

Running head: COGNITIVE TESTS FOR UGANDAN CHILDREN

Adaptation of Western Measures of Cognition for Assessing five-year-old Semi-urban
Ugandan Children.

^{1,2}M. Nampijja (margaret.nampijja@mrcuganda.org), ³B. Apule, ¹S. Lule, ¹H. Akurut, ¹L.
Muhangi, ^{1,4}A.M. Elliott, ²K.J. Alcock

¹MRC/UVRI Uganda Research Unit on AIDS, Entebbe, Uganda.

²Lancaster University, Psychology Dept Lancaster, LA1, 4YF, UK

³Entebbe General Hospital, Entebbe, Uganda,

⁴London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Abstract

BACKGROUND: The majority of available psychometric tests originate from the Western world and were designed to suit the culture, language and socio-economic status of the respective populations. Few tests have been validated in the developing world despite the growing interest in examining effects of biological and environmental factors on cognitive functioning of children in this setting.

AIMS: The present study aimed at translating and adapting Western measures of working memory, general cognitive ability, attention, executive function, and motor ability in order to obtain a cognitive instrument suitable for assessing five-year old semi-urban Ugandan children. This population represents a particular assessment challenge as school enrolment is very variable at this age in this setting and many children are unused to a formal educational setting.

METHODS: Measures of the above domains were selected, translated and modified to suit the local culture, education and socio-economic background of the target population.

The measures were piloted and then administered to semi-urban Ugandan children aged 4;6 to 5;6, who included children that had started and not yet started school.

RESULTS: Analysis of validity and reliability characteristics showed that eight (at least one from each domain) of the 11 measures were successfully adapted on the basis that they showed adequate task comprehension, optimum levels of difficulty to demonstrate individual and group differences in abilities, sensitivity to effects of age and education, and good internal as well as test-retest reliability.

CONCLUSION: Translation and adaptation are realistic and worthwhile strategies for obtaining valid and reliable cognitive measures in a resource limited setting.

Adaptation of Western Measures of Cognition for Assessing five-year-old Semi-urban Ugandan Children.

There is a growing interest in investigating the impact of social and biological factors on child growth and cognitive development particularly in the developing world where these factors are less favourable. This has brought researchers face to face with the scarcity of psychometric tests to use amidst the wide variety of culture, language, socioeconomic and educational backgrounds.

This manuscript describes a project which was conducted within a larger study, the Entebbe Mother and Baby Study (EMaBS). The EMaBS study was designed primarily to investigate effects of worms and de-worming during pregnancy and childhood on the efficacy of childhood immunizations and on susceptibility to infectious and allergic diseases in early childhood (Elliott A.M. et al., 2007). The parent study also aims to determine the effects of childhood worms and de-worming on cognitive abilities at age 5 years. For the latter objective, we needed cognitive tests that would suit children from Entebbe town and the neighbouring villages, with a variable experience to schooling. We sought to obtain appropriate measures through translation and adaptation of existing tests rather than develop novel tests as this would be more costly in terms of time and other resources. Translation means that an already existing test is administered in the local language but otherwise the original test content is left almost intact. Test adaptation on the other hand not only translates the test but also modifies it as much as is required to suit it to the culture, education and socio-economic status of the target population. This involves writing and trying out new test items to replace unfamiliar ones, construction of new norms (administration, instructions and scoring), examining the validity and reliability of the new versions and standardizing the scores on the target population.

The degree of transformation needed appears to depend on the nature of the test and the differences between the population of origin and the population for which it is being adapted. We were particularly interested in measures of working memory, general ability, attention, executive functions, and motor abilities. Working memory was the main target since this has been implicated by most of the previous studies to be sensitive to effects of worms (Sakti et al., 1999). For example, commonly affected functions include short-term spatial memory and short-term sequential memory (Boivin et al., 1993); verbal short-term memory and reaction time (Jukes et al., 2002); free recall and fluency (Nokes et al., 1999, and Ezeamama et al., 2005) all of which are components of working memory. General ability, attention, executive function, and motor ability are believed to affect working memory performance: tests of these functions were therefore also required. In earlier studies of health and cognitive development, often only global measures of intellectual ability have been tested, but recently researchers have begun to realise that more targeted measures are necessary (Hughes & Bryan, 2003).

There are many measures of the highlighted domains; however most of these originate from America and Europe and are therefore biased to Western culture which is different from the many cultures in Africa. We would therefore argue that in their original form, Western measures would not be appropriate for Ugandan children who speak a different language, have not had experience of testing or, interacting with unfamiliar adults, come from relatively poor families and have had little or no exposure to modern technology, manufactured toys, or teaching situations which cognitive testing greatly resembles. Indeed it has been argued that even though trajectories of cognitive development are similar across cultures children are not necessarily identical in terms of cognitive style (Mandler, Scribner, Cole, & Deforest, 1980; Jahoda, 1979). There are qualitative differences in development that are specific to a local environment and these variations are related to literacy, schooling, language, ethnicity, exposure to modern technology and relative socio-economic status (Jahoda, 1979; Mandler et

al., 1980; Wagner, 1978; Miller & Meltzer, 1978). For example, Scribner (1974) and Worden (1974) observed that young children and non-schooled populations do not use categorical structure to order their recall and researchers have argued that this is because they have not practiced using that recall strategy rather than having no knowledge of taxonomic categorical organization. In contrast, in the schooled population, this ability is fostered by the teaching methods and is therefore more developed. Moreover, several studies have reported that children who attend school perform better on cognitive tests than their age-mates who do not go to school (Cole & Scribner, 1977; Sharp, Cole, & Lavc, 1979; Wager, 1978; Baddeley, Gardener, & Grantham-McGregor, 1995; Ceci, 1991; Alcock, Holding, Mung'ala Odera & Newton, 2008). Based on the previous findings, we believed that for five-year-olds in Uganda, variability in schooling experience, and the other factors mentioned above were of paramount importance in test preparation.

