APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **15**: 433–443 (2001)
DOI: 10.1002/acp.715

A Computerized Test of Speed of Language Comprehension Unconfounded by Literacy

JON MAY,[1]* KATHERINE J. ALCOCK,[2] LOUISE ROBINSON[1] and CHRISTINA MWITA[3]

[1]*University of Sheffield, UK*
[2]*Partnership for Child Development, Oxford University, UK*
[3]*Makwami Project, Bagamoyo, Tanzania*

SUMMARY

A computerized version of the Silly Sentences task developed for use with children is found to be equivalent to the pencil-and-paper version from the SCOLP test with UK undergraduates, and is usable by a sample of young UK children. Because the sentences are presented aloud instead of being written, the computerized test is not affected by literacy skills. Translated into Kiswahili, the task was used in Tanzanian schools, despite the absence of an electricity supply and a very different cultural background. The decision latencies had a test–retest reliability of 0.69 over 5 months, and were independent of age and baseline decision speed. The task appears appropriate for longitudinal studies, including those in developing countries. Given its simplicity and the correlations with the original SCOLP version of the task, it may also be useful in studies on literate adults. Copyright © 2001 John Wiley & Sons, Ltd.

INTRODUCTION

The Speed and Capacity of Language Processing Test (SCOLP, Baddeley *et al.*, 1992) was originally developed to provide a brief and easily administered holistic measure of an individual's efficiency of language comprehension. A pencil-and-paper test, it consists of a 'speed of comprehension' task and a 'spot the word' task. The latter provides a measure of crystallized verbal intelligence to serve as a baseline for interpreting the first task, which requires people to work through a list of 100 sentences in two minutes and mark them as true or false. Since the 50 false sentences are often bizarre (e.g. 'Nuns can be bought in shops') this test is also known as the 'Silly Sentences' task (e.g. Baddeley *et al.*, 1995).

This task had its basis in an experimental programme investigating the Teachable Language Comprehender model of Collins and Quillian (1969), but its interpretation

*Correspondence to: Jon May, Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TP, UK. E-mail: jon.may@shef.ac.uk

Contract grant sponsor: Partnership for Child Development.
Contract grant sponsor: Ushirikiano wa Kumwendeleza Mtoto Tanzania.
Contract grant sponsor: James S. McDonnell Foundation.
Contract grant number: 94-13.

within this framework transpired to be too complex for analytical use. In the course of experimentation, however, it became apparent that the task could be useful as a 'sensitive indicator of the effects of environmental stress, and subsequently as a neuropsychological measure' (Baddeley *et al.*, 1992, p. 5). As a component of the SCOLP it has now been shown to be a useful assessment tool in detecting mild cognitive impairment following mild head injury (Hinton-Bayre *et al.*, 1997) and more generally as an indicator of 'dysexecutive syndrome' (Papagno and Baddeley, 1997), a deficit in planning and attention switching.

This application is particularly valuable because mild generalized neurological deficits may not be detectable by conventional neuropsychological measures that target impairments related to specific lesions or gross organic damage. Mild levels of disability can evade detection due to the raised motivation and effort associated with psychological assessment, while still impairing performance on continual, everyday tasks. This problem has dogged attempts to investigate the effects of malnutrition and parasitical infection upon cognitive development, where it is thought that enduring changes in day-to-day motivational levels may result in slower or impaired cognitive development (Strupp and Levitsky, 1995). The Silly Sentences task, in its brevity and simplicity, may offer a measure at which individuals always perform at their own particular 'ceiling', avoiding such motivational confounds, and so being more suitable for studies involving malnourished children (Gardner *et al.*, 1996).

