

Constructing Tests of Cognitive Abilities for Schooled and Unschooled Children

K. J. Alcock

Lancaster University

P. A. Holding

V. Mung'ala-Odera

C. R. J. C. Newton

Kenya Medical Research Institute

It is frequently necessary to assess children with little or no schooling to determine their level of cognitive functioning, especially in developing countries. It is not possible, however, to assume that assessments will hold equal validity for children with and without the experience of schooling. The authors, therefore, set out to create a battery of tests suitable for both schooled and unschooled children. They assessed 973 schooled and 645 unschooled children in rural coastal Kenya using culturally adapted cognitive tests. Significant effects of age and schooling were found on all tests. On some tests (verbal knowledge, speeded figure matching, and pattern copying), unschooled children did not improve as much with age as schooled children. The effects of length of exposure to schooling and of age were greater than that of initial enrollment in school. The authors conclude that it is possible to assess unschooled children, but test batteries must be carefully constructed and standardized.

Keywords: *cognitive assessment; cognitive development; schooling; developing countries; Kenya*

A major challenge to psychological assessment of children is their differential exposure to schooling, which may influence test taking and the interpretation of performance levels. Such influences are particularly important in developing countries where potential risk factors for cognitive impairment are significant, due to poor medical provision and increased incidence of birth complications and of infectious diseases (Olness, 2003), but where school attendance is not universal. In the world's 50 poorest countries, nearly 40% of children never attend school (Bellamy & UNICEF, 2004), and many more attend primary school for only a limited number of years, often starting late (Jukes, Grigorenko, Alcock, Sternberg, & Bundy, 2008).

Authors' Note: This study was supported by the Kenya Medical Research Institute (KEMRI), Wellcome Trust Collaborative Research Programme. The authors thank the mapping and census team, field staff, and assessors who made this study possible. In particular, they thank Joseph Gona, Godfrey Otieno, Elizabeth Obiero, Khamis Katana, Kenneth Rimba, Gladys Murira, Judy T. Dzombo, Francis Yaa, Douglas Konde, Mary Karisa, Francis Kanyetta, Silas Haro, Karen Konde, and Janet Chea. This article is published with the permission of the director of KEMRI. C.R.J.C. Newton holds a Wellcome Trust Career Post in Clinical Tropical Medicine (050533). Data collected for this study form part of the PhD of V. Mung'ala-Odera.

In other countries, psychologists are also becoming increasingly aware of the need to have appropriate methodologies for testing children whose previous educational experience or cultural exposure differs from that of the majority population, especially when early life experiences may have placed them at risk of cognitive impairment. The population structure of developed countries is changing, largely due to immigration, including immigration by children, with documented changes in many regions including the United States, the United Kingdom, Italy, Scandinavia, and Australia (Peña, 2007). Some children born elsewhere but living in developed countries will have had little, or interrupted, schooling, and some will have been exposed to risk factors for cognitive impairment. Naturally, it is important to assess children whose cognitive status is in question, to ensure that schooling, support, and other services are appropriately aimed at the children who need them. However, if these children are assessed on cognitive tests that assume exposure to specific cultural experiences, their skills and abilities may be underestimated and their needs misrepresented (Cummins, 1984). A major danger of test bias is its potential for overrepresenting children from minority cultural backgrounds in special needs facilities (Losen, Orfield, & Civil Rights Project, 2002). Conversely, the assumption that a poor test result is merely due to a lack of educational experience may lead to inadequate provision being made for educational support.

Currently, in many developing countries, the most commonly used method for the evaluation of cognitive delay is observations made by clinicians or teachers. For example, pediatricians might assess a child's neurological function, including his or her ability to follow simple commands during the assessment, and might then base their conclusions about a child's cognitive abilities on their general impressions of the child. In some cases, these observations may follow a semistructured protocol (Kenya Institute of Special Education, 1984; Njugana, Mung'ala-Odera, Chong, Meehan, & Newton, *in press*), but these are only available in very few settings. The sensitivity of this approach to assessment will be limited and is likely to be restricted to identifying children with more severe cognitive delay. In other cases, school test performance is used. This approach may lead to the misidentification of children as cognitively delayed whose poor performance is due, in fact, to lack of exposure to school instruction or to behavioral or emotional difficulties rather than impaired cognitive skills.

Where more structured assessment methods have been used, investigators have often applied unadapted materials possibly with translated instructions (Mulenga, Ahonen, & Aro, 2001). If test materials are inappropriate or children are unfamiliar with content and procedures, test scores are likely to show restricted within-population variability, limiting the ability of investigators to appropriately characterize development (Connolly & Grantham-McGregor, 1993). An example of the application of inappropriate tools can be seen in Gay et al. (1995), whose study of children of Haitian origin living in the United States used the Bayley Scales of Infant Development (BSID; Bayley, 1993), unadapted. Deterioration on verbal standard scores was observed over time, but rather than indicating a developmental deficit, this was more adequately explained by a lack of exposure to English in the home. Pollitt and colleagues, working in Indonesia, (Pollitt, Saco-Pollitt, Jahari, Husaini, & Huang, 2000) also used the BSID, largely unadapted, and found that standardized scores deteriorated with age. This deterioration was attributed to the effect of poor nutrition in the population. The investigators did not examine the possibility that an inappropriate test could become more inappropriate with age in children who lack exposure

to the educational and cultural experiences that are typical in the standardization population and, hence, fail to improve their raw scores with age at the expected rate. The ability to investigate cognitive performance and the development of cognitive processes depends, therefore, on having adequate and appropriate assessment tools.