Our hypotheses were:

1. That the cognitive and motor measures to be adapted or translated would be suitable for all of the children in the study sample irrespective of their school exposure.
2. That children with school experience would obtain higher scores than their age-mates who have not been to school. In Uganda, children enrol into formal education strictly after their sixth birthday. However, between age 3 and 5 years, some children attend one to three years of nursery (kindergarten) school before they then go to primary 1. In the first year of kindergarten, children are introduced to the school environment.

Learning at this level is dominated by games, rhymes, songs, stories and sounds which are meant to introduce concepts like numbers, the alphabet, and the English vocabulary. In second and third years, children learn more advanced number and letter concepts including addition, subtraction, reading and writing and their vocabulary grows. Although pre-primary schooling is optional and not regarded as formal

education, we think that it may improve children's cognitive competencies in the same way, and for similar reasons, to the effects of formal education. Therefore, given the same age of children, we would expect higher scores in those who have attended pre-primary school.

Following previously published guidelines for good test adaptation (Hambleton, 1994; Van de Vijver & Hambleton, 1996), we adapted a selection of psychometric measures of the target domains to Ugandan children and evaluated their reliability and validity with regard to the study population. We also sought to find out whether there were differences in performance based on educational experience.

Tests of working memory, general cognitive ability, attention, executive function and motor abilities were selected. Most of these measures were first developed in the US and the UK, and were selected if they measured the domains of interest and preferably had a history of easy transferability across populations. Some tests were simply translated into the local language, whereas others had to be modified to a variable extent; the choice of modification depended on the test in question and its acceptability to the population. A detailed description of individual test modification follows in the methods section.

Methods

Test Translation and Adaptation

Measures of Working Memory

Sentence Repetition. The task was adapted from the Developmental Neuropsychological Assessment (NEPSY; Korkman, Kirk, & Kemps, 1997) as a measure of the phonological component of working memory. The task was developed and standardized in the US among children aged 3-12 years so it was appropriate for our participants in terms of age. The task appears transferable, to non-Western settings as it has been used in various African cultures. For example in Central Cameroon, Diller & Diller (2002) developed a

French version of sentence repetition to assess a sample of Tuki speaking adults. The measure consists of 17 English sentences of increasing complexity, which a child repeats after the assessor. Direct translation of the phrases was not possible because many English words did not exist in the local language and some words were longer and more complex when translated. Novel Luganda phrases were constructed and instructions were translated. Children were asked to repeat sentences without omitting, changing or adding word, or changing a word order. A child scored two points if no errors were committed, one point for one or two errors, and zero for more than two errors. The maximum possible score was 34.

Verbal Fluency. This was translated from NEPSY (Korkman, Kirk, & Kemps, 1997) to measure the Supervisory Attention System of working memory. This task was originally standardized on a sample of 3-6 year old American children. It has been successfully used to assess children from various countries, including Tanzania (Jukes et al., 2002), the Philippines (Ezeamama et al., 2005), Indonesia (Sakti et al., 1999), and China (Nokes et al., 1999). In this task, all children (boys and girls) name foods, animals, boys' and girls' names as fast as possible in one minute. A point is given for each correct name and the total is calculated. Because the task was easily transferable we only translated its instructions and left the content unaltered. During the pilot phase many children gave responses in both English and Luganda so responses in English were accepted provided that they were not repeated in the local language.

It should however be noted that while many psychologists continue to use Verbal Fluency to measure the supervisory attentional system of working memory, other researchers argue that the task could be more cognitively complex. For example, there is evidence that it loads other executive processes such as mental flexibility (Rende, Ramsberber, & Miyake, 2002), the ability to selectively focus attention (Shimamura, 2002), and the ability to internally generate responses at the same time plan and follow rules (Elfgren & Risberg, 1998).

Measures of General Cognitive Ability

Block Design. This was adapted from the British Ability Scales - third edition (Elliot et al., 1996) to measure non-verbal general cognitive ability. The task was developed in the UK on children aged 2;6 to 7;11. Children are required to copy and construct designs with wooden blocks as demonstrated by the assessor. There was a total of 16 items all of which were administered to each child. Correctly constructed designs were awarded one point hence a maximum score of 16. Instructions were translated and scoring allowed for rotated designs only if the rotations did not exceed 45 degrees.

Picture Vocabulary Scale. This was adapted from the Kilifi Picture Vocabulary Test (Holding et al., 2004) as measure of general verbal cognitive ability. It consists of 24 items each with four pictures: one target picture, a phonological distractor, a visual or semantic distractor, and an unrelated distractor, all drawn in black and white (see example in the Appendix). Children had to point to one of the objects on each page as requested by the assessor. Each correctly identified picture is scored one point giving a maximum score of 24 points. All the items in this task are common in our setting and were phonologically similar in the two languages so we translated their names and administered the task in the local language.

