegal in English, such as /tlee/ or /zdree/. Two measures are of interest here: the number of correct repetitions of the stimuli and the number of times the subject's spelling agrees with his own pronunciation of the stimulus. The former measure relates to the segmental analysis hypothesis; if Chinese have trouble analysing sequences of segments, perhaps this deficit will make it difficult to learn new sequences of sounds. The latter measure bears on the generality of the selection test. If Chinese make errors in spelling single phonemes, this indicates that their deficit is not due only to difficulties with complex spelling patterns.

Twenty-one syllables beginning with consonant clusters that are illegal in English and ending with the vowel *e* were constructed. The syllables were recorded on tape by a female experimenter (R. T.), each syllable being repeated twice. Subjects were asked to listen to each syllable, repeat it, and then spell it. The experimenter recorded the subjects' pronunciations phonetically. The first three syllables were designated as practice items and were not scored. For each subject three measures were calculated: number of correct repetitions, number of correct spellings and number of times the subject's spelling agreed with his own pronunciation.

4 Tests of the hypothesis that Phoenicians have had more practice using rules

4.1 *Speech interference*

By one of our hypotheses, Phoenicians might use speech mediation more than Chinese when reading silently for meaning. One way to measure reliance on speech mediation in reading for meaning is to ask subjects to read for meaning while simultaneously hearing nonsense syllables. The nonsense syllables might be expected to make reading difficult, as indicated by slower speeds. Reading speed was measured in a task in which the subject had to check which of two alternatives fit best in each sentence in a list. The sentences were constructed so that some sentences did not contain the phonemes corresponding to *s*, *f*, *ch* and other sentences did not contain the phonemes corresponding to *p*, *t*, *k*. The nonsense syllables used as interferences contained either the first set of phonemes or the second. The measure of speech mediation is the difference between the times with different-phoneme interference and the times with same-phoneme interference. This technique assures that the interference effects found are not due to general properties of the nonsense syllables. The effects must be due to the fact that the nonsense syllables are related to the phonemes in the sentences. We consider this technique an improvement over previous uses of interference to measure phonemic mediation (Levy, 1975; Kleiman, 1975; Baron, 1977b). In many cases, the results of other interference manipulations could be due to other factors than interference with the phonemic properties of codes used in reading.

The reason that this specific interference interferes is not entirely clear. The mechanism may be related to the effect of phonemic confusability on short-term memory (Baddeley, 1966). However, this is not an explanation, since that effect is not understood either.

In the speech interference test, subjects were given sentences of the sort shown in the Appendix. (Only some of the sentences are shown; a complete list is available on request.) Each sentence contained a blank and two word choices; subjects were told to circle the word that made more sense in the blank. Times and errors were recorded for each page of 10 sentences. Two sorts of sentence lists were alternated in presentation. One type of list had many words containing the phonemes corresponding to *t*, *p* and *k*, and no words with the phonemes *s*, *f* and *ch*. The reverse was true of the other type of list. Throughout the experiment, all subjects heard taped nonsense words through headphones at a rate of about 6 syllables per second. Volume was set so that the stimuli sounded quite loud, but induced no discomfort. Subjects run in one condition heard nonsense words formed from the consonants *t*, *p* and *k*, combined with various vowel sounds (e.g. *ta*, *po*, *kee* . . .). In the other condition the phonemes *s*, *f* and *ch* were combined with the various vowels. On both tapes, syllables were read together in triplets, giving the impression of discrete 3-syllable nonsense words.

4.2 *Homophone sentences*

A second measure of speech mediation required subjects to decide whether each sentence in a list of sentences was true or false. In one type of list, the false sentences would be true if they were read aloud and listened to by another person. For example, the sentence, A BEECH HAS SAND, is true when listened to, but not when read. If a person uses speech mediation in reading silently, we would expect him to have trouble deciding that this *homophone sentence* does not make sense. This would be reflected either in errors or times. Similar techniques have been used by Baron (1973), Baron and McKillop (1975), and Meyer and Ruddy (1973). The present version of the procedure contains an improvement over previous versions. The control condition, in which the sentences do not make sense even when read aloud, contains words equally similar, visually, to the word that would make the sentence true. For example, a *control sentence* for the one above would be, A BENCH HAS SAND.

The control condition for this test and the speech interference test also provide us with measures of reading speed. If any measure of reading speed would be related to the Phoenician-Chinese dimension, it would probably be the speed of reading isolated sentences, since the reading and comprehension of whole paragraphs would be influenced by factors having even less to do with knowledge of spelling-sound rules. Thus, this test is a good one for our purposes.

Twenty-eight pairs of homophone and control sentences were constructed. The pairs of sentences differed in just one word, the word being a homophone of a word that would make the sentence true in one case, and a nonhomophone in the other case. The homophone and control words were always nouns. They were equated for visual similarity to the correct word, and they differed from the correct word in approximately the same number of letters in approximately the same positions. The homophone and control words were also equated for frequency according to Kučera and Francis (1967). Twenty-four unambiguously true sentences were also constructed. All sentences were short, simple declaratives (mean length, 5.5 words).

Lists of ten sentences were constructed by randomly interspersing seven false sentences with three true sentences. In homophone lists the false sentences were all homophone sentences; the control lists were identical except that the homophone sentences were replaced by their controls. One set of eight lists contained four homophone lists and their control lists; another set of eight lists contained the same homophone and control sentences, but in a different order and interspersed with different true sentences. Four practice lists, two containing homophone sentences and two containing nonhomophone sentences (different sentences from those used in the test lists), were also constructed.