It is possible, however, to appropriately identify children in developing countries who have cognitive impairment, by using careful adaptation of existing tests, construction of new tests, and attention to testing techniques (Alcock, 1998; Alcock & Bundy, 2001; Alcock, Jukes, Ngorosho, & Deus, 2008; Alcock et al., 2000; Baddeley, Gardner, & Grantham-McGregor, 1995; Holding, Stevenson, Peshu, & Marsh, 1999; Holding et al., 2004; Jukes et al., 2008; Jukes et al., 2002; Mung'ala-Odera et al., 2004). In this study, we set out to develop a test that could be used to do this, for children with limited or no schooling experience.

Why Might Unschooled Children Present a Challenge to Cognitive Assessment?

The association between length of school exposure and performance on IQ tests has long been acknowledged (Ceci, 1991; Ceci & Williams, 1997; Gordon, 1923), and there is evidence to suggest that school attendance has a significant effect on cognitive development over and above that of normal maturation. In other words, a year of education increases raw scores on IQ tests more than an added year of age (Cahan & Cohen, 1989). This is despite the fact that most IQ tests are deliberately structured to avoid school-dependent knowledge. A number of explanations have been put forward to explain this effect. School exposure may be accelerating the development of cognitive skills, such that schooling leads to improvement in specific skills, possibly as a direct result of methods of instruction and/or specific curriculum content. Benefits of school attendance have been observed in specific abilities such as memory skills (Cole, Gay, Glick, & Sharp, 1971) and phonological awareness (Morais, Bertelson, Cary, & Alegria, 1986; Morais, Content, Cary, & Mehler, 1989). Further research supports the suggestion that different types of education differentially improve different cognitive functions (Gustafsson, 2001).

Other factors that might contribute to the differences between performance on tests of cognition by schooled and unschooled children include differences in the types of questions teachers and parents might ask (Waterman, Blades, & Spencer, 2000). The differences between the cultures of home and school might be greater where parents also have had limited schooling (Jukes et al., 2008). However, an alternative possibility is that in developing countries, the poor physical state of schools and the large class sizes might mitigate against such a large effect of schooling, as might the fact that many children are delayed in starting school. Previous studies in a variety of developing country settings have, however, still found effects of schooling on cognitive abilities. These include one study in the former Zaïre, which has a lower per capita income and, therefore, where the school system is significantly poorer than in the country where this study is based, Kenya (Conant et al., 2003; Conant et al., 1999). Public schooling in Kenya also takes up a fairly large proportion of children's time, with attendance for, on average, 6 hours a day, 40 weeks a year. Many children attend nursery classes attached to primary schools theoretically starting from the age of 6, with the government-recommended age for starting the first grade being 7 years. Although very few start first grade with no nursery experience, a large proportion are delayed in enrolling for a variety of reasons, including financial.

Construction of tests of overall cognitive ability in a group of children who differ in both age and school experience will enable us to examine appropriate methods and help avoid inappropriate methods of evaluating cognitive performance in children who have had little or no schooling. We should, in addition, gain some insights into the relative contribution of school exposure to developmental change. It is clear that it is only possible to assess the effect of schooling on cognitive test performance, separate from the effect of age, in a setting where not all children enter formal schooling at the same time. In most settings, it is impossible to disentangle these two effects, but here we have the advantage that, because there is no one-to-one correspondence of years of schooling with years of chronological age, the differential effects of these two factors on test performance can be compared.

Aim of the Study and Hypotheses

The primary purpose of this article is, hence, to report on the process of construction and adaptation of appropriate assessment tools for both children who are schooled and children who are unschooled. Continuing from previous studies examining test performance in schooled and unschooled children, we hypothesize as follows: that a test of global cognitive skill will not necessarily have the same measurement properties in these two groups of children, so that extra care must be taken in constructing such a test; that not all subtests will necessarily be suitable for both groups or likewise have adequate measurement properties in both groups; and that separate norms may be necessary for each group.

Using data from the same study, we will be able to investigate our secondary hypothesis: that school exposure will contribute to differences in cognitive performance on the tasks developed as part of the study. As part of this, we hypothesize that even if raw scores increase with age in unschooled children, standardized scores in these children will appear to deteriorate with age when a test has been standardized on schooled children. This will reflect similar findings in younger children tested with infant developmental scales who have not been exposed to cultural and educational experiences that would be typical in the standardization populations (Gay et al., 1995; Pollitt et al., 2000).

Test construction and validation will be carried out using a large sample of children living in a rural area of Kenya, who were recruited as part of a community survey of neurological impairment (Mung'ala-Odera et al., 2004). Children were identified by parents as either having or not having some developmental concerns; both test development hypotheses and effect of schooling hypotheses will be examined using a sample of children that is representative of the general population in terms of the number of children who parents believe to have a developmental concern.