Measures of Attention

Picture Search. This was adapted from the Sky Search in the Tests of Everyday Attention for Children (TEA-Ch; Manly et al., 2001) as a measure of selective attention. The measure has proved adaptable for various populations, including Indonesian (Sakti et al., 1999), and Chinese (Nokes et al., 1999) school age children. Materials for the Picture Search consist of three A3 sheets, each with a target picture at the top; below are about 100 other pictures with copies of the target picture scattered among them. In the original version, the

total time taken to locate all the target pictures is recorded. In the modified version, children have to locate and touch as many copies of target picture as possible within 10 seconds. The score is the number of target pictures found in 10 seconds. A practice example was given before the test trials.

Measures of Executive Function

Wisconsin Card Sort Test. This task measures executive function including ability to execute a cognitive set, mental flexibility, and inhibition but it has been popular among different populations and age groups for testing mental flexibility (e.g. Renne, Bull & Diamond, 2004). There are many versions including computerized ones but a modified version of Berg's (1948) card sort test was used since playing cards are readily available and would be more acceptable in a setting where most children are more familiar with cards than with computers. In this task children are presented with four playing cards (numbers 4, 5, 6, 7 of different suits), and then given a pack of 12 cards to sort first by the number on the card (Block 1) and then by suits (Block 2). A correctly placed card is awarded one point: there is a maximum score of 12 for each block. No further instructions or corrections are given once the child starts sorting the cards.

Block 1 is intended to engage a child's mental processing with that task, whereas Block 2 is given to test the child's ability to shift to a new rule, having already encoded the initial rule. Block 2 would therefore be more challenging than Block 1. Scores on the second block are therefore used to assess the ability to shift cognitive set.

Knock-Tap-Game. This was adapted from the NEPSY as measure of executive function (inhibition). This measure has not been commonly used but it was selected for our study because it covered the age group of our participants. It was also cheap to use because other than the instructions and scoring form no more equipment was required. The task consists of two blocks each with 12 trials. The assessor taps on the table with their fingers,

knocks on the board with their fist, and makes a cutting motion in the air. In the first (imitation) block, children have to imitate the tap, knock and cut as the assessor carries them out. In the second (Opposite) block, children have to tap when the assessor knocks, to knock whenever the assessor taps and not to do anything when the assessor cuts. Block 2 is more complex than Block 1, and therefore we would expect higher scores on Block 1. Each block is preceded by a practice example. Each block is scored out of 12 but scores on Block 2 are used to compare performance on inhibition.

Tap Once Tap Twice Task. This is hand game similar to the Knock Tap Game. It was developed for our study to serve as an alternative measure of inhibition. The task has two blocks each with 12 trials, in the first (imitation) block children have to copy the assessor, tapping once or twice in imitation of the assessor. In Block 1 children have to tap once when the assessor taps twice and vice versa. Trials on the second block therefore should be more cognitively demanding than the first where children simply imitate the assessor. The score on the second block is hence a measure of the ability to inhibit prepotent responses. For both hand games described above, there is no data to our knowledge regarding their adaptability to other populations.

Measures of Motor Function

Bead Threading. This task measures fine motor function. It has been widely used in many assessment batteries for some time, such as in the Bayley Scale of Infant Development, for children from birth to 3;6 (Bayley, 1993), and in Indonesian children aged 8-13 years (Sakti et al., 1999). We translated the Bayley version into the local language and used it to assess our participants. The task requires children to thread as many beads as possible onto a shoelace as fast as they can in 20 seconds. Initially time allocated for threading was 60 seconds but because most children were threading all the 20 beads in 60 seconds, we reduced the threading time and found that 20 seconds was optimal. The score on a trial is the number

of beads threaded. Two trials are given and the average score is computed.

Coin Box. This task measures fine motor abilities. It was taken from the Kilifi Developmental Inventory (Abubakar, Holding & Van Baar et al., 2008), which was standardized on a sample of Kenyan preschool children. The task requires children to slot coins through a small horizontal slit on the box as fast as possible in 20 seconds. The score is the number of coins dropped into the box in 20 seconds. Two trials are given and the average score calculated. As for Bead Threading, we simply translated the instructions and did not need to make any other adaptations.

Balancing on one Leg. This test was taken from the Movement Assessment Battery for Children (Movement-ABC) (Henderson & Sugden, 1992). The test assesses gross motor functioning of the lower limbs and general physical fitness. Children have to balance on one leg for up to 60 seconds. There are two trials for each leg and these are preceded by a demonstration with emphasis on the rules including knees apart, the balancing leg stationary, the lifted leg well off the ground, and hands free. The assessor starts timing as soon as the child achieves balance and stops the clock when the child commits any of the above faults. The average balancing time for the four trials is computed. We gave the instructions in the local language but otherwise administered the test and scored participants as in its original form.

Training and Pilot Phase

The assessment team comprised six nurses and two doctors who had previously participated in child development assessment using the Kilifi Developmental Inventory (Abubakar et al., 2008.). The team was trained by the first author, supervised by the senior author. The training comprised of teaching as well as practical sessions. General guidelines for good assessment were emphasized; these included good participant and equipment preparation, ensuring maximum engagement of the child being assessed, dealing with needs of

individual children, maintaining speed and momentum, and accurate scoring and recording of observations. Instructions for administration of each task as described earlier were also emphasized. There are few graduate psychologists in the study area and it is important that assessment methods used in the region are easy for non-psychologists.

