Applied Psycholinguistics **21** (2000), 525–555
Printed in the United States of America

The development of reading tests for use in a regularly spelled language

K. J. ALCOCK and K. NOKES
University of Oxford

F. NGOWI and C. MUSABI
MAKWAMI Project, Bagamoyo, Tanzania

A. MBISE
University of Dar-es-Salaam

R. MANDALI
MAKWAMI Project, Bagamoyo, Tanzania

D. BUNDY
University of Oxford

A. BADDELEY
University of Bristol

ADDRESS FOR CORRESPONDENCE
Katherine J. Alcock, Department of Psychology, City University, Northampton Square, London EC1V 0HB, UK. Email: k.j.alcock@city.ac.uk

ABSTRACT
Data are presented on the development of tests of reading skill for primary school pupils in rural Tanzania. Instruction in these schools is in Kiswahili, a regularly spelled language. Using a translation of a standard reading test, children can read aloud all words once they have learned the sound–letter correspondences, regardless of comprehension. In addition, children can pass traditional comprehension tasks by decoding only some of the words. Three graded tests were developed to test children who had only some letter knowledge, could read single words, or were proficient readers. The tests required children both to decode and to understand the reading material in order to achieve high scores. The tests correlated well with scores on other educational achievement tests and showed age and school grade differences. It is suggested that these tests are useful measures of reading development in a regularly spelled language. Their adaptation to English and validation against standardized instruments are planned.

A good reading test should test whether pupils can both decode and understand the words they are reading. It is difficult to construct such a test in a regularly spelled language. Readers can decode words that they do not understand, and some forms of test, such as an oral single-word reading task, allow them to read

© 2000 Cambridge University Press 0142-7164/00 $9.50

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

aloud correctly words that they do not know. The reverse is also true in any language: given a passage to read with comprehension questions to answer, children give the impression of being able to decode words that in fact they cannot read because they are able to "fill in" words that they cannot decode. In this study we developed a graded test of reading for a regularly spelled language. There is no standardized test in existence for the language we are studying (Kiswahili); the reading tests that do exist are mainly those constructed by individual teachers for use in the classroom. In the absence of standardized tests, any attempt to validate a new test must rely on good score distributions and systematic variation in test scores with years of schooling and age.

To examine the alternatives that have previously been used, we review existing studies of reading in other regularly spelled languages; it seems that there are few standardized tests in such languages. This may be due to the general underserving of non-English speaking populations by standardized test development. However, it may also stem from a slightly more systematic cause: reading assessment methods used in English and other irregularly spelled languages are, by definition, unsuitable for use in regularly spelled languages. Thus, general efforts to develop tests for irregularly spelled languages will not assist those working in regularly spelled languages.

Experimental studies of reading in regularly spelled languages have, in fact, generally found this to be true: the types of tests intended for use in English are not suitable for use in such languages. Many first-pass reading tests in English are oral single-word reading tests where the primary measure is the error score. Some researchers have indeed used such tests to look at error scores in regularly spelled languages. Cossu, Shankweiler, Liberman, Katz, and Tola (1988), in a study with Italian children, examined oral reading of two-syllable words from children's first and second grade reading books and divided children into good, poor, and average readers. However, they did not report children's scores on this test. In another experiment, these authors looked at error scores for children in the first grade, but found that children in other grades had scores that were too high to use except for error analysis (Cossu, Shankweiler, Liberman, & Gugliotta, 1995). In a study in Spanish, which is slightly less regularly spelled than Italian, Carrillo (1994) found that many children scored at ceiling, even in first grade. However, these studies did not consider the possibility that children were reading aloud words that they did not comprehend. German, which is spelled more regularly than English but not completely so, seems to be the only language to have standardized single-word reading tests (Wimmer, 1996). These tests measure not only the number of errors, but also the speed of reading (Wimmer & Hummer, 1990).

This approach of looking at both speed of reading and error rates seems to be more successful in such languages. Wimmer and colleagues (Wimmer, Landerl, & Schneider, 1994) used a task that involved timed reading of both words and sentences. Porpodas (1986, 1991) timed the reading of texts in Greek, and Wimmer's (1996) German task was similar. In the latter study, even dyslexic children made few errors; however, in another study, such children were found

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

to be impaired on speed of reading (Wimmer, 1993). However, timed tests can be more difficult to administer than untimed tests.

Other studies have used reading tests that include nonwords to examine children's decoding accuracy. Working in Spanish, Valle-Arroyo (1989) found that there was an effect on reading accuracy of the length of a word but not, for real words, of frequency. It was concluded that children were using only the phonological route for reading. Working in Italian, Cossu et al. (1995) found that children reading words and nonwords did not confuse visually similar words and letters, but made phonologically based errors.

Klicpera (1989) found that German children did make more errors on unfamiliar words and nonwords, suggesting that they were using a form of the lexical route and that there may be some frequency, or at least familiarity, effect in a regularly spelled language.

In contrast to these single-word tasks, studies using text reading have tested reading comprehension. For South American Spanish-speaking children, de Manrique and Signorini (1994) used a standardized comprehension task, in addition to spelling of regular and irregular words and reading of single words used in the spelling test. A less skilled group of readers was found to be much more successful on spelling than on reading.

Naslund and Schneider (1991, 1996) gave German second graders a multiple-choice text comprehension task, as well as a test where they filled in the blank in a sentence with the correct word. The single-word tasks used by this group included matching real or distractor words to a picture and choosing the correct antonym or synonym from a set of choices. But as mentioned before, text comprehension tests are inherently problematic, owing to the ability of children to fill in words they cannot read using general knowledge and expectations.

In other studies, more general assessments of reading were used as part of the experimental study or as a means to separate readers from nonreaders or poor readers from good readers. Porpodas (1986) acknowledged the lack of standardized reading tests in Greek and used the class teacher to assess informally whether children were reading or not. Zucchermaglio and colleagues (Zucchermaglio, Pontecorvo, Tonucci, & Blachowicz, 1986) used a rather subjective 4-point scale for writing and reading in Italian, ranging from reading fluently without errors to refusing to read, and judged comprehension of a written sentence by whether children followed the instructions it gave. Again, such assessments seem to have many problems, not least of which is subjectivity.

Hence, in developing a reading test for a variety of skill levels in a regularly spelled language that might be adaptable for use in other languages, we had several options: oral single-word reading tests measuring speed, errors, or both; continuous text reading, either timed or with some kind of comprehension measure; or a more complex test involving multiple-choice responses, nonwords, or misspelled words.

None of the tests that have previously been constructed have looked at letter knowledge except indirectly as part of a word or text reading task. There is, however, the possibility of examining this aspect of children's reading develop-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

ment and of gaining a picture of very early or prereading literacy skills. Piloting revealed that, at this early stage, many children were unable to read words, but had some letter knowledge. These factors led to our decision to devise three graded reading tests at different levels to measure basic knowledge of letters, single-word reading, and reading of extended text.

Background to Kiswahili

The current study took place in an area of coastal Tanzania where Kiswahili is the most common language. Kiswahili is spoken as a native language by inhabitants of the coastal areas of Tanzania and Kenya. It is the national language of Tanzania as well as the language of instruction in primary schools, although most children in rural areas away from the coastal strip speak one of a large number of vernacular languages at home. A factor taken into consideration when choosing the study area was that a single language should be the children's first or primary language as well as the medium of instruction in primary schools.