The pencil-and-paper version has its limitations in this context, of course. It is dependent upon literacy and verbal ability (hence the need in the test for the 'spot the word' task), and since it was designed for British adults, the items assume a certain cultural background. To overcome these problems as part of an investigation into the effects of intestinal parasitical infection in Jamaican children, Baddeley *et al.* (1995) produced an oral version of the task, with simpler materials that they hoped would be understood by children world-wide, such as 'that fire is hot, and that the moon shines at night' (p. S179). The items were changed from sentences to questions such as 'Do cows live under water?' to make the task more natural to children. To administer the task, the experimenter read the forty questions aloud to the child, waiting for them to reply 'yes' or 'no'. The whole administration was timed, and divided by 40 to obtain a child's score (around 2.5 seconds in Gardner *et al.*).

Despite problems in consistently reading the items, in dealing with errors, and in accurately timing responses, test–retest reliabilities on 19 children over one week of 0.83 and 0.89 were achieved for the two testers. Over three months, the test–retest reliability over a sample of 145 children remained high, at 0.72. Despite being specifically designed to avoid any demands on literacy, the decision latencies correlated with measures of scholastic performance. Baddeley *et al.* concluded that they were 'happy with the acceptability, reliability, and validity of the test and would recommend its continued use' (p. S188). In studies such as theirs, where a small number of trained administrators conducted all the testing sessions, few problems would arise from any variability in the rate of speech, in intonation during the test, or in timing decision latencies. In larger-scale studies, extended over longer time periods, and spread over a wider geographical area, such variations in test administration could be problematic. Presentation variability could be controlled by using a portable, battery-powered tape recorder, leaving only the timing to the administrator. At more technological cost, a computerized version of the task could both present the test items and collect precise decision latencies, with a number of procedural advantages.

In any testing situation test reliability and appropriateness for the testing situation and the subjects are important factors that sometimes clash with each other. This is especially true when working with children, or working in rural situations in developing countries, where subjects can be naïve or nearly naïve not only of the standard psychological testing situation but of any testing situation or formal learning situation.

Children who are attending school in rural areas of developing countries may be more experienced in a formal learning situation but that situation is very structured, takes place in a large group, and making an error involves personal cost. They are therefore unfamiliar with the situation of being tested one-to-one and apprehensive about being tested. Computerized testing depersonalizes the testing situation, and by hiding the assessment and response recording aspects of the testing within the 'black box' of the computer, can reduce test anxiety.

Psychometric testers who are familiar with local language and customs, in addition, are difficult to find and train. Where there are no local training programmes, staff who are trained locally in another field may find the transition to psychometric testing, with all its demands, difficult; staff who have been trained abroad, even if they are from the research area, may find problems translating their skills to their home country. Standardization of test presentation and data collection is clearly facilitated through computerization, and training in use of the computer and the associated software and hardware is less problematic than training for psychometric test administration.

The aim of the studies reported in this paper was to develop a computerized version of the Silly Sentences task, and to check that the scores it produces correspond to those obtained from the existing pencil-and-paper task from the SCOLP test. We also wanted to assess the use of modern, laptop computers in developing countries, where testing conditions can be *ad hoc* and electrical supplies unavailable, and to determine whether the task survived translation into another language and cultural background.

THE COMPUTERIZED SILLY SENTENCES TASK

This version of the Silly Sentences task was programmed using Psyscope (Cohen *et al.*, 1993), to run on Macintosh computers. The 80 sentences prepared by Baddeley *et al.* (1995) were divided into two parallel forms (hereafter Forms A and B; materials are available from the authors—see Acknowledgements). The sentences were read by a UK-English speaker, and digitized into individual sound files. The Psyscope script allows either form to be selected, and following two practice blocks of six and ten sentences, plays forty test sentences in a fixed order, waiting for the listener to press a 'yes' or 'no' key on a CMU button box (this contains a millisecond timer and connects to the Macintosh serial port) before continuing with a two-second pause and then the next sentence. There is a rest break after twenty sentences, which the experimenter terminates when asked by the participant.

Instructions for the experimenter are displayed on the screen, which can be positioned some metres away, (depending upon the length of the cables available), so the respondent can focus on listening to the sentences and responding. Each response results in a 'beep', whether correct or incorrect. The Psyscope script records the keypresses made to each item and their latency (from the start of the sentence).