During piloting the tests were first administered to 30 children. These sessions were conducted by an assessor who gave the tests, with an observer present who watched and at the end of the session corrected the assessor's mistakes in the procedures. The paired assessments provided an opportunity for the assessors to learn from each other. Each of the assessors was then supervised by the trainer on at least two sessions; the trainer then judged whether the assessor was competent or needed more practice. Problems with the procedures were noted and appropriate changes made; these are detailed in the above descriptions of each test. The piloting and training exercise together took about twelve weeks and after this final versions of the measures were confirmed. Because we continued to modify the tests during the piloting phase, the data collected were not suitable for analysis and hence we do not report results from this phase.

Participants

Volunteers were sought through community-based field workers and through mothers of study participants who were asked to enrol older siblings of the study children. Inclusion criteria were that a child was aged 4;6 to 5;6, lived in the study area (Entebbe Municipality and Katabi Division), spoke Luganda at home, and was typically developing, defined as having no mental, sensory or physical handicap that was obvious to the parent or previously diagnosed by a clinician. None of the children were found to have any handicap; for the children who were excluded it was because of other reasons other than physical or mental disability. Parents were requested to bring their child's immunisation card or a birth certificate and these were used to verify the age of the child. Children were also excluded if

they had fever or were unwell as judged by a doctor at the clinic where testing took place.

A total of 65 children (30 boys, 35 girls) aged 4;6 to 5;6 were assessed. Their mean age was 5;2. One child was difficult to engage and did not attempt some of the tasks making the number of cases for these 64. Parents in urban parts of the country tend to send their children to school earlier than parents upcountry. In our study sample, over two-thirds of the participants had already enrolled into pre-primary education: 6 (9.2%) were not yet in school, 32 (49.2%) in kindergarten class1 (KG1), 12 (18.5%) in KG2, 10 (15.4%) in KG3, and 5 (7.7%) in Primary 1. Our participants were assessed early in term 1 of the academic year so those in KG1 had just 2 months of schooling at the time of testing. Children in KG2 and KG3 had attended school for 14 months and 26 months respectively by the date of assessment.

Testing Procedure

At the EMaBS study clinic, after obtaining parental consent and children's assent, eligible children were briefly assessed by a medical doctor. Children visited the toilet before the sessions and received a small snack for motivation and to avoid effects of short-term hunger. The child's mother was asked to encourage the child to participate but not give answers. The child was briefed about the 'games' they were going to participate in and was encouraged to co-operate. Tasks were administered in a fixed order unless children were difficult to engage. Sessions were conducted in an interactive play-like style in order to maximize child participation. The first 30 participants were requested to return after three weeks for retesting, and retest sessions followed the same procedure.

Ethical Approval

Ethical approval was obtained from the Uganda National Council of Science and Technology, Uganda Virus Research Institute Science & Ethics Committee, and Lancaster University Psychology Department Ethics Committee. Local council leaders were also approached for permission to recruit participants from their respective divisions - community

consent is important in the study area. Written informed consent and verbal assent were obtained from the parents and children respectively. The information sheet was read to illiterate parents who then voluntarily gave a thumbprint in presence of a witness.

Results

Descriptive Statistics

Before analysing for effects of specific explanatory variables, distribution of scores on all the tasks was examined. Descriptive statistics for all the tasks are summarized in Table 1. Performance in the various tasks generally yielded a near-normal distribution with little or no ceiling or floor effects. In the case of Balancing on one Leg, a plot of raw scores (balancing time) appeared skewed (skewness = 1.16) but after log transformation, a near-normal distribution was obtained. Ceiling effects in the initial blocks of the Knock Tap Game, Tap Once Twice, and the Wisconsin Card Sort Test were expected since these trials are designed to be easier than trials in the second blocks.

Where possible, we estimated a chance score to see if mean scores were above chance. For some measures (Sentence Repetition, Verbal Fluency, Picture Search and all the three measures of motor ability) was not possible to calculate a chance score. In the cases where a chance score was determined, we found that this was significantly smaller than the mean score (except for the second block of Tap Once Tap Twice), implying that it was not likely that children passed the various measures by chance or by guessing. Inspection of the distribution of scores on the second block of the Tap Once Tap Twice revealed that although the distribution was not clearly bimodal nor did it show a clear ceiling effect, there was some evidence that children divided into those that were guessing (hence a mean score not significantly different to chance) and those that grasped the principle of the task rapidly (with a modal score of 11, and 17 (26%) of children scoring either 11 or 12 out of 12). These results are shown in Table 1.

[Insert Table 1 about here.]

Effects of Gender, Age, and Education on Performance

We compared performance between boys and girls; there were no significant differences in mean scores between boys and girls except in Block Design where boys performed significantly better than the girls (mean difference 1.85; $p = .024$).

Using a single step regression analysis, the effects of age and schooling on performance were examined. Significant zero order relationships with age were found with all measures except Sentence Repetition, Picture Search and Wisconsin Card Sort Test. After adjusting for schooling, the age effect reduced slightly but remained statistically significant. These results are summarized in table 2.