Until the early part of this century, Kiswahili was written in Arabic script; it was then officially and fairly systematically transcribed into Roman lettering (Inter-territorial Language [Swahili] Committee to the East African Dependencies, 1956). As with many other African languages, most of which have no prior written form, its orthography is constructed (not evolved) and hence is regular. Kiswahili has only five distinct vowel sounds, and thus the five vowels of written English are sufficient to transcribe all the vowels of the spoken language with a one-to-one correspondence. Consonants are the same as in English except that there is no *q* or *x*; although there is a slight phonetic variation for some consonants, each has a direct one-to-one correspondence with one phoneme. In addition, some digraphs are used (ch (/tʃ/), sh (/ʃ/), th (/θ/), dh (/ð/), gh (/ɣ/), and kh (/x/)), but again these all have one-to-one correspondences. Stress is predictable, falling on the penultimate syllable without exception. Morphological features are transcribed phonologically. Hence, each written representation of a word can have only one phonological representation. Where one grammatical morpheme has more than one phonetic realization, the spelling follows the phonetic pattern, unlike in English. For example, the prefix /ma/ indicates a plural in some circumstances. This can have more than one phonetic realization, depending on the following phoneme: in *makubwa* 'big things', the prefix precedes a consonant, the phonetic realization is /ma/, and it is spelled *ma*; whereas in *mengi* 'many things', the prefix precedes a vowel (in this case /i/), the phonetic realization is /mɛ/, and it is spelled *me*. Contrast this with the uniform spelling *ed* for /id/, /d/, and /t/, which is the same grammatical morpheme (the regular past tense) in English.

Due to dialect differences, the pronunciation of the phonological representations may differ in different areas of the country. In some areas voiced stop consonants are implosive and unvoiced stop consonants are explosive, but in other areas all stop consonants are explosive. Children in the study area tend to use /r/ and /l/ interchangeably. Such dialectal variations would not affect children's ability to decode the written word because they would merely learn to associate two graphemes with one, potentially variable, phoneme. Teachers and

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

testers realize these issues and do not penalize oral reading based on dialectal pronunciation; the child's phonological representation would still correspond to a word in their lexicon.

To illustrate, in the case of *l*/*r* substitution there is only one minimal pair in the language: *kalamu* 'writing instrument' versus *karamu* 'feast'. If a child read aloud either of these words with /r/ or /l/, a tester would score the reading as correct. The child would know that the word just read was a real word because the phonological representation exists in his or her lexicon. *Karamu* is an infrequent word, and thus the child would probably give *peni* 'pen' as a synonym. However, in the context of a passage concerning kings, slave traders, and palaces, he or she would recall the less frequent meaning. If given a lexical decision task, the child would identify either form as a word. It is true that he or she might also identify *chakura* as a word when it is not (*chakula* is a word: 'food'). When testing beginning readers, testers and psychologists need to be aware of these issues, just as those working with some English dialects encounter children who think that *im* or *ca* is a word because initial /h/ or final /r/ is dropped in some dialects, or those working with Spanish-speaking children in Latin America encounter children who accept *lus* as a word because *s* and *z* are both pronounced /s/ in their dialect. However, given that *chakula*, *him*, *car*, and *luz* are all regularly spelled words, children speaking these dialects of Kiswahili, English, or Spanish can read these words based on phoneme–grapheme correspondences alone.

In a regularly spelled language such as Kiswahili, readers who know all possible phoneme–grapheme correspondences can then use the phonological route to read any word in the language. Once a child has cracked the code, it is theoretically possible for him or her to read every word, familiar and unfamiliar.

As there were no standardized reading tests for Kiswahili, it was necessary to develop one. Initially, we hoped to use an adaptation of the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1965), but following piloting we discovered this was not possible. Oral single-word reading tasks rely on children being able to read aloud only those words that they are actually familiar with in written form. In Kiswahili, as in any regularly spelled language, it is possible for children to appear to read better than they can in reality because they can read out loud words that they do not understand. Hence, normal children will exhibit symptoms of hyperlexia (Ellis, 1993) in an improperly designed test.

In research similar to ours carried out in English-speaking developing countries, reading tests used for screening are normally oral single-word reading tasks such as the WRAT (Simeon, Callender, Wong, Grantham-McGregor, & Ramdath, 1994). Other available tests depend on comprehension, such as the Schonell Diagnostic English Tests (Schonell, 1940). Such tests can be useful in some circumstances: for example, if it is explicitly desired to test children's inference about the meanings of words in continuous text. However, for children at the earlier stages of reading and where it is desired to test decoding skills as well as written word comprehension, these tests are not ideal; in addition, they are difficult to construct well and to use reliably. A good reading test should test whether pupils can both decode and understand the words they are reading.

There were additional factors to take into account in constructing these read-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

ing tests. In many countries, the lack of standardized reading tests is a result of little research into reading development in that language, which can be due to a dearth of educational or developmental psychologists working in the country or proficient in the language. This was the case in Tanzania. It was therefore necessary to develop tests that could be used with ease by nonpsychologists as well as by local teachers. In developing countries, large class sizes and lack of equipment are common; thus, an effort was made to develop tests that could be used quickly with large groups of children. In addition, large class sizes and poor facilities often mean that children's early reading development is so slow as to be almost imperceptible; thus it is desirable to measure some aspects of reading knowledge in preliterate children as well as more advanced skills in older children. Hence, an effort was made to construct a range of reading tests that would be usable across a range of reading skill.

The school system in Tanzania has seven years of theoretically compulsory primary education followed by either four or six years of secondary education, which about 5% to 10% of primary school pupils complete. In fact, only around 60% of children of the correct ages enroll in primary school. Reasons for nonenrollment are complex, but may include such things as lack of money for school fees, distance from school, and child labor.

Children should start school by age 7; some children enroll in nursery schools before then. Of the schools in our study, two have their own nursery schools, and one is very near a church-run nursery school. Of children in the study, 33% attended nursery school; all were at the three schools with a nursery on site or nearby. Late enrollment is also very common for the same reasons cited for failure to enroll and because parents may be unsure of a child's age. Late enrollment is more prevalent in rural areas due to higher poverty levels and greater distances to schools. Some children may enroll late purely due to the long distance from home to school, which makes parents reluctant to send a small child to school when this means a long walk each day. The current study was carried out in the first to fifth grades with children aged 8 to 14. All children were within two years of the modal age for their grade.

The current study

This article reports on two studies, the first consisting of data collected by using existing reading tests and the second discussing the development of novel, graded reading tests. These are reported separately. The first study includes an adaptation of an existing reading test for English, the WRAT (Jastak & Jastak, 1965), and data obtained from the reading tests currently in use in the first and second grade in the study schools. The second study reports the development of four new reading tests as well as performance on the established tests.

STUDY 1: EXISTING READING TESTS

Participants

For the translation of the WRAT reading test, 67 children from one school, in the first to fourth grades, were tested. For the school-based academic achieve-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

ment tests, data were collected for 455 children in eight schools, in the first and second grades (other grades did not take a separate reading test). All schools were in Bagamoyo District, Coastal Region, Tanzania. All children were aged 8 to 14, and the children who took the school-based academic achievement test were within two years of the modal age for their grade.

Methods

WRAT. The WRAT test is a graded task that involves individual oral reading of letters and words. As a first step in the construction of a Kiswahili reading test, the test was translated directly and then ordered by word length and roughly by familiarity, using the presence of a word in a lower or higher grade reading book as an index of familiarity. The reading books were not graded by word frequency or length, but as children in this study had very little access to written Kiswahili outside school it was felt that this was a realistic grading of the probability of a child knowing a word. This was similar to the procedure followed by Klicpera (1989) in developing a reading test for use in German, where words not found in the reading books for a particular class were classified as unfamiliar – this in a culture that is probably much more literate than that of coastal Tanzania. Words at higher levels of the test were not found in reading books and, in some cases, were of such low frequency that the English word was unfamiliar to the first author and the literal Kiswahili translation was unfamiliar to our Tanzanian colleague who translated the test.