Method

Study Site and Study Participants

The study took place at the Kenya Medical Research Institute, Centre for Geographic Medicine Research, which is located in Kilifi District, a predominantly rural area in coastal Kenya, East Africa. The area is inhabited mainly by people of the Mijikenda ethnic grouping, and a large proportion of families in the study area speak Kigiriama, an Eastern Bantu language. Economically, this area is uniformly poor and differences between households in

socioeconomic status are hard to measure. The study was part of an epidemiological investigation of neurological impairment in primary-school-age children. The construction of the test battery here described was to enable external validation of moderate or severe cognitive delay in children identified through parental report. The main project screened 10,218 children aged 6 to 9 years in the community, through demographic surveys (Mung'ala-Odera et al., 2004). Parents of these children were asked 10 questions concerning their child's motor, sensory, or cognitive development (for details of the Ten Questions Questionnaire [TQQ], see Durkin, Hasan, & Hasan, 1995). All children identified as having a delay or a problem by any of the 10 questions ("screened positive") were assessed using the battery of cognitive assessments, as well as an equal number of randomly selected children whose parents had indicated that their children were developing typically ("screened negative").

The study was approved by the National Ethics Committee in Kenya. Parents gave their informed consent for their child's participation in the study. Where a parent was illiterate, the study was explained to them orally. Children gave verbal assent to their participation.

During large-scale testing, the battery of cognitive tests was administered to 1,618 children in total (the "overall sample"); 860 children had screened positive on the TQQ, including 375 children (44%) who were not attending school and 485 children (56%) who were attending school. A further 758 children, randomly selected from the original survey population, screened negative on the TQQ, and of these, 488 (64%) were attending school and 270 (36%) were not attending school. School attendance was defined as enrollment in nursery class or higher. The highest grade represented was Grade 4. Table 1 shows the mean age in years and months of all children attending and not attending school, with the former broken down by the year of school in which they are enrolled. It is possible that the overall sample included fewer school-going children than might be found in the general population because of the overrepresentation of children who potentially suffer from developmental difficulties, those who screened positive. From the overall sample, therefore, a subsample was selected including all the children who screened negative on the TQQ, plus 78 children who screened positive. The size of the screened positive group was determined by the proportion of children from the general population who had screened positive (9.3%, because out of 10,218 children selected randomly from the community whose parents took part in the initial survey, parents of 955 answered "yes" to one or more questions; i.e., their child screened positive). This sample (the "representative sample"), therefore, intended to draw from the population in a random manner children whose screening status was representative of those in the general population. Figure 1 shows the numbers recruited, and in each of the samples and screening groups. The representative sample is used for all further analyses unless otherwise indicated. Breakdown into schooling, gender, and whole year age groups (number of completed years of age) is also included in this figure.

In the overall sample, there is a large proportion of children whose parents consider them to have difficulties and who may, for that reason, keep them at home or request that they be held back a class. This is not true of the representative sample, but it can also be seen in this sample that children do not move through school at a rate of 1 year of schooling per year of age. Nor do they necessarily start nursery class or Grade 1 at the recommended ages of 6 and 7 years, respectively. Despite this variability in amount of education for a given age, children of each age were simply divided into "attending" and "not attending" school

Table 1
Age (in years) and Schooling Distribution of Sample, Including Only
Those Children From Whom Valid Test Results Could Be Obtained

Schooling	Overall Sample			Representative Sample		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Not attending school	7.67	1.11	645	7.14	1.12	290
Nursery	7.62	1.06	571	7.10	1.05	303
Grade 1	8.50	.96	260	7.93	.93	137
Grade 2	9.03	.76	109	8.60	.67	70
Grade 3	9.36	.57	27	8.80	.52	20
Grade 4	9.46	.31	6	9.00	.00	5
Overall	7.91	1.15	1618	7.43	1.15	825

for initial analysis. This is because the aim of this study was to construct a cognitive test battery suitable for both groups of children, regardless of the number of years of schooling a child attending school happened to have completed. Following this analysis, further analyses to examine the relationship between age, school exposure, and cognitive abilities were conducted.