The decision latencies are timed from the start of each sentence, rather than from the end, to avoid the problem of missing responses when the child anticipates the answer before the sentence has ended. While this adds time to the response measure, it is at least consistent for each child, and avoids the difficulty of deciding exactly when a sentence has in fact 'ended', and in 'clipping' the sound file too abruptly.

EXPERIMENT 1: COMPARISON OF COMPUTERIZED AND PENCIL-AND-PAPER TASKS

Method

Sixty-four first year undergraduates at the University of Sheffield, including 42 females, 18 males, and 4 who did not identify themselves as male or female, took part in the first phase of the experiment. The students were approached at the end of a Laboratory Class and asked to participate. Those who agreed were then given one minute to attempt as many items as they could from the paper-and-pencil version of the Speed of Comprehension test (Version A) from the SCOLP battery. Four months later, the students were recontacted and asked to take part in phase two of the experiment, using the computerized task. Twenty-one students consented (19 female and 2 male, with ages ranging from 18 to 33, mean 19 years).

The participants were all tested individually, in a private testing cubicle containing a Macintosh 5200 PowerPC and a CMU Button Box. They were seated so that the button box was within easy reaching distance, but so that they could not see the computer screen. They were asked to press the green, right-hand button of the box in response to sensible sentences and the red, left-hand button box in response to silly sentences. They were warned that some of the sentences were very bizarre. They were then given a pair of headphones to wear, through which the Sentences were presented.

Eleven heard Form A, and ten heard Form B. The first sentence began one second after the start of the block, and each subsequent sentence began playing two seconds after the response to the previous sentence had been made.

Results

The first sample of sixty-four participants correctly answered an average of 69.4 sentences from the SCOLP Speed of Comprehension scale. The twenty-one recontacted for the second phase had obtained a mean of 66.5 sentences, which did not differ significantly from the full sample ($t = 0.92$, df $= 62$, n.s.). Those selected to receive Form A of the computerized task had a mean of 66 on the SCOLP scale, while those who received Form B had a mean of 67.1, and these means did not differ significantly ($t = 0.89$, df $= 19$, n.s.)

The mean latency to answer items on the computerized task was 1692 milliseconds for Form A and 1691 milliseconds for Form B, measured from the time at which each sentence began to be played. Unsurprisingly, these latencies did not differ significantly ($t = 0.99$, df $= 19$, n.s.) and so the two Forms can be regarded as equivalent. Most participants answered all items correctly (four made one error, one made two errors, and one three—latencies for these errors have not been included in these analyses).

Pooling all twenty-one subjects' data, the SCOLP score correlated significantly with decision latencies (Pearson's $r = -0.583$, df $= 19$, $p < 0.05$). The faster participants could respond to items in the computerized version of the task, the more items they could answer in the pencil-and-paper task.

Discussion

The correlation in the undergraduates data between the pencil-and-paper and the computerized tasks suggests that the two tasks are measuring the same capacity, even though the items are completely different and those within the latter task were designed for much younger participants. The two Forms of forty sentences in the computerized task resulted in almost identical latency measures, indicating that they are of equivalent difficulty, and so can be used to assess test–retest reliability and as pre- and post-intervention measures. As with all speeded measures of ability, however, this correlation could be due to factors such as age, general ability, or baseline response speed. The next two studies address these possibilities, and also evaluate the test with its intended population: young children, including some who are preliterate and from a non-Western culture. The first step is to try out the computerized task with a younger age group, and to check that it is not correlated with a measure of general intelligence. The expectation would, of course, be that higher scores on a test of general intelligence should be matched by faster sentence decision latencies, but if the correlation between intelligence and decision latency were too high, this particular task would not be measuring anything other than intelligence.

EXPERIMENT 2: USE OF COMPUTERIZED TASK WITH YOUNG CHILDREN

Participants

A sample of thirty-five children were recruited from a mainstream school, being asked to participate by their headteacher. Of these, 18 were male, 17 female, and ages ranged from 7–11 years, with a mean of 9 years.