One tester administered this test to 67 children. It was administered individually in one of two forms. Children were allowed to try reading a typed version of the test. If they could not read the letters on the page or if they could read the letters but not the words, they were transferred to a handwritten version. This used the same script as was used by the teachers of the first and second grades for classroom instruction. Some children who had never seen typed materials before had some success with handwritten materials but none with typed materials.

The test consisted of 7 letters plus 58 words, for a total of 65 items. Children discontinued the test if they failed 10 consecutive items.

School-based reading tests. The school-based reading tests were designed by the class teachers in the schools where the testing took place; the tests were part of the regular end-of-term testing that took place in every school. The children were in the first or second grade when the testing took place. These data are later compared to data collected halfway through the following year when the same children were in the second or third grade.

These tests took a similar form to the WRAT, although some teachers at some schools included oral reading of sentences or word or nonword digraphs (CV syllables, usually a consonant followed by the vowel *a*) in their task. In one school, the test given was oral reading of a page from the reading book used in class.

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

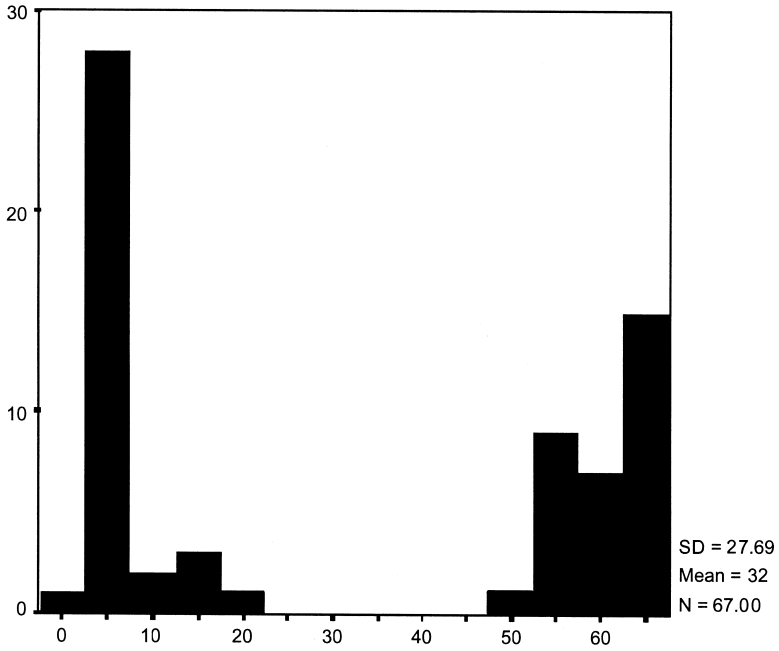


Figure 1. Reading scores, WRAT translation.

Results

The translation of the WRAT gave a bimodal distribution of scores, as can be seen in Figure 1. Only four children scored over 7 but under 48 (i.e., they could be classified as reading words but reading them poorly).

The school-based reading tests also tended to have a bipolar distribution. Figure 2 gives the results for one of the second grade classes who were reading sentences aloud. The distributions were similar for other oral reading tests.

Discussion

It can be seen from these data that both the adaptation of the WRAT and the existing tests used in schools have strongly bimodal distributions. Many children scored either at floor or at ceiling, and none of these tests differentiated children who could read words at different ability levels. In addition, all children who could read only letters tended to be grouped together, and in some of these tests a high proportion of children scored zero. It therefore seemed necessary to develop new reading tests with the following characteristics:

1. Children should not be able to achieve a score for words that they cannot understand or they cannot decode.

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

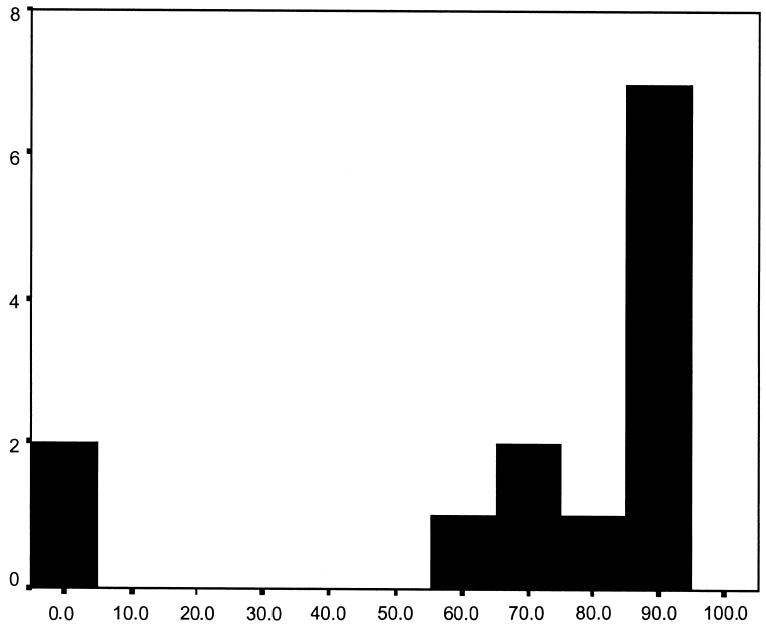


Figure 2. Scores on school-based reading tests, grade 2.

2. Poor readers who have some letter knowledge should be able to achieve a range of scores depending on their level of letter knowledge.
3. Good readers who can decode and understand most common words should be able to achieve a range of scores based on their fluency or speed of reading.

STUDY 2: NEW READING TESTS

Participants

A total of 314 children in four schools were tested for the main study. During pilot testing children in the first to sixth grades, from two additional schools, were tested to investigate floor and ceiling effects. Pupils in the main study were in the second to fifth grades and were aged 9 to 14. These children were all within two years of the modal age in the grade in which they were enrolled.

Methods

Three reading tests were developed. These are presented in order of increasing difficulty.

Letter reading test. In this test the child was asked to discriminate between letters and pseudo-letters. Both this test and the word reading test were origi-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

nally timed tasks, administered individually. During piloting it was found that the number of errors varied with the time taken, such that poorer readers made many errors and were very slow, whereas better readers made few errors but took roughly the same amount of time to do the task. This meant that time taken did not distinguish between participants any better than number of errors. It was therefore decide to carry out the task as an untimed classroom task.

Letters that seemed to the children to resemble either Roman or Arabic letters or rotated versions of them were omitted. The final version of the test consisted of two parallel forms, each with 24 items. All 24 letters of the Swahili alphabet were represented and were randomly assorted with 24 pseudo-letters, so that each parallel form had 12 letters and 12 pseudo-letters. All letters were written as capital letters in a form similar to the lettering used by primary school teachers. Items were not numbered, as it was found in pilot testing that some children confused real numbers with pseudo-letters. One parallel form of the test is shown in the Appendix.

It was explained to the children that on their papers were some letters that were real and some letters that were not real. Four sample letters with a small box to the right of each one were drawn on the blackboard, and the tester asked the children if the letters were readable or not readable. For the first letter and pseudo-letter, the tester explained that the children were to put a mark in the box. The tester emphasized that the children should put a tick if the letter was readable and a cross if it was not. The children were reminded that they should write a mark for each letter to show whether it is readable, without skipping any.

Test–retest reliabilities, at an interval of one week, and interform reliabilities (parallel form A vs. parallel form B), also at an interval of one week, were carried out prior to testing the main group.

Number correct, hits, and false alarms were scored. The maximum number correct was 24, and the maximum number of hits was 12. The main dependent score was hits minus false alarms.

Word reading task. The construction and development of this task, as well as the instructions given, were very similar to that of the letter reading task. The test, originally developed as an individually administered task, was altered to be a classroom task, with similar instructions and procedure. There were some differences, however.