The sample was then stratified into two groups in relation to schooling: minimal schooling [no schooling (6 children) or Kindergarten 1 (32 children)] versus more schooled [Kindergarten 2 (12 children), Kindergarten 3 (10 children) and Primary 1 (5 children)]. The effect of age was examined in each category. Results showed that in the more schooled group, a significant age effect ($p < .050$) was present in all the tasks except Sentence Repetition and Wisconsin Card Sort Test. In the less schooled group, the age effect was significant only for Block Design, Coin Box and Bead Threading. See Table 2 for these regression results.

[Insert Table 2 about here.]

Zero order relationships with schooling were significant in all the tests except Sentence Repetition and the Wisconsin Card Sorting Game and these effects remained significant after adjusting for age except in Picture Search. The schooled children were further categorized into two groups: category 1 comprised those in Kindergarten 1 or Kindergarten 2, and category 2 comprised children in Kindergarten 3 or Primary 1. Using the two schooling categories regression analysis revealed better performance for more schooled children (category 1) and the difference was statistically significant ($p < .050$) for all the tasks except

Sentence Repetition and the Knock Tap Game. Table 3 summarises the schooling effect.

[Insert Table 3 about here.]

Internal Consistency of the Measures

Internal consistency within each of the measures was examined. As shown by Table 4, the measures had good to excellent Cronbach's Alpha ranging from .65 in Picture Vocabulary Test to .90 in the Knock Tap Game and removal of some items from Sentence Repetition, Verbal Fluency, Block Design, and Picture Vocabulary Test did not change Cronbach's alpha appreciably.

Test-retest Reliability

A total of 19 participants (18 schooled) had a re-test three weeks after their initial testing and correlations between their initial and retest scores were examined. As displayed in Table 4, test-retest correlations were strong for Sentence Repetition, Verbal Fluency, Block Design, Picture Vocabulary Scale, Picture Search, Wisconsin Card Sort Test, Tap Once Twice, and Leg Balancing. Test-retest correlation coefficients were good nine and low for the remaining three ($r < .50$). There were generally better scores on the retest session possibly because of practice effects, however the differences were not statistically significant.

[Insert Table 4 about here]

Exploratory Factor Analysis

We entered all the measures into a factor analysis to see how many components would be extracted. Using an eigen-value cut-off of 1, we extracted only three components. After performing varimax rotation, component 1 showed strong loadings especially with cognitive measures namely Block Design, Picture Search, Picture Vocabulary Scale, Tap Once Tap Twice, Sentence Repetition, Knock Tap Game, and Verbal Fluency with values ranging from .47 to .79. Component 2 loaded highly on measures of motor abilities including Bead Threading (.67), Coin Box (.83), and Leg Balancing (.74) and moderately on Block Design

(.49). Component 3 loaded highly on the Wisconsin Card Sort Test (.90) and moderately on Picture Vocabulary Task (.47) and Block Design (.48). See Table 5 for a full display of factor loadings.

[Insert Table 5 here]

Discussion

Our results have revealed good psychometric properties for most of the translated and adapted versions of the measures that we chose. These features include normal distribution of scores, sensitivity to the effect of age and schooling on performance, good internal, and test-retest reliability, and meaningful associations between performance on measures within and across domains. Here, we critically evaluate each measure based on the above qualities and eventually select the tests that will be used and those that did not reach desirable standards.

Tests of Working Memory

The two working memory tasks, Sentence Repetition and Verbal Fluency achieved adequate comprehension and optimum difficulty as evidenced by wide dispersion, a normal distribution of scores and absence of floor or ceiling effects. The normal distribution also indicates that the tasks are likely to differentiate between individuals based on their abilities, and would therefore detect effects of factors that underlie these differences. Indeed Verbal Fluency demonstrated sensitivity to influences of age and schooling, but it is difficult to establish why Sentence repetition did not. The first items in Sentence Repetition were easy for most participants and could have limited the task's capacity to discriminate between individuals with small differences in the ability. Both tasks showed a high degree of stability as revealed by the high test-retest reliability coefficients and this implies that these tasks would be suitable for use in an intervention study. Based on the described values, Sentence Repetition and Verbal Fluency proved suitable measures of working memory in this sample of children. Verbal Fluency in particular has a record for good cross-cultural acceptability

(Baddeley, Gardener, & Grantham-McGregor, 1995). Other than in Western settings where it was developed, it has been successfully used in Tanzania (Jukes et al, 2002), the Philippines (Ezeamama et al., 2005), and Indonesia (Sakti et al., 1999).

Tests of General Cognitive Ability

Performance on the two measures of general cognitive ability (Block Design and Picture Vocabulary Task) was close to normal distribution and had a reasonable range suggesting that these tests have the capacity to show a distinction between individuals based on their overall cognitive ability levels. The tests are also capable of detecting effects of important exposures, and this is evidenced by their ability to show age and schooling effects on performance. Turning to test-retest reliability, these tasks appear to be stable as revealed by their fair to good test-retest correlations. Overall, both Block Design and Picture Vocabulary Task were successfully adapted and can be used to assess general ability in similar populations.

Tests of Attention

The wide range and normal distribution of scores on Picture Search implies a capacity to distinguish between more attentive and less attentive individuals. The task's sensitivity to differences in schooling reveals potential to detect effects of other important factors. Its good test-retest reliability makes it suitable for longitudinal and experimental designs with pre- and post- treatment assessments. Based on these positive features, Picture Search is considered successfully adapted to measure attention in this sample of children and similar populations.