Method

The testing in this experiment was carried out in the children's schoolrooms, using a Macintosh PowerBook 190, the CMU Button Box, and portable power-speakers (instead of headphones). The procedure was the same as for the computerized testing session reported in Experiment 1, except that after the computerized task each child also completed sets A and B of Raven's Coloured matrices (Raven, 1947). After completing A1 as a practice item (which they all did successfully) they were asked to complete the other 23 items, with the experimenter turning the pages.

Results

All but nine children made one or no errors on the Silly Sentences task, the mean number of errors being 0.43 out of 40 items. Errors were excluded from subsequent analyses. The remaining correct decision latencies were examined within-participant, and any greater than three standard deviations from a child's own mean were treated as outliers due to inattention or a missed button-press. A total of 22 times were excluded using this criterion, no more than one per child. The remaining mean correct decision latency for the 35 children was 2.48 seconds (with a standard deviation of 0.27 seconds), which is virtually identical to the 2.5 seconds reported by Gardner *et al.* (1996).

The children all performed well on the Ravens Matrices, with a mean of 20.3 items correct (standard deviation 2.3), although only seven scored the maximum 23. The Ravens score did not correlate significantly with the decision latencies (Pearson's $r = 0.09$, Spearmans's $r = 0.11$; both n.s.).

Discussion

This experiment has shown that the computerized task is capable of being used successfully with children, and that it cannot be predicted by a measure of general ability.

The possibility remains that these results could be due to baseline response speed, and so this is addressed in the next experiment by including a choice reaction time task that requires a similar response decision to be made, but which does not test any language comprehension. This study also examines the suitability of the task for use with a non-English-speaking sample of children, in a developing country.

EXPERIMENT 3: USE OF COMPUTERIZED TASK WITH NON-ENGLISH-SPEAKING CHILDREN

Participants

This study was carried out in Tanzania, in conjunction with a larger-scale study into the effects of intestinal parasitical infection (either hookworm or *Schistosoma haematobium*) upon cognitive performance, and data were collected by five local Kiswahili-speaking experimenters, none of whom had any previous experience in psychological testing or with computerized testing (more details about this study can be found in Alcock *et al.*, in press). A total of 618 children took part in the larger study, which included treatment for parasitic infections. Only the data from the 69 children who were uninfected are presented here. These children were aged between 9;6 and 15;1 (mean 11;10). Each child was asked to complete the task on two occasions, separated by around 5 months, all being tested in their schools. It was intended that each child would complete both versions of the task (i.e. the two Forms of 40 Sentences), one on each occasion, with half of the sample completing the versions in each order. One child missed the first testing session, four the second, and three inadvertently completed the same version of the test on both testing occasions, leaving 61 participants.

Method

The Sentences were translated by a local colleague, who spoke the same dialect of Kiswahili as the children, and digitized. It was possible to translate most of the sentences without difficulty, but we did have to alter two of the practice items and fifteen of the test items for cultural or linguistic reasons. These sentences replaced the UK-English language versions (a matter of deleting the original files and replacing them with the Kiswahili files, the Psyscope script remaining unaltered). The hardware and software used for the presentation of the stimuli and the recording of the results was identical to that used in Experiment 2, except that power was provided by batteries, there being no electricity supply available at the schools where testing was carried out. The speakers and the Button Box ran off standard commercial batteries, and while the Powerbook internal batteries needed to be recharged daily, they provided sufficient power for a day's testing of up to seven sessions.

Because of the potential difficulties of using computerized testing equipment in a developing country, with children who have never encountered such technology before, it is worth describing our testing procedures in some detail.

Before the testing began, the child had already done some other warm-up activities such as drawing a picture or playing a familiar game. Children sat at a low table in an empty classroom or other unused room in their own school with the tester, sitting at right angles to the tester. The tester arranged the equipment so that the button box was directly in front of the child and the speakers were behind that, facing the child. The computer was placed in front of the tester and the battery for the button box was placed behind the speakers so that the child could not disconnect it.