Both forms of the task had 12 real words, taken from first grade reading books, and 12 possible Swahili nonwords. The nonwords differed from real words by two or more phonemes. The words were all printed in lowercase letters, in the font that was most similar to the font used in the reading books. The items were numbered in this case, as it was felt that children could distinguish numbers from words. The task was therefore equivalent to a written lexical decision task.

Test–retest reliabilities, at an interval of one week, and interform reliabilities (parallel form A vs. parallel form B), also at an interval of one week, were again carried out prior to testing the main group.

Number correct, hits, and false alarms were scored. Again, the maximum

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

number correct was 24, and the maximum number of hits was 12. The main dependent score was hits minus false alarms.

Sentence reading task. This task was based on the Silly Sentences task (Baddeley, Emslie, & Nimmo-Smith, 1992), which in turn was a modification of the sentence verification task developed by Collins and Quillian (1969). It was intended to provide a speeded comprehension task for better readers. When adapting this task for group testing, it was decided that some element of timing was still important; rather than time each child for a short task, it was decided to construct a long task that no child could complete in the standard time allowed. A total of 125 sentences were chosen.

The sentences were based on fundamental knowledge that was assumed to be well known to even the youngest child (e.g., "Is your hand attached to your arm?"). The negative sentences were equally obviously false (e.g., "Is your hand attached to your leg?"). Sentences found to be ambiguous during piloting were dropped.

Half the sentences were true and the other half were "silly." They were randomly distributed so that children who read up to any point in the test would encounter roughly half true and half silly sentences in their set of sentences. All the sentences were in question format so that the children would not be faced with the dilemma of disagreeing with adult statements (Baddeley, Gardner, & Grantham-McGregor, 1995).

The introduction to the test was the same as that for the letter and word reading tasks except that the practice sentences were printed on a separate cover page and the test sentences were not visible during practice. The tester emphasized to the group that in this test they must read and respond as quickly as possible and that they had only 5 minutes to read a very large number of sentences. They were also told that they should not wait for the tester to tell them to turn over each page, as in piloting it was found that many children were reluctant to continue the test without being told to do so by the tester. The children were given 5 minutes to complete the test.

Only a proportion of the children took the sentence reading test; the total number was 166. Test–retest reliability, at an interval of one week, was carried out as part of piloting.

Number correct, total number of sentences read, hits, and false alarms were scored. The maximum possible number correct was 125, and the maximum possible number of hits was 63. The main dependent score was again hits minus false alarms.

Cutoff points. All children did the letter and word reading tasks, but during piloting it was seen that the sentence reading task was too difficult for some poor readers. These children almost all had no score on the sentence reading and found the task distressing and onerous. Cutoff points were therefore set for this task to determine which children should do which tasks.

All children in the fourth and fifth grades took the sentence reading task. Children in the third grade who scored 9 or above (hits minus false alarms) on word reading then took sentence reading. Too few children in the second grade

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

scored sufficiently well on the letter reading and word reading tasks; therefore, the sentence reading task was not used with children in this grade.

Additional tests administered. Half the children tested as part of this study (i.e., the results reported here) were also tested on a battery of tests of cognitive functioning, including a measure of verbal working memory (digit span forwards and backwards), a measure of nonverbal working memory (Corsi block span), category fluency (foods and animals), a measure of long-term memory (paired associate learning – Spanish words with pictures), an adaptation of the Stroop task for prereaders (the congruent condition consisted of "Say tick when you see a tick and cross when you see a cross"; the discongruent condition was to say the opposite), and the grooved pegboard task. For a discussion of the methods used to adapt such batteries for use in developing countries, see Baddeley et al. (1995).

Additional tests of educational achievement were administered to all children whose results are reported here, comprising a spelling test and two levels of arithmetic tests, both adapted from the WRAT (Jastak & Jastak, 1965).

Results

Letter reading test. Summary statistics for this test are found in Table 1. Figure 3 shows the distribution of hit minus false alarm scores on this test. Test–retest reliability at a one week interval for this test was 0.77, and interform reliability (form A vs. form B) was 0.76.

Table 1 shows differences by grade on the letter reading task. Nonparametric tests revealed that children's performance on this test improved significantly with school grade (Kruskal–Wallis chi square = 53.26, $p < .001$).

Scores on the letter reading task were found to correlate significantly with other tests administered, as shown in Table 2. Correlations were significant with all the other reading tests and with the other tests of educational achievement administered as part of this study. Correlations were also significant with reading and Kiswahili tests previously administered by the schools. Although the school-administered reading tests were devised by separate teachers for each class, correlations within classes were similar: overall correlations would be lower if the variable nature of the tests between classes were an issue.

Correlations with many of the tests of cognitive functioning administered were found to be significant, including digit span forwards, one measure of verbal fluency, and one measure of the Stroop effect.

Word reading test. Summary statistics for this test are found in Table 3. Figure 4 shows the distribution of hit minus false alarm scores on this test. Test–retest reliability for this test was 0.91, and interform reliability was 0.92.

Table 3 also shows differences by grade on the word reading task. Children's performance on this test improved significantly with school grade (Kruskal–Wallis chi square = 76.63, $p < .001$).

Scores on the word reading task were found to correlate significantly with

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Table 1. *Letter reading task, hits minus false alarms*

	M	*N*	*SD*
Overall results	8.06	314	4.21
Results by class			
Grade 2	6.14	101	4.89
Grade 3	7.65	94	4.22
Grade 4	9.64	61	2.68
Grade 5	10.41	58	1.96

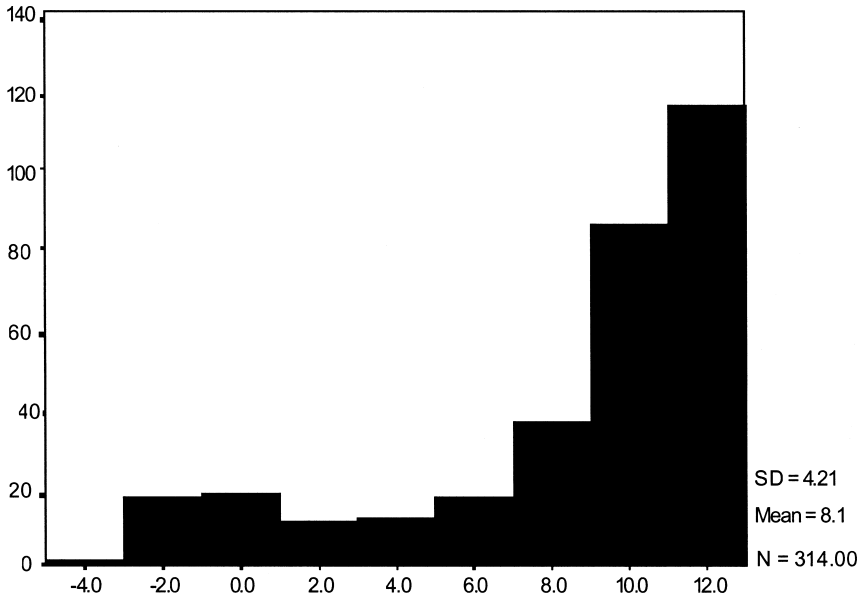


Figure 3. Letter reading, hits minus false alarms.

other tests administered, as shown in Table 4. Correlations were highly significant with all the other reading tests and with the other tests of educational achievement administered as part of this study. Correlations were also high with the reading and Kiswahili tests previously administered by the schools.

Correlations with almost all of the tests of cognitive functioning administered were found to be significant.

Sentence reading test. Summary statistics for this test are found in Table 5. Figure 5 shows the distribution of hit minus false alarm scores on this test. Test–retest reliability on this task was 0.89.