Each subject received four practice lists followed by sixteen test lists. Half the pairs of lists occurred with the homophone list first, and half with the control list first. Each subject received one complete set of homophone and control lists, followed by the other set; order of sets was balanced across subjects. Within these constraints, order of test lists was randomly chosen for each subject.

Subjects were instructed to read each sentence silently and to say

'Yes' if it was true and 'No' if it was incorrect in any way. Their time for each list of ten sentences was measured with a stopwatch.

4.3 *Categorised words*

Another way in which Chinese might get less practice with spelling-sound rules is their using meaning to extract pronunciation, thus circumventing the rules. Our measure of whether semantic codes were used in extracting sound from printed words consisted of asking subjects to read lists of words aloud. The categorised word list contained eight groups of five words in each group, all five from the same category (e.g. furniture, utensils, etc.). The control list contained the same words rearranged so that each group no longer contained words from the same category. If a person uses meaning in reading aloud, he ought to read aloud more quickly when the words are in categories, either because meanings are activated more quickly when related meanings have been activated (Meyer *et al.*, 1975) or because the subject can use knowledge of the category of a word to speed reading it aloud. There are other ways of interpreting such an effect. For example, it might be that phonemic codes of semantically related words are associated in memory, so that activation of one code permits activation of other related code more quickly. However, it is hard to see why Phoenicians and Chinese should differ in the magnitude of such an association effect. If the groups differ in the effect of categorising the words, the most likely interpretation would seem to be that they differ in use of semantic information in reading aloud. Whether this difference is a cause of or a result of their differential knowledge of spelling-sound rules is a much more open question.

In the categorised word test, subjects were asked to read four typed lists aloud as fast as possible without error. Two of the lists contained the words grouped into categories, in a vertical column, with spaces separating the groups. The control lists had the same format, but the words within a group were never from the same category. Two versions of each type of list were used. Each version contained the same words, but the order was changed, within the constraints described. The four lists were presented in a balanced order (either categorised-control-control-categorised or control-categorised-categorised-control) twice. Subjects were told in advance how the lists were constructed.

4.4 *Ambiguous patterns*

To test whether Chinese use rules to a lesser extent than do Phoenicians when reading aloud, we used the paradigm of Meyer *et al.* (1975). In their experiment, subjects were asked to read pairs of words aloud, such as FREAK-BREAK or COUCH-TOUCH. These words contain ambiguous spelling patterns (in some cases because one of the words is an exception word, such as TOUCH). Subjects were slower in reading the second member of each pair than when the pairs were rearranged to avoid the confusion, e.g. FREAK-TOUCH, COUCH-BREAK. We would expect such an effect to be larger in subjects who rely more heavily on spelling-sound rules in reading words aloud. (This test and the categorised word test thus measure together the use of rules and meanings in reading aloud.) Such a difference between Phoenicians and Chinese was in fact found by Baron (1979) for fourth grade children. The lists used were slightly different from those of Meyer *et al.* Instead of rearranging the pairs of words, the first word in each pair was always a homophone that could be spelled with or without the ambiguous spelling-pattern. Thus, one list would contain pairs such as MAID-SAID, and the control list would contain pairs such as MADE-SAID, in the same positions. Thus the actual sequence of sounds produced by a subject was identical for the two kinds of lists. These lists were used in the present experiment. This test can show that the groups differ in their use of rules in oral reading.

In the ambiguous pattern test, the four lists shown in the Appendix were given in the order indicated, four times, with the procedure otherwise identical to the categorised word test.

5 Tests of the segmental analysis hypothesis

5.1 *Free classification*

To measure segmental analysis, subjects were given triads of the sort used by Treiman and Baron (in press) (e.g. /bɪ/, /vɛ/, /bo/) in studying age differences in segmental analysis. Treiman and Baron had also found that this test could predict whether children would be Phoenicians or Chinese. Children who made more dimensional classifications (/bɪ/, /bo/) were more likely to make sound preserving errors in reading isolated words aloud.

In the free classification test, the subject heard triads of syllables and were asked to decide which two 'went together on the basis of sound'. Two types of triads were included. In the first type, classification on the basis of shared segments and classification on the basis of overall similarity of syllables gave different results. For example, in the triad /fI/, /se/, /fo/ (as in BEET, BAIT, BOAT) classification by shared segments would group together /fI/ and /fo/. These syllables have the same initial consonant, but have very dissimilar vowels. Use of overall similarity would lead to the classification /fI/ and /se/. These syllables are similar in both the consonant and the vowel, but identical in neither. Twenty-four such triads were constructed, as described in Treiman and Baron (in press). In sixteen, the shared segment was in the same position in the two syllables (e.g. /fI/, /se/, /fo/). In eight, the shared segment was in a different position in the two syllables (e.g. /fI/, /se/, /of/).

In triads of the second type, none of the three syllables had any segments in common, but one pair was more similar overall than any other pair. For example, in the triad /Is/, /ez/, /bo/ (as in BEAT, BAIT, BOAT) the first and second syllables are most similar. Twenty-four such triads were constructed.

The triads were recorded on tape by a female experimenter (R. T.). One tape contained two occurrences of each Type One triad, the second tape contained one occurrence of each Type One triad and one occurrence of each Type Two triad. (The purpose of the second tape was to reduce the number of dimensional responses to Type One triads; however, this purpose was not achieved.) The syllables were spoken slowly and distinctly, and each triad was repeated twice.