The child was first asked what they thought the speakers were. Most children replied that they were a radio or speakers, as all children were familiar with these items. The children were then told that the button box would give out some sounds via the speakers, which they would hear and respond to by pressing the buttons on the box. The children were invited to press the buttons themselves and if they were reluctant to do so the tester showed them how to. The computer was then set up and the children asked if they knew what it was. Following the inevitable hesitation they were told it was called a computer and were asked if they had ever heard of a computer. They were then told that the computer was used to type words, such as their name, and also would play various sounds to them.

The test started with the choice reaction time task. In this task children were required to press one button when they heard a dog barking and another when they heard a bird chirping. Before starting the test, therefore, children were allowed practice time to try pressing the buttons, and to ensure that the buttons were pressed quickly and then released. Above each button was a small cartoon of the appropriate animal, drawn by a local artist, and the children were asked what these pictures represented while the tester put the pictures in place. Most children recognized the pictures. The children were then asked to demonstrate for the tester what sound the two animals would make. Any children who were reluctant to do so were cued by the tester. Almost all children produced sounds that were similar to the sounds used in the test.

Children were then given some off-line practice by the tester, with the sounds produced by the tester while the child practised pressing the buttons as quickly as possible. It was emphasized to the children that they must press as fast as possible. This pre-practice session was introduced after it was found that children were a little confused when confronted with the computerized practice session immediately. It also helped to demonstrate to the children that there could sometimes be duplications of a sound, rather than strict alternations, which was the impression some of them got from the fairly short practice session. The sounds were, however, presented randomly at 6-second intervals during both practice and test sessions by the program.

The children were then given a 20-trial practice session. During this they were encouraged to go fast but to try not to make mistakes. This exhortation was repeated after the practice session, which was generally given to the children twice unless they seemed to be making few or no mistakes and were obviously making a great effort to go fast. If the children seemed to be making a large number of mistakes or merely alternating pressing the two buttons, then the practice session was repeated a third time. Following this the child undertook the test session of 60 randomized trials.

After completing the choice reaction time tasks, the two pictures of the dog and the bird were removed from the button box to reveal a cross and a tick mark over the left- and right-hand buttons respectively. The child was asked what these marks represented—in Kiswahili the names for these marks translate as 'wrong mark' and 'correct mark' respectively, so that the mapping of the name of the mark to the generally accepted meaning in a school context is very direct. The children were then asked two questions similar to the questions used in the Silly Sentences task and asked to say if these were true or untrue, e.g. Do people have 25 heads? and Do you eat food?

Following the child's response, they were shown the correct button and told 'OK, if the question is (true/untrue) you should press the button next to the (tick/cross) mark'. The need for speed and accuracy was emphasized during the practice sessions. After the first set of 20 sentences if the child seemed to be tired or their attention was wandering, they were reminded of the task and told that it was nearly finished. However, fatigue was rare. Children enjoyed the task and only one child out of 618 refused to do the task.

Results

Although error rates were low overall, some of the translated sentences were clearly ambiguous, or had changed from Yes to No. 'Can stones bite you?', for example, had been translated as 'Can stones hurt you?' which they clearly can, if thrown accurately. 'Do cows have wings?' was mistranslated as 'Do goats have fur?'. 'Yes' answers to these items were scored as correct. Although this did unbalance the number of Yes and No answers in Form B, it is unlikely to have affected the results. 'Is water wet?' became the ungrammatical 'Is water soaked?' which received almost equal numbers of 'Yes' and 'No' responses (the expected Yes response being scored as correct). Responses to this item have been retained within the analyses reported here for reasons of conservatism, although it would be preferable to revise the materials before the test is used again.

The mean response time and number of errors (out of 40) for each testing occasion are shown in Table 1. Paired t-tests showed that there was statistically significant speeding in the responses over time, although the absolute improvement is small (134 ms). The number of errors did not decrease significantly (there being 0.77 fewer errors, on average). The correlations between the two testing sessions are strong. One child responded over a second slower than the others on the first testing occasions, but omitting this child from the analysis does not change the correlation between the reaction times.