Table 5 also shows differences by grade on the sentence reading task. Chil-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Table 2. *Correlational table for letter reading task*

Relationships with other reading tests						
		Word reading (hits – false alarms)	Sentence reading (hits – false alarms)			
Spearman's rho correlation coefficient	Letter reading (hits – false alarms)	0.507** $n = 314$	0.260** $n = 166$			
Relationships with other tests of educational achievement						
		Written arithmetic score	Spelling score	Oral arithmetic score	Marks in class Kiswahili test	Marks in class reading tests
Spearman's rho correlation coefficient	Letter reading (hits – false alarms)	0.404**	0.399**	0.281**	0.356**	0.304**

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Relationships with tests of cognitive function							
		Digit span forwards (total correct)	Digit span backwards (total correct)	Corsi block (total correct)	Fluency (foods)	Fluency (animals)	Category fluency total
Spearman's rho correlation coefficient	Letter reading (hits – false alarms)	0.163*	0.182*	0.298**	0.052	0.204**	0.146
		Spanish learning (total correct)	Stroop correct per minute forwards	Stroop correct per minute backwards	Stroop effect (corr/min)	Pegboard, dominant hand (time/secs)	Pegboard, nondominant hand (time/secs)
		0.136	0.266**	0.170*	0.175*	–0.078	–0.169*

*Significant at the .05 level (two-tailed); **significant at the .01 level (two-tailed).

Table 3. *Word reading, hits minus false alarms*

	M	*N*	*SD*
Overall results	7.10	314	4.85
Results by class			
Grade 2	4.62	101	4.57
Grade 3	6.80	94	5.06
Grade 4	8.38	61	4.12
Grade 5	10.57	58	2.80

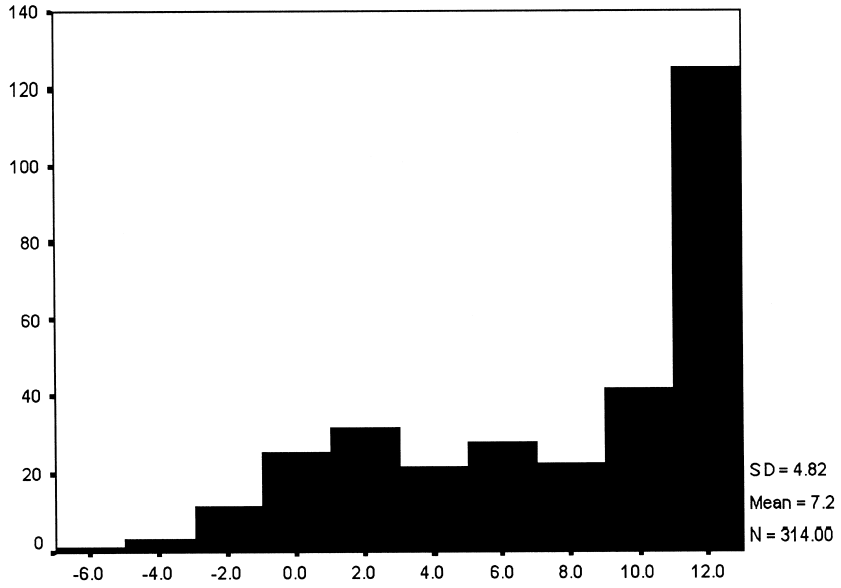


Figure 4. Word reading, hits minus false alarms.

dren's performance on this test improved significantly with school grade (Kruskal–Wallis chi square = 56.862, $p < .001$).

Scores on the sentence reading task were found to correlate significantly with the other tests administered, as shown in Table 6. Correlations were significant with both of the other reading tests. Correlations were also highly significant with the other tests of educational achievement administered as part of this study. Again, too few children took both oral arithmetic and sentence reading for this correlation to be calculated. A correlation was also significant with written Kiswahili tests previously administered by the schools.

As with the other reading tests, correlations with many of the tests of cognitive functioning administered were found to be significant. This included digit

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

span forwards and the Spanish word learning task but not any of the measures of category fluency.

Reading letters, words, and sentences. Hit minus false alarm scores on these three tests were added together to give one total score for reading letters, words, and sentences. The distribution is shown in Figure 6. During piloting, almost all children who were in the second or the third grade and scored less than the cutoff point (9 hits minus false alarms on the word reading test) had no score on the sentence reading test. These children did not take the sentence reading test in the main study. However, since they were expected to achieve no score on the sentence reading test, their score was set to zero for this test. Hence, the graph and statistics for the total reading score includes all 314 children in the study, with a score of zero on the sentence reading test for the children who did not take this test.

As with the individual tasks, children's performance improved significantly with school grade (Kruskal–Wallis chi square = 136.772, $p < .001$). This is shown in Table 7.

DISCUSSION

These new tests present a method of assessing reading in a regularly spelled language. The reading tests are graded, which ensures that children who perform at floor or at ceiling on one test are distinguished from one another on the next higher or lower test. Although there were groupings of children at ceiling on the easier tests (letter reading and word reading), the summary score for all three tests provided a useful distribution, without clustering at either end. These tests could be readily adapted to other regularly spelled languages.

It was decided not to continue to develop individual timed reading tasks because of the amount of time needed to administer them (the overall study included over 1,000 children) and because of the difficulty of reliable administration by nonpsychologists. The current tasks were reliably administered by three testers who had primary school teaching diplomas (equivalent to four years of secondary education) and who received approximately two weeks of training and piloting in schools. Piloting showed that it was not necessary to administer all the tests to all children, thus reducing testing time, and that all three tests could be administered in classroom settings.

Relatively high test–retest and interform reliabilities were reported, and the intercorrelation between scores on the individual tests was high. The tests also correlated well with standard tests of cognitive functioning, particularly those with a working memory component. They were also found to correlate significantly with established tests of educational achievement, both those in use in the schools and those adapted from the WRAT. Children's performance on the test was better if they had made more progress through the school system. Together, the findings suggest that these tests not only are internally consistent, but also have some external validity.

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Table 4. *Correlational table for word reading task*

Relationships with other reading tests						
		Letter reading (hits – false alarms)	Sentence reading (hits – false alarms)			
Spearman's rho correlation coefficient	Word reading (hits – false alarms)	0.507** $n = 314$	0.534** $n = 166$			
Relationships with other tests of educational achievement						
		Written arithmetic score	Spelling score	Oral arithmetic score	Marks in class Kiswahili test	Marks in class reading tests
Spearman's rho correlation coefficient	Word reading (hits – false alarms)	0.666**	0.830**	0.443**	0.585**	0.485**

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Relationships with tests of cognitive function

		Digit span forwards (total correct)	Digit span backwards (total correct)	Corsi block (total correct)	Fluency (foods)	Fluency (animals)	Category fluency total
Spearman's rho correlation coefficient	Word reading (hits – false alarms)	0.477**	0.518**	0.352**	0.206**	0.212**	0.263**
		Spanish learning (total correct)	Stroop correct per minute forwards	Stoop correct per minute backwards	Stroop effect (corr/min)	Pegboard, dominant hand (time/secs)	Pegboard, nondominant hand (time/secs)
		0.285**	0.517**	0.406**	0.178*	–0.124	–0.249**

*Significant at the .05 level (two-tailed); **significant at the .01 level (two-tailed).