Before beginning the test, subjects were given practice triads of the types they would hear on the test tape. All subjects heard the two tapes in the same order, and results from the two tapes were pooled.

For Type One triads, each subject's number of 'dimensional' classifications, or classifications based on shared segments; 'similarity' classifications, or those based on overall similarity; and 'anomalous' classifications, or those apparently based neither on shared segments nor on overall similarity, was calculated. For Type Two triads, number of correct classifications, which are by necessity based on overall similarity, was calculated.

5.2 *Segment comparison, word comparison and word reading*

A second measure of segmental analysis was designed to be as close as possible to the hypothesised ability in question, the ability to recognise that two different words have the same segment in the same position. It is this recognition that would seem to underlie the recognition of the significance of the fact that these two words are spelled with the same letter in the same position. Thus, our idea was to ask subjects to judge whether or not two words had the same segment in a prespecified position (first, middle, or last – with 'middle' always referring to a vowel cluster surrounded by two consonant clusters). Ideally, we would have liked to use auditory presentation of the words. However, lack of time and handy apparatus forced us to design a pencil-and-paper form of the test, which, in retrospect, seems to have certain advantages. In this *segment comparison* test, sets of words were chosen so that the spellings of the words were completely useless in making the judgment required. For example, in lists in which the subject had to decide whether the first segments of two words were identical, one list contained the word pair CHASE–CHOIR, and another list contained CHOSE–CHAIR. In this case, the degree of visual difference between the members of each pair is constant for the two pairs, since one pair is made from the other simply by switching letters; all pairs were constructed in this way. Also, if the subject used the identity of the initial letters to make his judgment, this would help him with CHOSE–CHAIR but hurt him with CHASE–CHOIR. (In other sets of words, the subject would be helped when the answer was *different* but hurt when the answer was *same*.) This kind of design goes a long way toward solving another problem in studies of segmental judgments. Specifically, when people have learned to spell words, or when they have learned spelling-sound rules, they can often make segmental judgments by imagining the spellings of the words and comparing the spellings rather than the sounds. It is hard to believe that subjects could be using such a process in this test. To do so, they would have to imagine phonetic spellings of the words, e.g. CHASE–QUIRE, and compare the relevant parts of these spellings even while they were looking right at the (misleading) correct spellings. It would seem much easier to do what they are instructed to do, i.e. compare the segments, even if this were difficult.

The *word comparison* test required subjects to make exactly the same kinds of comparisons as the segment comparison test. However, the subjects knew in advance that the two members of each pair would be phonetically identical except for the critical segment. All the *same* pairs consisted of homophones, and all the different pairs consisted of words differing only in a single segment, in a previously specified position. The same kind of design was used as in the segment comparison test, so that degree of visual similarity was completely useless as a guide to the correct response, and for each pair in one list for which the spellings suggested a correct answer, there was a pair in another list for which the spellings suggested an incorrect answer.

The word comparison test was included out of curiosity, since our hypotheses make no prediction about the extent to which it would differentiate the two groups. Whether it would differentiate the groups would depend on how the task is done. It is done in the same way as the segment comparison task, we would expect it to differentiate the groups to the same extent. If, on the other hand, if it is done by comparing sounds of whole words, our segmental analysis hypothesis would not predict any difference between the groups. (However, it is conceivable that Chinese have trouble comparing *any* speech sounds, and that this difficulty is most clearly manifest in comparing segments only because these are the most difficult for anyone to compare. By this account, Chinese ought to have difficulty even with similarity classifications in the triads test, but their impairment should be less with these classifications than with dimensional classifications.)

The segment comparison and word comparison tests might be more difficult if the subject were unfamiliar with the words used. Since the design required the use of a number of infrequent words, subjects were given practice reading all the words used in these two tests before the tests were given. This *word reading* test, the word comparison test and the segment comparison tests were all given in a single session. The word reading test served as a measure of the speed of reading the words in the last two tests. To take advantage of another chance to collect data, the words were divided into groups according to degree of regularity. Most of the words were entirely regular, but there was one list consisting of real exception words and another list consisting of words using ambiguous spelling patterns. These two lists were combined for purposes of analysis. These lists were thus far from optimal for getting a good measure of the speed of reading exception words versus regular words.

The word reading test required the subject to read 17 columns of words aloud, with about 30 words per column, as quickly as possible as in the other tests involving the reading of isolated words. After all the columns were read once, they were read again. The columns were typed, double spaced.

The word comparison and segment comparison tests are given in full in the Appendix, along with the instructions given to the subjects and the items used for practice. When the practice items were presented, any misunderstandings about the task were corrected. Each list of items contained 20 pairs of words, in a column. The subject was to place a check next to those pairs in which the critical phoneme was the same in the two words and an X next to those pairs where it was different. (Pilot studies in which subjects were asked to say 'yes' or 'no' aloud suggested that this was considerably harder, presumably because the mode of making the response conflicted with the processes used to represent the sounds of the two words.) There were two columns on each page. The column on the right contained the words matched to those on the left. For example, if the column on the left contained CHOSE–CHAIR in the 10th place, the column on the right contained CHASE–CHOIR in the 10th place. No subject reported noticing this relationship. There was one page for each test and for each segment position. The order of presenting the pages was: first segment, word comparison; first segment, segment comparison; second, word; second, segment; third, word; third, segment. The subject went through the six pages twice in the same order. Note that the subject practiced reading all the words used in the word comparison and segment comparison tests.