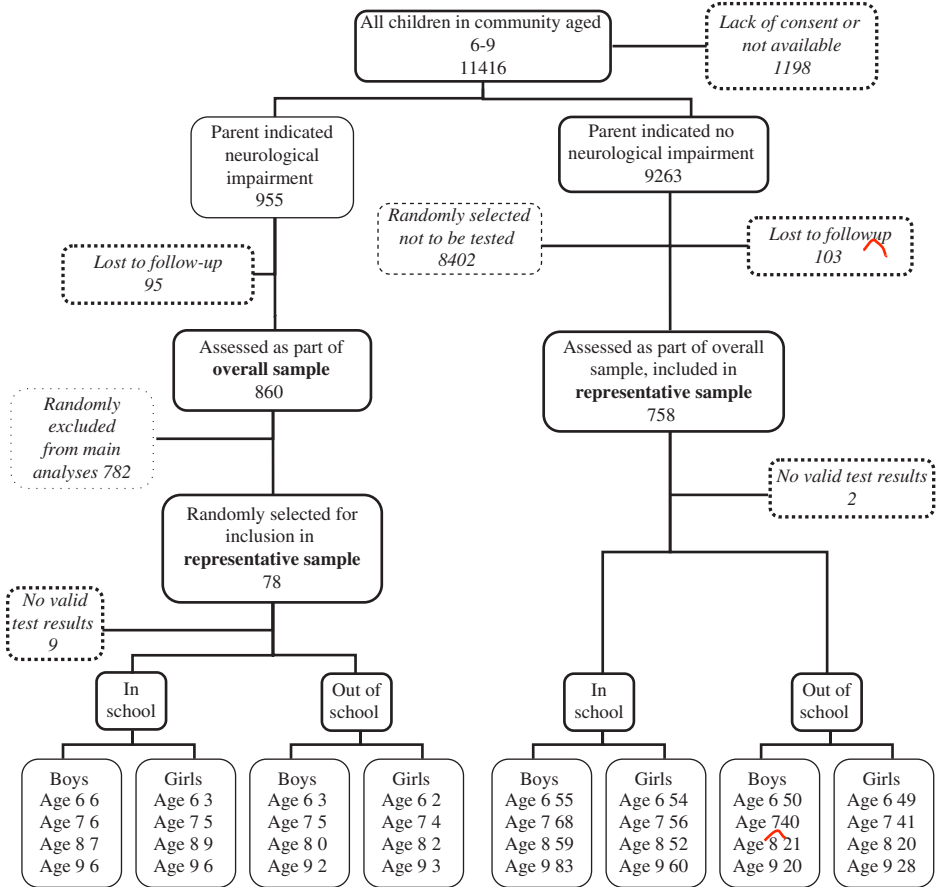
Children aged 6 to 9 years old, recruited both through the census database held at the research unit and through schools, were included in an earlier test development phase of the study. This phase involved the piloting of new test materials and the assessment of test reliability where it had not previously been carried out in the target population. Only children deemed to have no developmental, motor, or sensory problems, on parent or teacher report, were included in the pilot testing. For both the pilot and the main testing phase, children's ages were determined from parental report of date of birth, supplemented by the child's infant health and immunization record card (given to each child at or shortly after birth) where families still had these cards.

Description of Test Battery

Tests were selected to provide an evaluation of general cognitive ability (*g*) and of two subdomains, verbal and nonverbal abilities. Almost all published tests of general cognitive ability, although acknowledging an underlying general intelligence factor (Spearman's *g*), draw a distinction between just two types of cognitive processing, in the main corresponding to a verbal and to a nonverbal factor (Kaufman & Kaufman, 1990; Raven, Court, & Raven, 1992; Wechsler, 1991), although the interpretation and labeling of these factors differ. For each subdomain, an attempt was also made to include tasks that involve a speeded component, as speed of processing has been hypothesized to be closely related to *g* (Deary & Crawford, 1998).

Tests were drawn from those previously developed for application in the region (Holding et al., 1999; Holding et al., 2004; Jukes et al., 2002; Nokes et al., 1992) and supplemented by new adaptations. Tests are described in detail below, and maximum possible scores on each test can be seen in Table 2. All testing was carried out individually by a tester from

Figure 1
Flow Chart Showing Breakdown of Samples of Children Who Participated in the Study



the study area who was familiar with testing children, including those who had not attended school and were, therefore, not familiar with a teaching situation. Children were accompanied by their parents during testing. Tests were administered in a fixed order, starting with tasks that did not require a verbal response from the child or prolonged eye contact with the tester. Testing was carried out on the same day, with breaks if necessary. Given the number of children tested, children were tested at different times of year.

Verbal Tests

Picture Vocabulary Test (PVT). Tests of vocabulary generally have consistently shown high correlations with full-scale measures of IQ (Jensen, 1980). The PVT requires the child to select the correct picture from a choice of four. This test was originally constructed for use in

Table 2
Distribution of Scores on Each Test

Test		Mean Score Obtained	Standard Deviation of Scores	Minimum Score Obtained	Maximum Score Obtained	Maximum Possible Score
PVT	schooled	37.17	6.29	8	49	50
	unschooled	31.44	7.84	5	46	
Information	schooled	5.35	4.49	0	19	20
	unschooled	2.48	2.55	0	12	
Total PVT and Information	schooled	42.49	9.62	10	66	70
	unschooled	33.89	9.15	5	54	
Digit Span total correct	schooled	7.87	2.01	2	16	24
	unschooled	6.68	1.86	0	14	
Category Fluency total score	schooled	17.64	6.66	1	38	—
	unschooled	13.83	6.29	0	34	
Construction total correct	schooled	11.26	5.23	0	24	24
	unschooled	6.15	4.08	0	24	
MFFT mean reaction time for correct items	schooled	4.52	2.32	.85	17.42	—
	unschooled	3.52	1.76	.79	10.99	
MFFT total number of errors	schooled	13.02	4.32	4	27	36
	unschooled	16.22	5.27	1	28	
Build a Man total score	schooled	8.26	3.73	0	20	20
	unschooled	5.25	3.48	0	18	

Note: PVT = Picture Vocabulary Test; MFFT = Matching Familiar Figures Test.

the study area by Holding et al. (2004). Further modifications were incorporated to extend upward the age range to which the test could be applied. The test has four sets of 12 items each. In this study, testing started at Set 2, and if the child failed 8 or more items in Set 2, Set 1 was administered. If the child failed 8 or more items in subsequent sets, testing was discontinued. Children were given credit for Set 1 if they failed fewer than 8 items in Set 2.