Table 5. *Sentence reading, hits minus false alarms*

	M	*N*	*SD*
Overall results	15.5	166	7.69
Results by class			
Grade 3	7.78	94	8.78
Grade 4	12.00	61	7.73
Grade 5	19.20	58	7.38

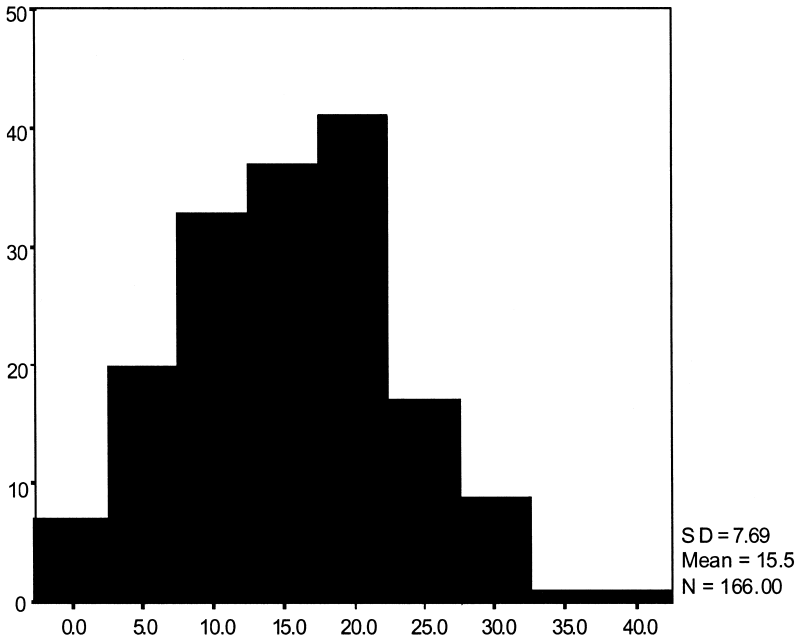


Figure 5. Sentence reading, hits minus false alarms.

CONCLUSIONS

We developed tests of a new design which ensure that children who are learning to read a regularly spelled language can be tested reliably on whether they are decoding and understanding the words they are reading. On the word reading task, it is not possible to score well unless the words are both decoded and understood. A child who is not capable of decoding print into sound must be merely guessing as to whether a word is real or not. The nonwords used here are visually and phonologically similar to real words, so that decoding at the phonemic level is necessary to distinguish them from real words. In contrast, a child who can decode the words that he or she is reading but has never encoun-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

tered before will score at chance because these words will not be familiar to him or her.

Other researchers have used word reading tasks to distinguish children who are decoding words from those who are decoding and understanding words. The tests used by Naslund and Schneider (1991, 1996) seem to be most similar to those that were used in the present study. They asked children to match real words or distractor words (pseudowords) to a picture, gave them multiple-choice questions about single words, and asked them to select synonyms and antonyms. Children who could only decode words would not be able to do these tasks, whereas those who could understand would perform well. However, these tests, especially the latter, require metalinguistic knowledge as well as reading skill. Our tests were designed to test reading skills (not other metalinguistic knowledge), using children's knowledge about everyday objects in a comprehension task (the sentence reading task).

The sentence reading task was designed with this aim in mind: to test decoding as well as understanding. Children who could only decode the words would not understand the true/false questions in the sentences. The opposite situation – children who "filled in" words they could not decode – would prompt children to insert the most likely word into any given sentence. However, in a Silly Sentences reading paradigm, sentences were not predictable. A reader could not simply insert the most likely word and still answer correctly; nor could the reader guess the content of a passage or sentence from one word or a few words, as in many comprehension tests. It was necessary both to decode and to understand, at a minimum, all the content words in a sentence in order to answer correctly.

Although the scores on the overall set of tests (total hits minus false alarms) were normally distributed, on some of the tests a large proportion of children reached the maximum score. However, it was still possible to see significant correlations within the different reading tests and between these tests, other tests of educational achievement, and tests of cognitive functioning, which suggests that our tests had good discriminatory power. Administering only a single oral word-reading test with a simple error score measure may not provide all the information necessary to distinguish between good and poor readers once the skill of decoding from print to sound has been attained. The wider range of single-word reading scores and total number of sentences read suggested that competent readers differed in their speed of reading, and that this speed measure (speed of reading sentences) distinguished among children who could read but who did so at various skill levels. This test was an improvement on previous speeded reading tests (e.g., Wimmer, 1993, 1996), which also distinguished between good and poor readers but had to be administered individually. All our tests could be administered in a group setting.

Our tests not only measured these two desirable abilities (decoding and comprehension), but also demonstrated external validity. They correlated with scores on tests of reading used in the schools, with other educational achievement tests adapted from standardized tests, and with locally appropriate adaptations of commonly used tests of cognitive functioning. The test scores varied in the expected direction with years of schooling. However, given the lack of standard-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Table 6. *Correlational table for sentence reading task*

Relationships with other reading tests				
		Letter reading (hits – false alarms)	Word reading (hits – false alarms)	
Spearman's rho correlation coefficient	Sentence reading (hits – false alarms)	0.206** $n = 166$	0.534** $n = 166$	
Relationships with other tests of educational achievement				
		Written arithmetic score	Spelling score	Marks in class Kiswahili test
Spearman's rho correlation coefficient	Sentence reading (hits – false alarms)	0.588**	0.709**	0.609**

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Relationships with tests of cognitive function							
		Digit span forwards (total correct)	Digit span backwards (total correct)	Corsi block (total correct)	Fluency (foods)	Fluency (animals)	Category fluency total
Spearman's rho correlation coefficient	Sentence reading (hits – false alarms)	0.269**	0.435**	0.196	0.063	0.129	0.121
		Spanish learning (total correct)	Stroop correct per minute forwards	Stoop correct per minute backwards	Stroop effect (corr/min)	Pegboard, dominant hand (time/secs)	Pegboard, nondominant hand (time/secs)
		0.275**	0.297**	0.284**	–0.022	–0.022	–0.273**

*Significant at the .05 level (two-tailed); **significant at the .01 level (two-tailed).

Table 7. *Reading of letters, words, and sentences, total hits minus false alarms*

	M	*N*	*SD*
Overall results	23.4	314	15.80
Results by class			
Grade 2	10.8	101	7.97
Grade 3	22.2	94	15.60
Grade 4	30.0	61	12.00
Grade 5	40.2	58	9.86

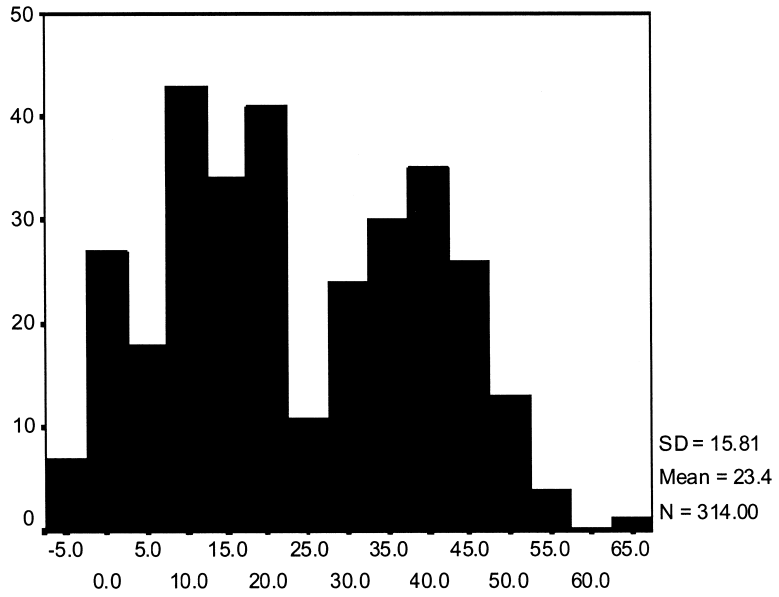


Figure 6. Total scores for reading of letters, words, and sentences.

ized tests in the language, it is difficult to draw too many conclusions about the validity of the tests. Baddeley and others are currently undertaking a validation of these tests in English, comparing them to existing standardized tests.