6 Test of the hypothesis that Phoenicians are more prone to learn rules in general

6.1 *Rule learning*

One of our hypotheses is that Phoenicians are generally better at learning rules. The rule learning test is relevant to this hypothesis. This test was a paired-associate learning experiment in which the stimuli were consonant trigrams and the responses were other consonant trigrams. Most of the letters in each response could be derived from letters in the stimulus by simple letter-letter correspondence rules. One stimulus letter, however, was paired with different letters in each response. After subjects had learned a list of criterion, they were presented with the same four pairs, plus four new pairs, and they were asked to guess at the responses to all the items. Their responses to the new pairs served as one measure of whether they had learned rules, since these responses could in large part be derived by the rules. Subjects were also asked at the end of the experiment whether they had discovered any rules during the learning phase.

The subject was told that the task was to learn four paired associates. He was shown each of the four stimuli, typed on an index card in upper case letters, and told what its correct response was. Then he was shown the four cards, in a different random order on each trial, and asked to give the response, guessing if necessary. If the subject was wrong, the experimenter gave the correct response. After the subject was correct on all four items given in a trial for two trials, the next phase of the experiment began. Here, the subject was shown eight cards (in a different random order for each subject). He was asked to produce the responses to all cards. If he did not know the response, he was told to guess, and that it might be possible to give correct responses to items he had not seen before. Finally, the subject was asked if he had noticed any correspondence between letters in the stimuli and letters in the responses. Note that all letters in the stimuli but one had a corresponding letter in the response.

All subjects were paid $2.50 per hour for the tests. The experiment was explained to each subject after all tests were completed.

7 Results and discussion

The reliability of the selection test was assessed by dividing the test into two halves, using alternate items for each half; thus, the first half consisted of the items SEAL, GOOD, ROOF and SOAK, in both the production and recognition parts of the test, and the second half consisted of HAM, SCHOOL, STAFF and JANE. The correlation across the 94 subjects between halves was 0.51, with means of 76% and 78% correct on the two halves, respectively, and standard deviations of 14% and 12%, respectively. For the production part of the test, the correlation between halves was 0.48, and for the recognition part, 0.28; the standard deviations for the reading part were about half of those in the spelling part, although the means were not much higher (78% vs 75%).

For those wanting to measure ability to use spelling-sound rules, some recommendations are in order. First, a great deal of variability in the test seems to be due to the adoption of different criteria for a 'good' response. For example, many subjects gave GOULD as a good re-spelling of the word GOOD. We scored this as incorrect, since GOULD is actually a well-known proper name not pronounced the same as GOOD. Many subjects, when asked about this, said that they had not thought of the name or had not tried of other readings of what they wrote. In a pilot study, when subjects were asked to go back and check their responses for other possible pronunciations after completing the production part of the selection test, their scores (percentage correct) improved considerably. Further, the criterion problem affected different items differently. There was only one re-spelling GOOD that we counted as correct, namely GOODE, while ROOF had a large number of possible re-spellings (resulting from various combinations of initial R, WR and RH with EUF, UEF, OOPH, EUPH, UFE, UPHE, OOF, OOFE etc.). Thus, a subject who tried hard to think of all possible answers and didn't worry much about checking them would have his score raised by ROOF and lowered by GOOD. The same problem with the criterion existed in the recognition test. In retrospect, it seems that the best test for selecting subjects would require subjects to give the single best re-spelling of a larger number of words, where 'best' is defined as the least ambiguous as to its pronunciation. For the recognition test, it might be possible to design a test in which subjects were to indicate which two of three nonsense words or words were pronounced alike. (These suggestions are offered with the need for group administration in mind; simpler tests could be designed for individual administration.) A second recommendation concerns the scoring of tests. We – as experimenters – are not necessarily Phoenicians, and books on spelling-sound rules don't contain *all* the rules. At several points in the course of the experiment, it was necessary to go back and rescore all the data when we realised that we had made errors in deciding which answers were correct. We recommend the method described above, even if it is difficult. (We will be glad to help others on similar problems.)

The main results consist of the point biserial correlations between each of the tests and the (dichotomous) Phoenician-Chinese dimension. These correlations are shown in Table 1, along with their significance levels, the group means and the number of subjects involved in each correlation. For all tests involving times, logarithms were used to calculate the correlations: the times used were also the minimum times for each *list* used in a condition. (When there were two or more lists in the same condition, the times were summed before taking the logs. The point of using logs was to adjust effects for overall speed differences among subjects, on the assumption that the effect of some variable on a subject's time would generally be proportional to the subject's time, other things equal. In general, this use of logs is conservative with respect to the hypotheses of interest.) Error rates were not transformed.

We will first discuss the results of the tests that acted more to check the selection of subjects than to find out why the groups differed. One encouraging result is the difference between the groups in the number of errors in spelling illegal words (where an error is scored when the subject's spelling is inconsistent with his own pronunciation – note, however, that the difference between groups is still significant when errors are scored when the subject's spelling is inconsistent with the experimenter's pronunciation). This indicates that the Chinese have difficulty not only with complex spelling-sound rules of the sort tapped in the second part of the selection test, but also with the most elementary letter-sound associations. Undoubtedly, the unusual context makes the task harder, leading to more errors than would otherwise occur. But many of the errors on the selection test were also of this sort: for example, Chinese gave such responses as GHAM and AM for HAM and GUD for GOOD.