Information Test. The Information Test was developed for this study to increase the level of difficulty measured by the verbal knowledge tests. The item pool was developed from a parental survey carried out to determine the areas of knowledge considered by parents to be important for young children to acquire. The areas most commonly suggested by parents were good behavior, good diet, family relations and status, causes of sickness and dangers, cleanliness, games and play activities, religion and prayers, and the natural world. Other topics raised with lower frequency included knowledge of emotions and family relationships and school-based knowledge such as knowing colors, how to spell, and how to count.

The item pool was then supplemented with other items from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and Wechsler Intelligence Scale for Children, Third Edition (WISC-III) information subtests (Wechsler, 1989), the Bagamoyo Daily Living Skills questionnaire (Jukes et al., 2008), and the Orientation section of the Kilifi Creek Behavioural Memory Battery as used in the study described by Carter (2003). Questions were translated into Kigirima where necessary and back-translated for verification. The

resulting questions were then validated by groups of parents and teachers, the groups being asked to rank each question as being of major, minor, or no importance for children to be able to answer. The questions of major importance were further classified into whether a younger or an older child would be likely to know the answer. Only questions achieving 100% agreement both in designation as of major importance and in age ranking were incorporated into the item pool.

Two groups of children, aged 5 to 6 years or 9 to 10 years, were asked the final 40 questions (20 from each age group). Twenty questions that showed sensitivity to age were retained. These were administered to an additional 75 children. The 10 questions that achieved the highest correlations with age and with the total test score were retained in the final question schedule. The final questions are listed in the appendix. All items were administered to all children and the number of items correct was scored; for some items, a partial answer gained half a point. Test-retest reliability was evaluated with 16 children aged 9 and 10.

Digit Span. Digit Span, a measure of phonological working memory, has been found to correlate highly with verbal IQ (Gathercole & Baddeley, 1993). In adapting this test for languages other than English, word length should be taken into account, as digit span is related to the number of syllables, and duration of the pronunciation of words, as well as the number of words to be remembered (Ellis & Hennelly, 1980). The version used here was modeled on the procedures described in the K-ABC (Kaufman & Kaufman, 1983), incorporating an extension of the demonstration items and instructions that emphasized the importance of the order of the numbers. If a child had no correct items on a section (three items), the test was discontinued; the total number of items correct was scored.

Category Fluency. Category Fluency, or semantic fluency, measures speed and efficiency of retrieval from long-term verbal memory. Three categories, “things you can eat,” “girls’ names,” and “animals,” were selected from a list of common words developed through a survey of 190 local schoolchildren. For all ages, and both schooling groups, foods and girls’ names produced the longest lists, although boys produced shorter lists in both categories. This difference was larger for girls’ names. Alternative strategies were piloted, first using “names of people.” This elicited words describing relationships, such as *mother*, *father*, and *auntie*, creating a ceiling on potential responses. Instead, we asked girls to list girls’ names, and boys to list boys’ names. Although girls continued to produce longer lists, the difference was reduced. Therefore, this method was used.

Nonverbal Tests

Construction. The Construction task is based on an idea suggested in Rutter, Graham, and Yule (1970) and is designed to measure perceptual reasoning. The task requires children to assemble wooden sticks to match a model shape to which they are exposed. Piloting revealed that children could not copy a drawing of the shape accurately, but when presented with the model in the form of sticks pasted onto a card, they were more successful. This test was also originally developed for use in the study area by Holding et al. (2004). The task was lengthened for the older age group included in this study. Responses were given 1 point for shape and 1 point for orientation, and the test was discontinued if a child had no

score on one section (two items) after Section 2, or if a child had two consecutive sections with a score of less than 2 (out of a maximum possible 4).

Build a Man. This test is a parallel task to the Goodenough-Harris Draw-a-Person Test that is hypothesized to assess intellectual maturity (Harris, 1963). Studies in Zambia and the United Kingdom suggest that clay modeling measures an equivalent construct but in materials more familiar to African children (Serpell, 1979). The task has been standardized on a Zambian population (Kathuria & Serpell, 1998) from ages 7 to 12 years, providing age and grade norms. Instructions were translated into Kigiriama, but otherwise, the test was used in the same form. During testing, children were allowed to continue working with the clay material until they appeared to have finished, or until 15 minutes had elapsed, and the scoring procedure outlined in the standardization was used.

Matching Familiar Figures Test (MFFT). Although originally designed to assess impulsivity (Kagan, 1964), this task has also been used by Willatts et al. (2004) and Jacobson, Jacobson, Padgett, Brumitt, and Billings (1992) as a measure of processing time, measuring the time to correct response or the ratio of errors to correct response. The MFFT was included in our test battery to provide a measure of processing speed that was straightforward to administer and did not require specialist equipment that might be unreliable in our setting. As many of the original items were unfamiliar to the children in the study area, changes were made to the pictures used by Willatts et al. The telephone, bear, giraffe, boat, and cowboy were removed and were replaced with a basket, chair, cow, lorry, and school-girl, respectively. The number of errors made in response to each item (up to a maximum of three), as well as the reaction time for the first response, were recorded.