From this data it is possible to draw a few conclusions about the development of reading in Kiswahili and other regularly spelled languages. Thorstad (1991) carried out a comparative study of reading development in regular and nonregular orthographies, looking at initial teaching alphabet (i.t.a.), Italian, and traditional English orthographies. Thorstad found that children learning to read in either of the regular orthographies (Italian or i.t.a.) read earlier than those learning traditional English orthography. Their data concurs with our findings (especially on the translation of the WRAT and the school-administered reading tests)

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

that children who can decode any words can decode all words. The development of decoding accuracy in regularly spelled languages is a rapid and all-or-nothing process. After learning to decode words, children may increase their speed of decoding, but there is no period in which they can decode some but not all words (e.g., familiar but not unfamiliar words).

It may be that some particular features of Kiswahili lead to this all-or-nothing decoding phenomenon. In particular, the invariant simple syllable structure (only open syllables and very few consonant clusters) and completely predictable stress pattern may lead children to this extreme ease of decoding. Although we have not carried out cross-linguistic comparisons, there is a large amount of similar data from other such languages to suggest that this phenomenon is not restricted to Kiswahili.

Spanish has a more complex syllable structure than Kiswahili, including several consonant clusters and closed syllables as well as diphthongs and some unpredictable or context-dependent accents. In English and in other irregularly spelled languages, individuals can read irregularly spelled words only if these are already a part of their vocabulary. Hence, single-word reading tests with irregularly spelled items that vary in frequency are often used with adults as proxy tests for vocabulary knowledge and hence as easy to administer estimates of IQ.

In Spanish it is not possible to construct a single-word reading test because all words, even those not known, are perfectly decodable by skilled readers (Del Ser Quijano, Gonzalez-Montalvo, Martinez-Espinosa, & Delgado-Villapalos, 1997). To construct such a test, these authors found it necessary to rely on a small class of infrequent, irregularly accented words that were presented in an unnatural, accentless form. This evidence emerged from adult practiced readers. However, skilled beginning readers in Spanish displayed a similar pattern (Signorini, 1997); they were able to read aloud 75% of pseudowords (compared with 94% of words), suggesting that they were almost as capable of decoding pseudowords as words. Unskilled beginning readers were only able to read 17% of pseudowords, suggesting a bimodal distribution very similar to that found in our data.

In Turkish, which has fewer consonant clusters than in English as well as many closed syllables, beginning readers were found to read pseudowords and real words equally well. The two types of words were thus collapsed into one test (Oeney & Durgunoglu, 1997). Like the beginning Italian readers discussed earlier (Thorstad, 1991), these beginning readers were able to learn very rapidly to read both words and pseudowords. In October of their first grade year they could read 26% of words and pseudowords, and by February they were able to read 72% of these items (though no breakdown of distribution is reported), reaching a ceiling level of 93% by May (Oeney & Durgunoglu, 1997).

An ongoing study conducted in Malay – which has some consonant clusters, some diphthongs, some closed syllables, and no accent on any syllable in a word – suggests that children learning to read this language can also decode words they do not yet understand once they have cracked the code (S. Rickard Liow, personal communication, 1999).

Of all of the regularly spelled languages discussed here, Greek may have the syllable structure closest to English. It has many consonant clusters, including

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

most of those permissible in English and some that are not; it also has many closed syllables and unpredictable stress, which can be minimally distinctive but is marked on the written word. Giannouli and Harris (1997; Harris & Giannouli, 1999) found that children at the end of first grade performed at ceiling when reading all types of words – familiar words from school reading books, unfamiliar words, and pseudowords.

One of the few languages in which some frequency or familiarity effect has been found in beginning readers – German – is in fact one of the less regularly spelled languages of this group of languages. Beginning readers were found to be more accurate at reading familiar words (taken, as in this study, from beginning reading books) than unfamiliar words or pseudowords. However, after only a few weeks of reading instruction, children could read two-thirds of unfamiliar words and pseudowords. It was only in a subset of poor readers that the pattern of very poor reading of unfamiliar words and pseudowords, compared to reading of familiar words, persisted past the first few weeks of instruction (Klicpera, 1989).

These comparisons between very different regularly spelled languages are striking, both qualitatively and quantitatively. In languages where a comparison between skilled and unskilled beginning readers has been done (e.g., Kiswahili, Spanish, and Turkish), a bimodal distribution has been found in single-word reading skill on unfamiliar words or pseudowords. The levels of performance for skilled and beginning readers on such words are very similar in proportion in every language: less than 25% for unskilled readers and at ceiling, in many cases, by the end of first grade. In Greek, Italian, and English taught with i.t.a., unskilled beginning readers were not tested, but the rapid acquisition of the ability to decode almost all words by the end of first grade was also noted (including unfamiliar words or pseudowords where these were tested) once the code of the written language had been cracked.

From this data it may seem that Tanzanian children are learning to read relatively late, especially given the fairly transparent nature of the task and the excellent performance of children in other regularly spelled languages by the end of first grade. However, it should be borne in mind that, although the official enrollment age for primary school is 7 years, many children do not enroll until later, even up to age 10. Class sizes are large (40 is standard, and where two or more streams are taught together the class can be over 100), and the physical environment of the school is usually poor, with a leaking roof and an earth or sand floor. There are few desks, and the only teaching aid generally available is a blackboard; books are either not available or too few. Most teachers have the equivalent of four years of postprimary education in a teacher training college; some have four years of secondary education followed by two years of teacher training, and some older teachers have no postprimary education at all. Given the large class sizes, individual contact with the teacher is minimal. Under these circumstances it is surprising that the children are able to learn to read at all; the transparent orthography of Kiswahili may well help them in their task.

These tests were developed as part of a large study that aimed to investigate the impact of parasitic helminth infections (intestinal worms and schistosomes)

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

on cognitive functioning and educational achievement and the efficacy of treatment for such infections at improving these cognitive and educational outcomes. It was necessary to develop tests that could be easily administered by staff with a minimal amount of training and materials and that could be quickly administered to the very large number of children in an epidemiological study. These features mean that the tests, as they appear here or as adapted to other languages, could be readily used for other large-scale studies on the impact of various interventions on educational achievement. Adaptation itself would be easily implemented, especially since the methods used for construction of these tests did not rely on frequency tables and complex educational statistics (which were not available for the language in question) but on readily available reading books and general knowledge. Following adaptation and standardization, the tests could be used to assess the reading ability of individual children who are learning to read in a regularly spelled language. The reading tests are thus flexible and easily adaptable to other languages.

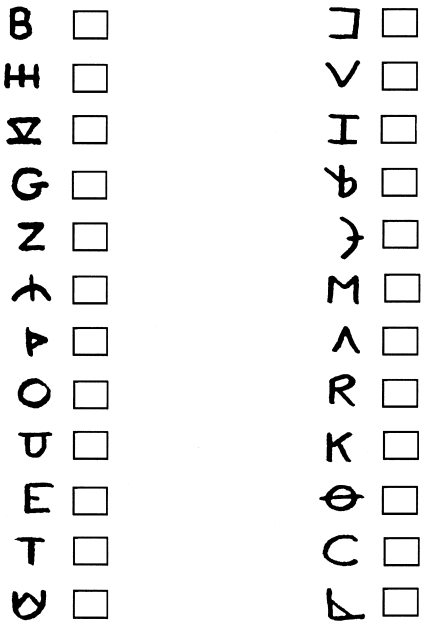
While single-word reading tests are principally used to evaluate the progress of beginning readers, the irregular spelling-to-sound characteristics of English have allowed the development of tests that implicitly measure vocabulary. The most widely used of these, the National Adult Reading Test (NART), was developed by Nelson and O'Connell (1978) to estimate the premorbid verbal intelligence of brain-damaged patients. It relies on the fact that, once learned, the capacity to read and to pronounce correctly a familiar word is extremely robust, withstanding the effect of brain damage that is sufficient to impair most measures of intelligence. It is, of course, difficult to produce equivalent tests in regularly spelled languages for which all words are readily decodable. In some cases, such as Spanish, it is possible to identify a small class of infrequent, irregularly accented words that can then be presented in an unnatural, accentless form (Del Ser Quijano et al., 1997). However, in Spanish even skilled beginning readers are able to read aloud 75% of pseudowords and 94% of words (Signorini, 1997), suggesting that this test would be much less sensitive than the NART. As in the present study, however, it is possible to bypass this problem by using lexical decision making. The Spot-the-Word test (Baddeley, Emslie, & Nimmo-Smith, 1993) presents the reader with pairs of items, each including a real word and a pseudoword. By varying the frequency of the words it is possible to vary the difficulty, thus providing an indirect test of vocabulary which correlates highly with standard tests of verbal intelligence. In this test, the choice can be based on a number of factors, such as knowledge of meaning or recognition at the orthographic or phonological level. Again, it appears to be highly resistant to the effects of brain damage and, as in the case of our reading test, can in principle be applied to any language, regardless of regularity.