The word spelling test also confirms the selection of the subjects, and shows further that the Phoenician-Chinese dimension has something to do with more everyday sorts of spelling tasks. While the groups did not differ significantly in total errors, separation of errors into those that followed the rules (phonetic errors) and those that did not (nonphonetic, see Appendix) revealed the pattern of group differences we hoped to find. Specifically, the proportion of errors (that could be clearly classified

TABLE 1 Group differences on tests

Test and measure	Point biserial correlation with group	Phoenician mean (&N)	Chinese mean (&N)
Tests primarily for selection and validation			
Word spelling (42 items)			
total errors	0.28	8.30 (13)	11.8 (10)
phonetic/(phonetic & nonphonetic)	0.38[1]	0.74 (13)	0.62 (10)
Illegal spelling (18 items)			
correct repetitions	0.33	16.3 (12)	15.5 (11)
correct spellings	0.38[1]	15.5 (12)	13.9 (11)
spelling and repetition agree	0.63[4]	17.6 (12)	15.0 (11)
Tests relevant to practice hypothesis			
Speech interference			
(interference time)/control	−0.23	1.04 (8)	1.08 (8)
(interference errors)/control	0.06	0.13 (8)	0·00 (8)
sec per list	−0.12	31.9 (8)	30.3 (8)
Homophone sentences			
(homophone time)/control	−0.27	1.00 (11)	1.04 (12)
(homophone errors)-control	−0.18	1.70 (11)	2.80 (12)
sec per list	−0.01	10.8 (11)	10.7 (12)
Categorised words			
(uncategorised time)/categorised	0.37[1]	1.05 (16)	1.09 (12)
Ambiguous spelling patterns			
(ambiguous time)/control	0.23	1.07 (17)	1.04 (13)
Tests relevant to segment-analysis hypothesis			
Free classification of syllables			
dim./(dim.+sim.), Type One	0.52[3]	0.93 (13)	0.82 (11)
dim./(dim. + sim.), same position	0.37[1]	0.97 (13)	0.93 (11)
dim./(dim. + sim.), different position	0.54[3]	0.86 (13)	0.61 (11)
dim. + sim. total	0.25	0.99 (13)	0.97 (11)
sim., Type Two	0.22	0.94 (11)	0.90 (11)
Word reading			
total sec (minimum)	0.21	106 (17)	115 (13)
ln (exception/regular)	0.07	0.10 (17)	0.09 (13)
Word comparison			
sec per item	0.42[2]	1.27 (16)	1.55 (13)
percent errors	0.38[2]	3.90 (16)	6.00 (13)
Segment comparison			
sec per item	0.54[4]	1.62 (16)	2.17 (13)
percent errors	0.43[2]	11.6 (16)	17.7 (13)
Test relevant to general rule-learning hypothesis			
Rule learning			
transfer letters (out of 18)	0.30	14.6 (14)	11.1 (10)
rules reported (out of 4)[5]	0.36[1]	3.10 (15)	2.00 (10)

[1] $p<0.05$ [2] $p<0.025$ [3] $p<0.005$ [4] $p<0.001$: all one-tailed.

[5] Subjects were scored correct on the stimulus letter associated with different response letters if they recalled two of four responses.

as phonetic or not) that were phonetic was higher for the Phoenicians. Subsequent analysis showed that this group difference was apparently due entirely to the Phoenicians' tendency to spell exception words as if they were regular; the groups did not differ in error-type tendencies on regular words. In general, the fact that Phoenicians and Chinese make different kinds of errors suggests that rules are normally used by at least some people for spelling real words.

The groups did not seem to be drastically different in intelligence or educational background. The number of responses given in the first half of the selection test averaged 19 responses for the Phoenicians and 18 for the Chinese ($r = 0.08$ for number and group membership). The number of errors made during rule learning likewise did not distinguish the two groups ($r = 0.07$). These are the best measures of intelligence we have in the data.

The main results are described simply: of the tests bearing on the major hypotheses, the ones that distinguished the groups best were Type One triads of the free classification test and the segment comparison test. Both of these results support the segmental analysis hypothesis.

More detailed analysis confirms this general account. The free classification test distinguished the groups when analysed separately for those Type One triads with the identical segment in the same position in the syllable and for those Type One triads with the position changed. Thus, this result cannot be due to a peculiarity of either type of item. Further, the groups did not differ on Type Two triads, in which the response had to be based on overall similarity.

The segment comparison test and the word comparison test bear closer examination. One might argue that the Chinese deficit in segment comparison was due to a deficit in reading isolated words. Of course, the word reading effect was smaller than the segment comparison effect, and the word reading effect was not even significant. But still, a small problem in reading isolated words could be magnified when the additional problem of segmental comparison is added to the task. To test this, we compared the correlation between group membership and segment comparison time with the correlation between group membership and word reading time (taking into account the correlation of 0.47 between word reading and sound comparison). The former correlation was significantly higher than the latter ($p < 0.05$). However, by the same method of analysis, the correlation between group membership and word comparison time was not higher than the correlation between membership and word reading time. Nor was the correlation between membership and segment comparison time higher than the correlation between membership and word comparison time. The import of these last two comparisons is that the status of the word comparison test is unclear. Because of this unclarity, two possible interpretations of the major result cannot be distinguished. One, which we prefer because of its simplicity, is that Phoenicians and Chinese differ only in their ability to compare segments, an ability used in the word comparison task as well as the segment comparison task. A second interpretation is that the groups differ both in comparing segments and in comparing whole syllables, but, since comparing syllables is easier for everyone than comparing segments, the group differences in segment comparison are greater than the differences in syllable comparison. The results most difficult to reconcile with this latter interpretation are those from the free-classification test, where all subjects preferred dimensional over similarity (whole word) classifications, but Phoenicians showed a stronger preference. (However, this result alone might be due to use of spellings of the syllables rather than sounds for classifications. Spellings would be useless in the segment-comparison test. Further, the correlation between group membership and dim./(dim. + sim.) was not significantly higher than that between membership and Type Two similarity classifications.)