Analysis

Test Reliability

Test-retest reliability. Reliabilities for those tests that had already been used in the study area (PVT, Digit Span, and Construction; Holding et al., 2004) were completed on approximately 20 6-year-old children at an interval of 2 to 3 weeks. For the remaining tests (Category Fluency, Information, MFFT, and Build a Man), test-retest reliabilities were performed at an interval of 1 week (Baddeley et al., 1995) on between 18 and 32 children within a 1-year age group (for each test, the oldest child tested was no more than 1 year older than the youngest child tested). All test-retest reliability testing was carried out by the same tester, at the same time of day, and in the same testing situation.

Internal consistency. For the examination of internal consistency (as well as change in children's abilities by age, schooling, and of the underlying factor structure), data were analyzed from the representative sample (see above for details). Cronbach's alpha is reported separately for children in school and out of school for Information, Construction, MFFT (both errors and time taken), and Build a Man.

For PVT and Digit Span, however, Cronbach's alphas were calculated on a sample of 50 and 45 control children, respectively, as rescoring of the original test administration was necessary to extract scores for individual items. For these tests, the statistic reported is for all children together, both those in and those out of school.

Parallel form reliability. For Category Fluency, parallel form reliability was calculated with the two available forms, separately for children in school and out of school, using the representative sample as with internal consistency.

Interrater reliability. For Build a Man, interrater reliability was also carried out. For this test, the main tester scores the child's model at the end of the testing session, retaining the model; the second rater scored the model at the same time point but independently of the first rater and blind to the first rater's judgments. This procedure was carried out with 16 children aged 9 and 10 years, during piloting.

Standardization of Scores and Determination of Delay

In each age group, ζ scores were initially computed for individual tests for all children in the representative sample (see above), taking each year of age separately. Outliers (those in any age group whose ζ score on an individual test was greater than +3.0 or less than -3.0) were then excluded and ζ scores were recalculated using the data from the remaining children in the representative sample, on a testwise basis. Every child from each group was then given a ζ score for each test based on these values. Hence, all children received a ζ score for all tests in which they had participated, based on the distribution of scores of children who both (a) were in the representative sample and (b) had a score on a specific test that was no more than 3 standard deviations away from the mean for their age.

To determine whether children were moderately or severely impaired on the cognitive test battery, and hence to enable validation against physician assessments of cognitive delay, three different methods were used: first, if their mean ζ score for all tests was below -2 (or -3 for severe impairment; Wechsler, 1991); second, if their mean ζ score on all the verbal tasks they performed was below -2 (-3 for severe impairment); and finally, for those children with a hearing impairment, if the ζ score for the Construction task alone was below -2 (again, -3 for severe impairment). A variety of methods were used to establish cognitive delay more accurately in children with communication or hearing impairments. For further discussion of the number of children in the sample who were found to have cognitive delay, and the implications of these findings, see Mung'ala-Odera et al. (2006).

Test Validity

As no standardized methods for cognitive assessment were available for the study population, test sensitivity was compared to the designation of impaired development made through a physician's assessment (common practice in the region). Clinical assessments by a pediatrician were made for all children in the overall sample (see Figure 1). Assessment comprised a neurological exam that included a developmental history taken from parental report; direct assessments of the ability to comprehend and follow simple, yet nonstandardized, instructions (e.g., "Can you shake hands?"; "Can you walk over there and back again?"); assessment of gross and fine motor abilities and coordination, as well as muscle tone, power, and reflexes; and a record of any dysmorphic features (visual signs of chromosomal abnormalities).

Where the clinical exam revealed some indication of a learning disability (either a chromosomal abnormality compatible with a learning disability, or clinical observation of

inability to respond and interact either verbally or nonverbally that could not be explained solely by a sensory or motor disability), the child was deemed to meet clinical criteria for a learning disability. These children were then compared with those who were deemed to be delayed by the cognitive tests outlined above, and specificity and sensitivity were calculated.

Results

Primary Hypothesis Set: Test Properties

Test Score Distribution

The distribution of raw scores for each test, with separate figures for schooled and unschooled children, is shown in Table 2.

Raw scores for Category Fluency, Digit Span, Construction, and Build a Man demonstrated a normal distribution for both children in school and those out of school. Further analyses were, therefore, carried out using the raw scores. Deviation from the norm was observed for scores on the PVT and the Information tasks, suggesting a ceiling and floor effect, respectively. When the two verbal test scores were combined, as intended in the original development of this battery, the total verbal score showed a normal distribution. Subsequent analyses examining the effects of age and schooling were, therefore, carried out using a combined PVT/Information score, although because the measurement properties of the PVT have already been reported (Holding et al., 2004), measurement properties of the Information test are reported here for the sake of completeness. The distribution of the MFFT time score had a very long positive tail, as is usual with reaction time scores. For subsequent analysis, log transformed scores were used. MFFT errors, however, showed a normal distribution.

Test Reliabilities

Correlations for test–retest, parallel form, and interrater reliabilities ranged from fair to excellent (Kline, 2000) and are shown in Table 1.

Internal consistency reliability (Cronbach's alpha) is reported for Build a Man, MFFT (errors), Construction, Information, PVT, and Digit Span. Alpha levels are near or above .8 with the exception of Digit Span and of MFFT errors and Information questions for out of school children only.