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

APPENDIX

EXAMPLES OF TESTS

Letter reading form A



Word reading form A

Date _____ **Name** _____

Class _____ **School** _____.

1. safi	☐	13. lanua	☐
2. migwa	☐	14. kiazi	☐
3. tefe	☐	15. mwalimu	☐
4. shamba	☐	16. chajoka	☐
5. kila	☐	17. kitu	☐
6. homa	☐	18. makindo	☐
7. umbiga	☐	19. shida	☐
8. mbapo	☐	20. afya	☐
9. maharage	☐	21. fako	☐
10. mchicha	☐	22. mbuva	☐
11. wagoto	☐	23. ngoma	☐
12. kumoma	☐	24. kigo	☐

Sentence reading practice

Do snakes live in the sky?	☐
Is it good to eat buns?	☐
Do people live in houses?	☐
Do dogs crow early in the morning?	☐

ACKNOWLEDGMENTS

We would like to express our thanks for many helpful comments on test development to Sally McGregor of the Institute of Child Health, London, and for facilitating the project to Elizabeth Yona of the Ministry of Education, Dar-es-Salaam, Tanzania. We also thank all of the staff of the MAKWAMI project in Bagamoyo, Tanzania, for their help in preparing the test materials and facilitating the collection of the data reported in this article. The study was carried out under the auspices of Partnership for Child Develop-

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

ment (PCD), Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, Oxford University, and Ushirikiano wa Kumwendeleza Mtoto Tanzania (UKUMTA; The Tanzania Partnership for Child Development). Funding was provided by the James S. McDonnell Foundation, grant reference 94-13.

REFERENCES

Baddeley, A., Gardner, J. M., & Grantham-McGregor, S. (1995). Cross-cultural cognition: Developing tests for developing countries. *Applied Cognitive Psychology, 9*, S173–S195.

Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1992). *The Speed and Capacity of Language Processing (SCOLP) Test: Manual*. Bury St. Edmonds: Thames Valley Test Company.

(1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology, 32*, 55–65.

Carrillo, M. (1994). Development of phonological awareness and reading acquisition: A study in Spanish language. *Reading and Writing, 6*, 279–298.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour, 8*, 240–247.

Cossu, G., Shankweiler, D., Liberman, I. Y., & Gugliotta, M. (1995). Visual and phonological determinants of misreadings in a transparent orthography. *Reading and Writing, 7*, 237–256.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*, 1–16.

Del Ser Quijano, T., Gonzalez-Montalvo, J.-I., Martinez-Espinosa, S., & Delgado-Villapalos, C. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition, 33*, 343–356.

de Manrique, A. M. B., & Signorini, A. (1994). Phonological awareness, spelling and reading abilities in Spanish-speaking children. *British Journal of Educational Psychology, 64*, 429–439.

Ellis, A. W. (1993). *Reading, writing and dyslexia*. Hove: Erlbaum.

Giannouli, V., & Harris, M. (1997). *The relationship of phonemic awareness to reading and spelling in Greek preschool and primary school children*. Paper presented at the European Developmental Psychology Conference, University of Rennes, France.

Harris, M., & Giannouli, V. (1999). Learning to read and spell in Greek: The importance of letter knowledge and morphological awareness. In E. Margaret Harris, E. Giyoo Hatano et al. (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 51–70). New York: Cambridge University Press.

Inter-territorial Language (Swahili) Committee to the East African Dependencies. (1956). *A standard English–Swahili [and Swahili–English] dictionary (founded on Madan's English–Swahili [and Swahili–English] dictionary)*. Oxford: Oxford University Press.

Jastak, J., & Jastak, S. (1965). *The Wide Range Achievement Test*. Wilmington, DE: Guidance Association.

Klicpera, C. (1989). The reading development of normal and poor readers in the first grade: How helpful is the concept of developmental stages for the understanding of reading acquisition in German-speaking children? In M. Brambring, F. Losel, & H. Skowronek (Eds.), *Children at risk: Assessment, longitudinal research, and intervention* (pp. 97–118). Berlin: de Gruyter.

Naslund, J. C., & Schneider, W. (1991). Longitudinal effects of verbal ability, memory capacity, and phonological awareness on reading performance. *European Journal of Psychology of Education, 6*, 375–392.

(1996). Kindergarten letter knowledge, phonological skills, and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology, 62*, 30–59.

Nelson, H. E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the new adult reading test. *Cortex, 14*, 234–244.

Oeney, B., & Durgunoglu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. *Applied Psycholinguistics, 18*, 1–15.

Porpodas, C. D. (1986). The effect of articulatory suppression on Greek children's reading of Greek. *Educational Psychology, 6*, 115–122.

(1991). The relation between phonemic awareness and reading and spelling of Greek words in the first school years. In M. Carretero, M. L. Pope, P. R. Simons, & J. I. Pozo (Eds.),

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

Learning and instruction: European research in an international context (Vol. 3, pp. 203–217). Oxford: Pergamon.

Schonell, F. E. (1940). *The Schonell Diagnostic English Tests*. Edinburgh: Oliver & Boyd.

Signorini, A. (1997). Word reading in Spanish: A comparison between skilled and less skilled beginning readers. *Applied Psycholinguistics, 18*, 319–344.

Simeon, D., Callender, J., Wong, M., Grantham-McGregor, S., & Ramdath, D. D. (1994). School performance, nutritional status and trichuriasis in Jamaican schoolchildren. *Acta Paediatrica, 83*, 1188–1193.

Thorstad, G. (1991). The effect of orthography on the acquisition of literacy skills. *British Journal of Psychology, 82*, 527–537.

Valle-Arroyo, F. (1989). Reading errors in Spanish. In P. G. Aaron & R. M. Joshi (Eds.), *Reading and writing disorders in different orthographic systems* (pp. 163–175). Dordrecht: Kluwer.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14*, 1–33.

(1996). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology, 61*, 80–90.

Wimmer, H., & Hummer, P. (1990). How German-speaking first graders read and spell: Doubts on the importance of the logographic stage. *Applied Psycholinguistics, 11*, 349–368.

Wimmer, H., Landerl, K., & Schneider, W. (1994). The role of rhyme awareness in learning to read a regular orthography. *British Journal of Developmental Psychology, 12*, 469–484.

Zucchermaglio, C., Pontecorvo, C., Tonucci, F., & Blachowicz, C. L. (1986). Literacy and linguistic awareness: A study of Italian first grade students. *Reading Psychology, 7*, 11–25.

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716400004069 Published online by Cambridge University Press