One other test bearing on a major hypothesis distinguishes the groups, the categorised word test. [Note also that the effect of categorisation itself was significant across subjects; $t(28) = 6.1$.] As mentioned earlier, however, this result has two interpretations. One interpretation is that since Chinese use meanings more often in reading aloud (or in extracting phonemic codes), they have had less opportunity to learn spelling-sound rules. The second interpretation reverses the direction of cause and effect: the Chinese rely more heavily on meanings *because* they are less skilled at using rules. This second interpretation predicts that Chinese ought to be slower at reading words, since they are less skilled at one process used in reading words. The Chinese were in fact slightly slower at reading words aloud, although the difference was not significant. It seems likely that the fact that all subjects were college students acted to restrict the range of word reading speed; those who were poor at using rules were thus somewhat better than others at using other processes required for reading (a proposal consistent with the fact that Chinese were slightly *faster* in silent-reading tasks).

Another test that may have distinguished the groups is rule learning, although the results are marginal. Conceivably, the single test used was a poor measure of general rule-learning tendency. This hypothesis seems worthy of further research.

In sum, while certain small differences between groups on other tests are suggestive, the results suggest that Phoenicians are better at segmental analysis, and that there may well be no other differences between the groups that bear on the question of how the differences arose.

A comment is in order about the hypothesis attributing the group difference in use of rules to differential use of phonemic mediation in reading for meaning. The failure to find the hypothesised difference here is unlikely to be due to phonemic mediation not being manifest in the tests, since the effect of homophone sentences was significant across all subjects [$t(22) = 1.54$ for times, 3.55 for errors], and the effect of specific phonemic interference was also significant [$t(15) = 3.31$ for times, 0.22 for errors]. However, firm conclusions cannot be drawn, since the two measures of phonemic mediation were not correlated with each other across subjects. It seems likely that while these tests were sensitive to the existence of phonemic mediation, they were not good measures of individual differences in phonemic mediation. Of the two, the homophone sentence test is probably the more accurate, since it is similar to the test of Baron and McKillop (1975) on which reliable individual differences were found.

The disappointments in the data were the failure to find differences in the relative speeds at reading exception and regular words in the word reading test and the failure to find a difference in the ambiguous patterns test. The exception-regular comparison was based on only two lists of exception words (one of which consisted of words with ambiguous spelling patterns, which might not function the same as true exception words), so this need not disturb us greatly. The ambiguous pattern test also yielded differences in the predicted direction [and also showed an overall effect of list type, $t(25) = 5.4$]. Because the words used were chosen so as to be familiar to fourth grade children (Baron, 1979), it is possible that the words were so familiar to the adults that they were all read by using word-specific associations to a large extent. Thus, this failure to replicate Baron (1979) need not be very disturbing either. Possibly a test using less familiar words (e.g. VENUS–MENUS) would yield the expected effect. (It is also encouraging that the *difference* between the effect of categorisation in the categorised word test and the effect of ambiguous patterns in the ambiguous pattern test *did* distinguish the groups significantly; this at least tells us that the two groups were differentially affected by these two manipulations of difficulty.)

On the whole, the simplest interpretation of the entire set of results is that people differ in their ability to perceive segments as identical attributes of different words. Differences in the perception of speech might arise through differential experience with speech, differential development of the brain, or perhaps through more general tendencies to compare stimuli in terms of identical attributes as opposed to overall similarity (Baron, 1978; Smith and Kemler, 1977). There is little evidence on any of these points. Saffran, *et al.* (1976) have found evidence that the right hemisphere tends to perceive speech in terms of overall similarity; specifically, an aphasic patient with a left-hemisphere lesion could perceive many words but could not distinguish minimal pairs of words on the basis of phonetic distinctions. Possibly, Chinese have less developed speech areas of the left hemisphere.

We must also acknowledge, however, that the dimensional perception of phonemes might still be a result of learning spelling-sound correspondences rather than a cause. The results from the segment-comparison test made it unlikely that subjects were actually comparing spellings. However, it is possible that learning to spell by rules had essentially provided practice in detecting segments and had improved the phonemic perception of those who did learn to spell by rules. By this account, differences in knowledge of spelling-sound rules would arise from other sources, such as general differences in rule learning, or reliance on other ways of learning to read and spell. The only way to settle the issue is to do experimental studies, showing that manipulation of phonemic perception can affect learning of spelling-sound rules relying on that perception. A beginning in this direction has been made by Rosner (1971), who found that children trained in phonemic manipulations such as counting phonemes were able to learn to read about twice as fast as children who were not so trained. However, Rosner did not show that this effect was specific to the learning of spelling-sound correspondences. We hope that our present results have increased the plausibility of the causal link from phonemic perception to spelling-sound rule learning so that other studies of this question are done.

Another direction for further research concerns the individual differences we found in segmental analysis itself. These differences may have consequences for ability to learn second languages (see Carroll, 1958). They also suggest that there are two mechanisms for perceiving speech, just as there are two mechanisms for reading printed words (Baron, 1977b). One mechanism requires analysis of speech into phonemes, just as printed words may be analysed into letters. The other mechanisms might be characterised as a 'whole word' mechanism for speech perception, a process analogous to recognition of whole printed words.