Tests of Executive Function

In the executive function domain, the Wisconsin Card Sort Test, despite its complexity in interpretation of scores, achieved reasonable representation of individual differences in mental flexibility, as indicated by the wide dispersion and normal distribution of scores. The two parallel measures of inhibition (Knock-Tap Game and Tap-Once-Tap-Twice) both had

normal distributions and wide dispersion of scores indicating that they represented differences in the ability reasonably well. Both measures revealed effects of schooling and age, revealing a potential to detect other effects. All the three measures in this domain had good internal reliability. However, the Knock Tap Cut Game and the Wisconsin Card Sort Test had low test-retest reliability. In this domain therefore we will retain the Tap Once Tap Twice and the Wisconsin Card Sort Test for measuring inhibition and mental flexibility respectively.

Tests of motor ability

Like in many of the measures discussed above, measures of motor ability displayed normal distributions, good dispersion, good internal reliability and sensitivity to age and schooling effects. However in this domain only the Balancing on one Leg showed adequate test-retest reliability; Coin Box and Bead Threading exhibited unsatisfactory test stability and for that reason they will not be retained. It should however be noted that for motor measures and all tests of cognition, the sample used to examine test-retest reliability was small. A bigger sample would probably have yielded more accurate test-retest correlations.

The Role of Gender, Maturation, and Schooling on Cognitive and Motor performance

Our results show no differences in performance between male and female participants. Although this has been reported by previous studies and is therefore not a novel finding (see also Kerr & Zelazo, 2004; Capitan, Laiacon, Gori, & Gruppo, 1991), it reveals that our measures may be reliable with regard to gender differences.

As expected, an age effect was seen in most of the measures; progressive maturation of abilities with age has been demonstrated by almost all previous studies (e.g. Leon-Carrion, Garcia-Orza, & Perez-Santamaria, 2004; Armstrong, 2006). That we were able to replicate the age effect further supports validity of our cognitive and motor measures.

In our sample, however, schooling appeared to have a stronger influence than age especially since there many grades of schooling in a narrow age-band (4;6-5;6). Children in higher

grades of nursery school probably perform better because they have had more experience with pictures, vocabulary, recall skills, and performance strategies that enhance speed and accuracy, which would not only boost their competence in task taking but also enhance development of various cognitive abilities. Such enhancing effects of schooling on cognitive performance reported in our study have been demonstrated by many other researchers (e.g. Baddeley et al., 1995; Ceci, 1991; Cole & Scribner, 1977; Sharp et al., 1979; Wagner, 1978); this replication therefore further supports the validity of our tests.

Internal Reliability and Factor Analysis

The high Cronbach's Alpha (good internal reliability) exhibited by the measures indicates a high degree of construct validity. This feature is further revealed by the strong within-domain correlations such as those observed among measures of executive functions, even though they measured different components within the domain.

Our results also reveal close associations between the cognitive domains.

Inter-domain associations are uncovered by factor analysis which reduces all 11 measures into just three underlying components, where component 1 can be described as “general cognitive ability”, component 2 can be described as “motor ability” and component 3 suggests a more obscure latent ability common to the Wisconsin card sorting and the two measures of general intellectual ability. Based on the complexity of the three measures that load on component 3, it might be that the component describes a higher level mental ability that is not found in the rest of the measures. It should however be noted that these interpretations are based on exploratory factor analysis rather than confirmatory factor analysis which was limited by the small sample size.

Both the good internal reliability and meaningful results of factor analysis provide evidence for construct validity of the measures. That the inter-domain associations reported in previous studies have been replicated in our study supports construct validity of these measures and is

evidence that they were successfully transferred to our setting.

Summary and Conclusion

We have evaluated the suitability of the translated and adapted versions of the tests to a sample of children. Our results show that, eight of the 11 measures were successfully transferred to our setting. These included Sentence Repetition, Verbal Fluency, Block Design, Picture Vocabulary Task, Picture Search, Wisconsin Card Sorting Test, Tap Once Tap Twice, and Balancing on one Leg. Three measures including the Knock Tap Cut, Coin Box, and Bead Threading were not successfully adapted, specifically because of poor test-retest reliability. It is important that we have at least one measure of fine motor function in the battery; we hope that through improved tester training we might improve test-retest reliability of the three measures. We believe that the successful measures satisfy the standards for test adaptation as recommended by Hambleton (1994), and Van de Vijver & Hambleton (1996), and that they will effectively measure the respective functions in the Entebbe Mother and Baby study participants. The implication of our findings is that translation and adaptation are realistic and worthwhile strategies for developing valid and reliable cognitive measures in a resource limited setting.