The CRT task carried out on the first testing occasion resulted in a mean baseline CRT of 677 ms ($\pm$ 170), and this correlated significantly with the Silly Sentences response times obtained on the first occasion (CRT $\times$ first SS rt: $r(61) = 0.27$, $p < 0.05$) but not the second (CRT $\times$ second SS rt: $r(61) = 0.233$, n.s.). Despite these low correlations, the regression

Table 1. Descriptive and inferential statistics for the 61 uninfected Tanzanian children's two testing occasions

Occasion	Sentence decision time (ms)	Sentences wrong (out of 40)
First	3896 ± 355	5.16 ± 3.43
Second	3762 ± 259	4.39 ± 2.61
	$t(60) = 4.09$, $p < 0.001$	$t(60) = 1.80$, n.s.
	$r(61) = 0.69$, $p < 0.001$	$r(61) = 0.41$, $p < 0.001$

equations were used to compute residual scores for the Silly Sentence response time for each testing occasion, and these were still found to correlate significantly ($r(61) = 0.67$, $p < 0.001$).

The age of the children also correlated significantly with the response time on the first Silly Sentences session ($r(61) = 0.33$, $p < 0.05$) but not the second ($r(61) = 0.04$, n.s.). Residual scores for the Silly Sentence response times using Age as a predictor were also calculated for each testing session, and were found to correlate significantly ($r(61) = 0.68$, $p < 0.001$).

Using these two variables together in forced multiple regressions (inspection of the data suggested that assumptions of normality, homoscedascity and linearity were met) to predict the two decision latencies had similar results (session one, $r = 0.34$; session two, $r = 0.27$), with the residual latencies still correlating significantly ($r(61) = 0.67$, $p < 0.001$). These three regression analyses, and the test–retest correlations between the residual decision latencies, indicate that, even when variance predicted by the child's age and choice reaction time is excluded, the decision latencies are reliable across testing sessions.

The presence of correlations for the first testing session, when the task and the testing situation was novel, and their absence on the second session three months later when the task, although strange, had at least been completed once before, raises the possibility of practice effects. Despite the 16 practice trials, the less able children might take longer to reach a stable level of performance during the task, and we might be measuring this 'speeding up speed' rather than the hypothesized holistic speed of language comprehension. To eliminate this, we examined the childrens' performance on the two halves of each testing session separately (i.e. Sentences 1 to 20 and 21 to 40). Latencies from the first 20 sentences correlated across testing sessions ($r(61) = 0.56, p < 0.001$), as did latencies from the last 20 sentences ($r(61) = 0.46, p < 0.001$). Although the relationship reduces slightly, it is still acceptably reliable.

GENERAL DISCUSSION

Taken together, the results of these three studies lead us to be confident that this computerized task is tapping the same language comprehension skills as the Silly Sentences task from the SCOLP battery, without being confounded by literacy skills. In UK undergraduates, sentence decision latencies were related to the number of the written SCOLP items completed in two minutes. The computerized task was used successfully with a sample of young English-speaking children, and presented no major difficulties for the Tanzanian children, despite the novelty of the equipment and the task itself, the very different testing conditions, and the translation of the materials into Kiswahili. It has proven possible to collect reliable data from preliterate and non-English-speaking subjects. Given the correlation in the undergraduate sample between this 'children's' version of the task and the original pencil-and-paper version, it would also seem appropriate to use this computerized test with adult, literate participants. For all participants, it is a natural task that is easy to comprehend and perform.

The variation in sentence decision latencies between participants is not attributable to baseline decision speed on a non-linguistic choice reaction time task, nor is it simply related to the participants' age. At a reliability of 0.69 across the two testing sessions in Experiment 3, the measure is as reliable as that obtained by Baddeley *et al.* (1995), where the reliability was 0.72. At this level, the measure is suitable for intervention or longitudinal studies. The persistence of the correlations across time when only half of the test items are included suggests that it might even be possible to use four parallel forms of the test, each based on 20 items, to obtain more measuring points per participant.