8 Summary

Two groups of adult subjects were selected: 'Phoenicians' were those who were good at spelling-sound rules to produce and recognise correct

spelling of words; 'Chinese' were those who made many mistakes in these tasks. Phoenicians' errors in spelling tended to be consistent with rules. Chinese tended to make errors in spelling phonologically illegal nonsense words.

Three hypotheses were advanced to explain the group difference: differential reliance on rules in reading, differential ability to discover rules in general, and differential ability to recognise common sound segments in different words. There was no evidence for the first two hypotheses, as the groups did not differ consistently in measures of speech mediation in reading or in the tendency to use rules in an artificial rule learning task. However, they did differ in two measures of the ability to recognise common segments – a test of free classification of syllables and a speeded test of judging whether two words contained an identical segment.

Acknowledgements

This work was supported by grant number MH 29453 from the National Institute of Mental Health; Jonathan Baron is the Principal Investigator.

9 Appendix: tests used

Spelling test (part of selection test)

Try to re-spell each of the words below as many different ways as you can. For example, if the word were *rye*, you could re-spell it *ri*, *rhi*, *wri*, *wrigh*, and so on. Make sure that each re-spelling is correct, so that if someone else wrote it you would pronounce it correctly (i.e. like the given word) on the first try. (*Note:* The words used are the same as in the pronunciation test.)

Pronunciation Test (part of selection test)

Underline each nonsense word on the right that you would pronounce the same as the word on the left (on your first try at pronouncing each nonsense word). (*Note:* italicised words are correct.)

seal – cel, sel, *seel*
ham – *hamb*, *hamn*, *hamm*, haim
good – gud, gudd, ghud, gude
school – skul, sceul, scheul, *skool*, scule, skewl, scoul, ckool
roof – ruf, *rufe*, *wroof*, *rhoof*, *rooph*, roogh
staff – *staph*, *staffe*, staphe, *staf*
soak – soce, souk, sauk, *soke*, coak, sok, sowk, *soac*, soche, *soack*, *soque*, *psoak*, soc, soch
Jane – Gane, *Jaign*, *Jeign*, Jan, Jaghn, Ghane, *Jain*

Word spelling test (* = nonphonetic error)

Target	Error
absence	abscence
acquitted	aquitted, acquited*
beggar	
colonel	colornel*, coronel* colnel*, coloniel*, colonel*, cornal*, colonial*
occasion	occassion*, ocassion*, occaison
pageant	pagent, pagant*
counterfeit	. . . fiet*, . . . fit
picnicking	picnicing*, picknicing*
occurrence	occurrance, occurance, occurence
endeavor	endevour, endever, endevor
unnecessary	unnecsessary*, unecessary*, unneccisary
rhythm	rhythmn, rhythem, rhythum, rythm, rythym
tomorrow	tommorrow
parliament	parliment, parliement*
until	
liar	lier
solder	sauter*, sodder, soddar
committee	

Target	Error
fiery	firey, firery*, fierey, firy
misled	mislead
ninetieth	nineith*, nineth*, ninetyth*, nintieth*, nineteeth*, ninetyeth
margarine	margerine, margerin
apparent	
angel	
psychic	psychik
inoculate	innoculate, annoculate*
stationery	stationary
principle	principal
recipe	receipe*, reciepe*
conscious	conscience*, concious
pneumonia	pnewmona*, pnemonia*, pnemonia, nemmonia
indict	endyte, indite, enditied*
penicillin	penacillen, pennicillen, penicillan, penecillin, penecilin*, pennicillan, pennicillin, penicillen, penicillian
balloon	ballon*
courageous	couragous*, corageous

Target	Error
prove	
beginning	
seize	siege*, ceased*, sieze, cease*, ceize
lose	loose*
parallel	parellel, paralell
dilemma	dilemna, dileama*, dilema*, delemia*
conscience	conscienous*, conscence*, concience, consciense

Illegal spelling test

nree	vree	zdlee
vwee	dlee	zvree
zdree	smree	zwee
smlee	zvwee	znwee
zbwee	dwee	mlee
thlee	hmee	sthlee

Speech interference test, examples of sentences with p, t *and* k

He talked to Ted, tripped him and told him to pay Paul or he'd kill barn / him .
Take your car to town and pay your parking lot / ticket .
To avoid making trouble took up my every minute / potion .
A tear rolled down onto her lap / dream .
A crooked banker can totally control your pocketbook / avenue .
Put your bet on our track team to win every meaning / event .
It turned out to be a bitter pill to take / beat .
Ball in hand, he took a turn toward a nearby hoop and went up to make a dunk / top .
Lately I have been pouring time into planning our trip / ruling .
To paint your boat completely will take a day or two / bucket .