References

- Abubakar A., Holding P., Van Baar A., Newton C.R.J.C., & Van De Vijver F.J.R. (2008). Monitoring psychomotor development in a resource limited setting: An evaluation of the Kilifi developmental inventory. *Annals of Tropical Paediatrics*, 28, 217-226.
- Alcock, K. J., Holding, P. A., Mung'ala Odera, V., & Newton, C. R. J. C. (2008). Constructing tests of cognitive abilities for schooled and unschooled children *Journal of Cross-cultural Psychology* , 39, 529-551.
- Armstrong D.F. (2006). Neurodevelopment and chronic illness: Mechanisms of disease and treatment. *Mental Retardation and Developmental Disabilities Research Reviews*; 12, 168-173
- Baddeley A.D., Gardener J.M., & Grantham-McGregor S. (1995). Cross- cultural cognition: developing tests for developing countries. *Applied cognitive Psychology*, 9,173-195.
- Bayley N. (1993). *Bayley Scales of Infant Development (2nd Edition)*. New York: Psychological Corporation.
- Berg E.A.(1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, 39, 15-22.
- Boivin M.J., Giordan B., Ndanga K., Makakala M.M, Manzeki K.M., Ngunu N., et al (1993). Effects of treatment for intestinal parasites and malaria on the cognitive abilities of school children in Zaire, Africa. *Health Psychology* 12, 220-226.
- Capitan E., Laiacon M., Gori E., & Gruppo I., (1991). Sex differences in spatial memory: A re-analysis of block tapping long term memory according to the short-term memory level. *Italian Journal of Neurological Sciences*, 12, 461-466.
- Ceci S.J. (1991). How much does schooling influence general intelligence and its cognitive components? A re-assessment of the evidence. *Developmental Psychology*, 27(5), 703-722.

- Cole M. & Scribner (1977). Cross-cultural studies of memory and cognition. In Kail R. V & Hagen J.W (Eds), *Perspectives on the development of memory and cognition*, Hillsdale, N.J: Lawrence Erlbaum.
- Daneman M. & Merike P. M. (1996). Working memory and language comprehension. A meta- analysis. *Psychonomic Bulletin & Review*, 3, 422-433.
- Diller J. & Diller K.J. (2002). Sentence Repetition Testing (SRT) and language shift survey of Tuki language. *SIL international*.
- Elfgren I.C., & Risberg J. (1998). Lateralized blood flow increases during fluency tasks: influence of cognitive strategy. *Neuropsychologia*, 36, 506-512.
- Elliott A.M., Kizza M., Quingley M.A., Ndibazza J., Nampijja M., Muhangi L., et al. (2007). The impact of helminths on the response to immunization and the incidence of infection and disease in childhood in Uganda: design of a randomized, double-blind, placebo-controlled, factorial trial of de-worming interventions delivered in pregnancy and early childhood. *Clinical Trials*, 4, 42-57.
- Elliott C.D., Smith P., McCulbough K. (1996). *The British Ability Scales* (3rd ed). Windsor: National Foundation for Psychological Research.
- Ezeamama A.E., Friedman J. F., Acosta L.P., Bellinger D. C., Langdon G.C., Manalo D.L. et al, (2005). Helminth infection and cognitive impairment among Filipino children. *American Journal of Tropical Medicine & Hygiene*, 72 (5), 540-548.
- Hambleton, R., K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-244.
- Henderson S.E.& Sugden D.A.(1992). *The Movement Assessment Battery for Children (Movement-ABC)*. New York: Psychological Corporation.
- Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., Mbonani, L. and Stevenson, J. (2004) Assessing cognitive outcomes in a rural African

- population: development of a neuropsychological battery in Kilifi, Kenya. *Journal of the International Neuropsychological Society*, 10, 1-15.
- Hughes D. & Bryan J. (2003). The Assessment of Cognitive Performance in Children: Considerations for Detecting Nutritional Influences. *Nutrition Reviews*, 61(12), 413-422
- Jahoda G. (1979). On the nature of difficulties in spatial perception task: ethnic, and sex differences. *British Journal of Psychology*, 70, 351-362.
- Jukes M.C.H., Nokes C.A., Alcock K. J., Lambo K.J., Kihamia C., Ngorosho N., et al (2002). Heavy schistosomiasis associated with poor short term memory and slower reaction times in Tanzanian school children. *Tropical Medicine and International Health*, 7(2), 104-117.
- Kerr A. & Zelazo P.D. (2004). Development of hot executive functions: The gambling task. *Brain and Cognition*, 55,148-157.
- Korkman M., Kirk U.& Kemps A.(1997). *A developmental Neuropsychological Assessment*. London: Harcourt Assessment.
- Leon-Carrion J., Garcia-Orza J., & Perez-Santamaria F.J. (2004). Development of inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience*, 114, 1291-1311.
- Mandler J., Scribner M., Cole M. & Deforest M. (1980). Cross-cultural invariance in story recall. *Child Development*, 51, 19-26.
- Manly T., Nimmo-Smith I., Watson P., Anderson V., Turner A. & Robertson I. H. (2001). The differential assessment of children's attention: The test of everyday attention for children (TEA-ch), normative sample and ADHD performance. *Journal of Child Psychology*, 42 (8), 1056-1081.
- Nokes C., McGarvey S.T., Shiue L., Guanling W., Hawai W., Bundy A.P., et al. (1999).

- Evidence for an improvement in cognitive function following treatment of *Schistosoma japonicum* infection in Chinese primary school children. *American Journal of Tropical Medicine and Hygiene*, 60(4), 556-565.
- Rende B., Ramsberger G., Miyake A. (2002). Communalities and differences in working memory components underlying letter and category fluency tasks: A dual task investigation. *Neuropsychologia* 16; 309-321.
- Sakti H., Nokes C., Subagio W.H., Hendratino S., Hall A., Bundy D.A. & Satoto (1999). Evidence of an association between hookworm infection and cognitive function in Indonesian school-children. *Tropical Medicine and International Health*, 4(5), 322-334.
- Sharp D., Cole M. & Lavc C. (1979). Cognitive consequences of education. *Monographs of the Society for Research in Child Development*, 44(178), 1-2.
- Shimamura PA. (2002). Memory retrieval and executive control processes. In *Principles of frontal lobe functions*. Edited by Stuss DT, Knight RT. New York: Oxford University Press; 210-220.
- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89-99.
- Wagner D. A.(1978). Memories of Morocco: The influence of age, schooling and environment on memory. *Cognitive Psychology*, 10, 1-28.