The Powerbook computers stood up well to the rigours of testing in the Tanzanian schools, although by the end of the study the screen of one had broken, and we needed to transplant its hard disk to another machine to access the data. The computers had been in use for 2½ years at this point, which is an acceptable lifespan for a portable computer in constant daily use in any circumstances. We also found that computerized test administration had advantages beyond those of presentational consistency and accuracy of data collection. Responding to a recorded voice by pressing buttons is much less socially demanding than interacting with a strange and imposing adult, there did not seem to be any of the 'performance anxiety' that is often evident with pencil-and-paper tests.

The Psyscope program provides simple methods for allocating participants to different forms of a task, and for using different materials without needing to reprogram or otherwise modify the task. All that is necessary is to replace the files containing the digitized sentences with different files. Once a portable computer is included within the testing inventory, a range of additional cognitive psychology tasks become available for baseline measures of cognitive function, such as the Choice Reaction Time task included here. In other situations, we have used them to administer questionnaires, collecting the data directly and avoiding the need for time-consuming and error-prone data entry, and the need to transport and keep secure paper copies of response sheets.

Concurring with Baddeley *et al.* (1995), our view is that the measure obtained from the Silly Sentences task is a valid measure of language comprehension skills, and we recommend the use of the computerized task. In this form, which overcomes problems of consistency, test administration and data collection, and where participants are able to perform at or near ceiling accuracy without undue effort, it may be particularly useful in situations where motivational effects have made it difficult to detect cognitive deficits in the past, such as in studies of malnourishment and intestinal parasitic infection, or of dysexecutive syndrome or fatigue effects. The nature of the administration and response may also make the task suitable for use in restricted situations such as those inherent in imaging studies.

ACKNOWLEDGEMENTS

We would like to express our thanks to Alan Baddeley, Don Bundy, Sophie Scott, Kate Nokes, Professor A. Mbise, and 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 paper. We also acknowledge the helpful comments made by an anonymous referee of a previous version of the paper. The Tanzanian study was carried out under the auspices of Partnership for Child Development (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 for the Tanzanian study was provided by the James S. McDonnell Foundation, grant reference 94–13.

Test materials and Psyscope Scripts can be obtained from the first author's web pages: http://www.shef.ac.uk/~pc1jm/sentences.

REFERENCES

Alcock KJ, Mbise A, Nokes K, Musabi C, Ngowi F, Mandali R, Baddeley A. In press. The development of reading tests for use in a regularly spelt language. *Applied Psycholinguistics*.

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

Baddeley AD, Gardner JM, Grantham-McGregor S. 1995. Cross-cultural cognition—developing tests for developing. *Applied Cognitive Psychology* **9**: S173–S195.

Cohen JD, MacWhinney B, Flatt M, Provost J. 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods* **25**: 257–271.

Collins AM, Quillian MR. 1969. Retrieval time for semantic memory. *Journal of Verbal Learning and Verbal Behaviour* **8**: 240–247.

Gardner JM, Grantham-McGregor S, Baddeley A. 1996. *Trichuris-trichiura* infection and cognitive function in Jamaican school-children. *Annals of Tropical Medicine and Parasitology* **90**: 55–63.

Hinton-Bayre AD, Geffen G, McFarland K. 1997. Mild head injury and speed of information processing: A prospective study of professional rugby league players. *Journal of Clinical and Experimental Neuropsychology* **19**: 275–289.

Papagno C, Baddeley AD. 1997. Confabulation in a dysexecutive patient: Implication for models of retrieval. *Cortex* **33**: 743–752.

Raven JC. 1947. *Coloured Matrices*. Heron & Co. Ltd: London.

Strupp BJ, Levitsky DA. 1995. Enduring cognitive effects of early malnutrition: A theoretical reappraisal. *Journal of Nutrition* **125**: 2221S–2232S.