Homophone Sentences, negative items only

Homophone	Control
A SUN is male	A sin is male
SOUL is a kind of fish	soil
An ORE is used for rowing	orb
A REIN can be a downpour	ruin
HARE is on the head	harm
An AIR can inherit money	ear
Letters and postcards are MALE	malt
A BEECH has sand	bench
A BEAT is a vegetable	belt
Boats may have SALES	salts
A BOW is an admirer	bog
Bread is made from DOE	dot
A person's way of walking is his GATE	gain
A PAIN is part of a window	pawn
A PLANE is where cattle graze	plant
A BEET is a measure of rhythm	bead
A CENT is a smell	scene
A blind man has lost his SITE	sigh
A PAIR is a kind of fruit	pier
FUR is a kind of tree	fire
The LOOT is a musical instrument	lift
STARES are in a house	starts
Three tones form a CORD	chore
A BEACH is a kind of tree	belch
A TALE is part of an animal	talk
A PANE is a hurt	pair
A 747 is a PLAIN	plate
Two things are a PEAR	pail

Free classification test

pI	te	po	be	vɛ	ke	ɪθ	es	oθ	še	čɛ	eš
be	vɛ	bo	fɪ	se	mI	ɪp	et	ɪl	še	ɛč	eš
fI	se	fo	še	čɛ	te	ab	ɛv	em	fɪ	se	of
še	čɛ	šo	θɪ	se	bɪ	ɪθ	es	ɪb	še	če	oš
θɪ	se	θo	ɪp	et	op	fɪ	se	ɪf	fɪ	se	ɪm
pɪ	te	mɪ	ab	ɛv	ob	fɪ	es	ɪf	še	čɛ	et

Key: a, b*a*t; e, b*ai*t; ɛ, b*e*t; I, b*ee*t; ɪ, b*i*t; o, b*oa*t; θ, *th*in; š, *sh*in; č, *ch*in

Ambiguous patterns test

son	maid	sun	made
on	said	on	said
none	dough	nun	doe
bone	cough	bone	cough
knows	great	nose	grate
cows	meat	cows	meat
been	some	bin	sum
seen	home	seen	home
four	no	for	know
hour	to	hour	to
to	steak	too	stake
go	leak	go	leak
rows	tow	rose	toe
cows	cow	cows	cow
sew	pear	so	pair
few	fear	few	fear
one	know	won	no
bone	how	bone	how

Word comparison and segment comparison test

Practice:
Say whether the words have the same sound in the indicated position:

First position	Middle (vowel) position	Last position
jim gym	through threw	peak pique
you ewe	week weak	dice dies
cue queue	pair par	flu flue
ode odd	die dye	damn dan
by buy	ways was	side sighed
phil pill	foul fool	sing sink
whole high	said wet	cow tow
west when	face bathe	fire store
wing who	head male	quartz has
shoe sue	food roof	lamb tim
cap kin	look mood	knife half
fork spoon	time pin	ring fun

First position: Word comparison		Segment comparison	
knot not	keel eel	cite sap	site cap
wheel heel	wring ring	chord cap	cord chap
cell sell	cold sold	sake cent	cake sent
whole hole	when hen	ends and	eyes aye
ill kill	new knew	are wart	one wont
where here	wrap rap	cell sold	sell cold
wrote rote	what hat	uses ease	urns earn
cake sake	cent sent	wine whore	whine wore
knight night	kink ink	cent scat	scent cat
chap cap	chord cord	grit rip	gnat nap
scat cat	scent cent	gum get	gut gem
cite site	sat cat	whose hose	when hen
gnat nat	gin in	thin then	tin ten
eye aye	end and	chow chute	cow cute
wink ink	wrung rung	hour out	hear eat
one won	are war	gone jean	joan gene
keel eel	knit nit	shoe sure	sue shore
wow owe	wry rye	chose chair	chase choir
whine wine	thin tin	thank than	tank tan
use ease	urn earn	heir hare	here hair

Middle position: Word comparison		Segment comparison	
comb calm	bomb balm	ware hear	wear hare
been bean	cheep cheap	cheep bean	been cheap
course coarse	sour soar	sour fore	four sore
deed dead	deer dear	bomb calm	comb balm
some sum	rome rum	leak greet	leek great
feat feet	head heed	coarse sour	course soar
thrown throne	gown gone	nun home	hum none
meat mate	grate grate	heal deed	heel dead
ron run	sun son	pear fair	pair fear
heart hart	fear far	some rub	sum robe
hare hear	ware wear	brake lean	break lane
meat meet	great greet	feat heed	feet head
four fore	sour sore	howl bone	bowl hone
none nun	gone gun	thrown town	thrown tone
fair fear	pair pear	peer lear	pier liar
break brake	bleak blake	grate meal	male great
howl hole	bowl bole	taut bought	taught bout
tees ties	peer pier	sun rot	son rut
shown shone	town tone	shone town	shown tone
bought bout	taught taut	heart far	hart fear

Last position: Word comparison		Segment comparison	
arc ark	lace lake	arc lake	ark lace
blue blew	sue sew	mate mat	sage sag
tow to	low lo	blue sew	blew sue
sin sink	tack tac	roll tie	role til
shoo shoe	too toe	lo tow	low to
pin pink	doc dock	inn bar	in barn
sew so	new no	tac sink	tack sin
hose hoes	dose does	tow hoe	how toe
ear earn	dam damn	shoo toe	shoe too
rage rag	bee be	tick tan	tic tank
show shoe	throw throe	dock pin	doc pink
we wee	stag stage	bloc sink	block sin
hoe how	toe tow	so new	sew no
roll role	til tie	tough doe	toe dough
barn bar	inn in	hose does	hoes dose
tic tick	ban bank	bam damn	bar darn
block bloc	sink sin	throe show	throw shoe
sage sag	bee be	lab to	lamb tom
plum plumb	so sob	stage fig	state fit
tough toe	dough loe	sag rage	sat rate

Note that in word comparisons, for same response, both words sound the same. For different response, they differ only on the critical phoneme. For segment comparisons, for same response, only the critical phonemes are the same. For different response, all phonemes are different. *Ed.*