In an attempt to improve the reliability of these last two tests for out of school children, on MFFT (errors), removing 5 of the 12 items only raises the reliability to .40, suggesting that only the reaction time should be scored for children out of school. For the Information questions, removing 3 of the 10 items, again, only raises the reliability to .50. Hence, although reliability for these two scores is relatively low, minor adjustments to the item sets do not raise the reliability appreciably, so further analyses include all items. It is worthwhile to report further data on these tests, however, as the purpose of this study was to examine methods of constructing valid test batteries for this population, which necessarily includes reporting those subtests that are not valid.

Table 3
Reliabilities for Tests in the Battery

Test	Test-Retest Reliability [<i>r</i> * (<i>N</i>)]	Internal Consistency for all Children ^a	Internal Consistency: Schooled Children ^b	Internal Consistency: Unschool- ed Children ^c	Other Reliability [<i>r</i> * (<i>N</i>)]
PVT	.55 (20)	.86 (50)			
Information	.74 (36)	.73	.74	.48	
Digit Span	.70 (26)	.63 (45)			
Category Fluency	.63 (36)				Interform: .52 (826) Schooled: .48 (534) Unschool- ed: .49 (292)
Construction	.73 (20)	.88	.87	.82	
MFFT	.81 errors, .84 RT (26)	.55 errors, .91 RT	.59 errors, .91 RT	.26 errors, .87 RT	
Build a Man	.57 (26)	.82	.78	.82	Interrater: .90 (20)

Note: PVT = Picture Vocabulary Test; MFFT = Matching Familiar Figures Test; RT = reaction time.

a. *N* = 837 where not specified.

b. *N* = 537 where not specified.

c. *N* = 300 where not specified.

**p* < .05 for all reliabilities reported.

Validity With Respect to Clinician Assessment

As discussed above, this is currently the primary means of assessing cognitive delay in use in the study area. Hence, if our battery of tests is appropriate for examining cognitive functioning in this population, it should demonstrate validity against any currently used means of detecting learning impairments.

The specificity of the test battery (using the criteria described in the Method section) in detecting learning impairments, compared with detection of learning impairments by the clinician, was 92% and its negative predictive value was 99%, both of which are very high, with sensitivity fair at 61% and positive predictive value low at 18%.¹ Hence, if physician assessment suggested that a child was impaired or unimpaired, this was very likely to be confirmed by the cognitive battery, but a larger number of children were shown to be impaired on the cognitive battery than were assessed as such by the physician, leading to the low positive predictive value.

Secondary Hypothesis Set: Schooling Effects and Their Interaction With Age

Looking at the representative sample, analyses of variance (ANOVAs) revealed main effects of schooling (enrolled vs. not enrolled) and age on all tests. These are shown in Table 4. Note that MFFT correct reaction time (RT) becomes longer as children get older.

Interactions between age and schooling were found on MFFT (number of errors made and the reaction time for correct answers) and Information/PVT. The interaction for Construction approached significance. These are shown in Figures 1 through 5 as well as in Table 4.

Table 4
Effects of Age, Schooling, and Their Interaction

Test	<i>F</i> Values (partial η^2)		Interaction Between Age and Schooling [<i>F</i> _{3, 762} values, <i>p</i> (partial η^2)]
	Effect of Age (<i>F</i> _{3, 762})	Effect of Schooling (<i>F</i> _{1, 762})	
PVT and Information combined	98.54** (.12)	107.42** (.12)	2.78, <i>p</i> = .040 (.01)
Digit Span	19.00** (.07)	42.40** (.05)	.244, <i>p</i> > .05 (.00)
Category Fluency	30.72** (.04)	32.99** (.04)	.693, <i>p</i> > .05 (.00)
Construction	148.33** (.16)	165.22** (.17)	2.313, <i>p</i> = .075 (.01)
MFFT correct RT	14.67** (.06)	28.09** (.03)	4.702, <i>p</i> = .003 (.02)
MFFT errors	31.01** (.11)	55.14** (.06)	4.366, <i>p</i> = .005 (.02)
Build a Man	40.31** (.14)	76.16** (.09)	.914, <i>p</i> > .05 (.04)

Note: PVT = Picture Vocabulary Test; MFFT = Matching Familiar Figures Test; RT = reaction time.

***p* < .001.

Post hoc analyses, with Bonferroni correction applied, were then carried out on data from schooled and unschooled children separately. These revealed that the scores of children in school improved on all tests as they grew older (Information/PVT $F_{3,506} = 105.99$, $\eta^2 = .31$; Digit Span $F_{3,506} = 28.99$, $\eta^2 = .08$; Category Fluency $F_{3,506} = 63.57$, $\eta^2 = .19$; Construction $F_{3,506} = 80.57$, $\eta^2 = .24$; MFFT RT $F_{3,506} = 26.94$, $\eta^2 = .12$; MFFT errors $F_{3,506} = 41.35$, $\eta^2 = .19$; Build a Man $F_{3,506} = 26.75$, $\eta^2 = .31$; all *ps* < .001).