Table 1. Descriptive statistics for all measures.

Measure	N	Min	Max	Mean	sd	Chance score	Max possible
Sentence Repetition	64	9	30	20.55	5.24		34
Verbal Fluency	65	0	33	15.97	8.42		
Block Design	65	3	16	10.14	3.33		16
Picture Vocab. Scale	64	8	24	18.25	3.06	6.00**	24
Picture Search	65	1	10	4.4	1.82		
Wisconsin Card1	64	0	12	8.30	4.08	6.00**	12
Wisconsin Card2	64	0	12	7.06	3.66	6.00*	12
Knock Tap Cut 1	65	9	12	11.72	0.65	4.00**	12
Knock Tap Cut 2	64	0	12	7.14	4.50	4.00**	12
Tap OnceTwice1	64	0	12	10.93	2.11	6.00**	12
Tap Once Twice2	64	0	12	6.67	3.89	6.00 (n.s)	12
Bead Threading	64	1.5	7.0	4.14	1.11		20
Coin Box	64	4	15	10.50	2.18		20
Leg Balancing	64	2	56	15.75	11.21		60

* $p < .05$; ** $p < .01$ (for the difference between mean and chance scores).

n.s – not statistically significant .

Table 2. Regression analysis showing zero order and adjusted effects of age on performance.

Measure	Raw effect		Adjusted		Unschoolied		Schooled	
	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>
Sentence Repetition	.24	.17	.22	.17	.22	.18	.28	.18
Verbal Fluency	.73**	.26	.65	.26	.54	.07	.77**	.27
Block Design	.31**	.10	.27**	.10	.38**	.10	.35**	.10
Picture Voc. Task	.24*	.10	.20*	.09	.16	.12	.30**	.09
Picture Search	.11	.06	.07	.05	.04	.06	.14*	.06
Wisconsin Card Sort	.08	.12	.06	.12	.10	.12	.07	.13
Knock Tap Cut	.37*	.02	.33*	.02	.30	.06	.43**	.10
Tap Once Tap Twice	.28*	.12	.23	.05	.23	.15	.36**	.12
Bead Threading	.13**	.03	.12**	.03	.09*	.04	.15**	.03
Coin Box	.17*	.02	.15*	.07	.19*	.08	.18*	.07
Balancing	.88*	.36	.77*	.34	.49	.36	.99**	.36

* $p < .05$; ** $p < .01$

Table 3. Regression analysis showing zero order and adjusted effects of schooling on performance.

Measure	Raw effect		Adjusted for age		Education categorized	
	b	SE	b	SE	b	SE
Sentence Repetition	.92	.59	.83	.59	1.90	1.60
Verbal Fluency	2.29	.92	1.98*	.89	5.00	2.48
Block Design	1.19**	.35	1.22	.34	2.79**	.93
Picture Vocabulary	1.22**	.32	1.15**	.35	2.78**	.86
Picture Search	.99**	.17	.96**	.17	2.46**	.43
Wisconsin Card Sort	.44	.42	.41	.42	2.29*	1.07
Knock Tap Cut	1.04*	.13	.86	.14	1.33	.32
Tap Once Tap Twice-	1.19**	.42	1.08*	.42	2.18	1.07
Bead Threading	.46**	.11	.41**	.10	.99**	.30
Coin Box	.53*	.24	.47	.24	1.64	.63
Balancing	3.55**	1.21	3.23**	1.18	11.15**	3.08

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

Table 4. Internal consistency and test - retest correlations

Measure	Cronbach's alpha	Retest r (n=19)	p-value
Sentence Repetition	.88	.88	<.001
Verbal Fluency	.70	.88	<.001
Block Design	.82	.74	.001
Picture Vocabulary Test	.65	.51	.032
Picture Search	.84	.76	<.001
Wisconsin Card Sort		.23	.429
Knock Tap Game	.94	.28	.240
Tap Once Tap Twice	.88	.74	.001
Coin Box	.76	.45	.054
Bead Threading	.76	.34	.155
Balancing	.80	.61	.006

Table 5. Principal component analysis: loadings on the 3 components.

Measure	Component		
	1	2	3
Verbal Fluency	.79	-.18	.10
Block Design	.44	.49	.48
Picture Vocabulary	.54	.39	.47
Picture Search	.47	.34	.39
Wisconsin Card Sort	.04	.04	.91
Knock Tap Cut	.73	.31	-.12
Tap Once Twice	.56	.46	.05
Bead Threading	.33	.67	-.06
Sentence Repetition	.59	.21	.09
Coin Box	.10	.83	.10
Leg Balancing	.05	.74	.31

Appendix

An example of items in the Picture Vocabulary Scale.

Item 10. The target picture is the book (ekitabo); the phonological distractor is the bed (ekitanda); the semantic distractor is the pencil (kalaamu); and the unrelated distractor is the spoon (ekijjiiko).