For children out of school, a significant improvement in test scores with age was found on all tests except MFFT RT (Information/PVT $F_{3,256} = 21.13$, $\eta^2 = .20$; Digit Span $F_{3,256} = 6.90$, $\eta^2 = .07$; Category Fluency $F_{3,256} = 13.80$, $\eta^2 = .14$; Construction $F_{3,256} = 17.35$, $\eta^2 = .16$; MFFT RT $F_{3,256} = 2.38$, *p* = .14, $\eta^2 = .03$; MFFT errors $F_{3,256} = 5.49$, *p* = .002, $\eta^2 = .06$; Build a Man $F_{3,256} = 18.72$, $\eta^2 = .18$; all *ps* < .001, except where otherwise noted).

The ANOVAs suggest differential rates of improvement with age on MFFT and Information/PVT for children in and out of school, as well as possibly for Construction. Children who were not in school performed more poorly at all ages for all tests. Post hoc comparisons, using Bonferroni adjustments, were further carried out to examine on which tests children out of school perform more poorly at the same age than children who are in their 1st year of schooling. This should help to determine whether there are initial differences between children in and out of school, separate from the effect of schooling. These data are shown in Table 2.

At ages 6 and 7, children in their 1st (nursery) year of school perform better on most tests (excluding MFFT at age 6) than children who are not in school. At ages 8 and 9, however, children who are out of school and those in their 1st (nursery) year of school perform similarly on the majority of tests in the battery, apart from Construction at both ages, and MFFT RT and Information/PVT at age 9. Note that the children in the nursery grade have a mean of only 5.76 months of education (*SD* = 3.28). It appears that either (a) there is a different effect of entering school at age 6 or 7 from the effect of entering school at age 8 or 9, or (b) the children who end up entering school at age 6 or 7 are different from those who remain home at age 6 or 7, whereas there is not a similar difference between the children who remain at home and those who start school at age 8 or 9.

Figure 2
Age, Schooling, and Total Score on the Information/Picture Vocabulary Test (PVT) Tasks Combined

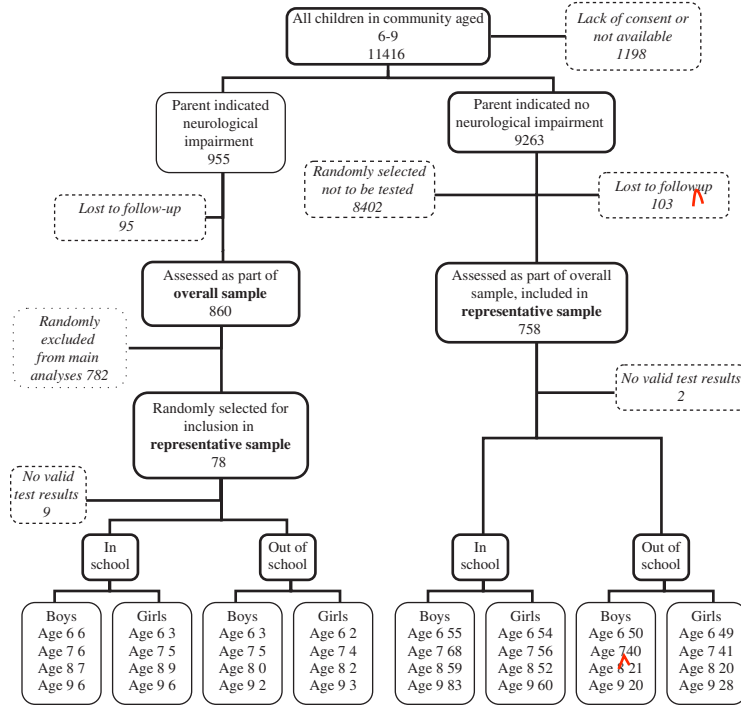
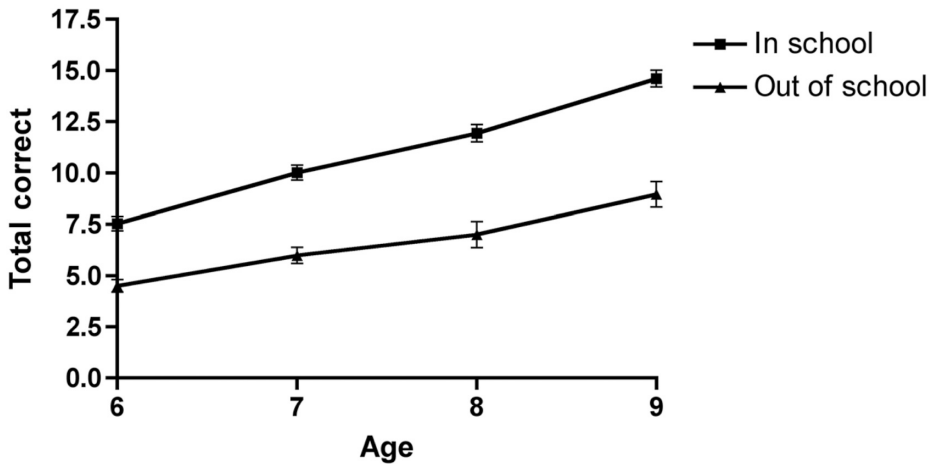


Figure 3
Age, Schooling, and Total Score on the Construction Task



Note: Error bars on all graphs indicate standard errors.

Figure 4
Age, Schooling, and Errors on the Matching Familiar Figures Test

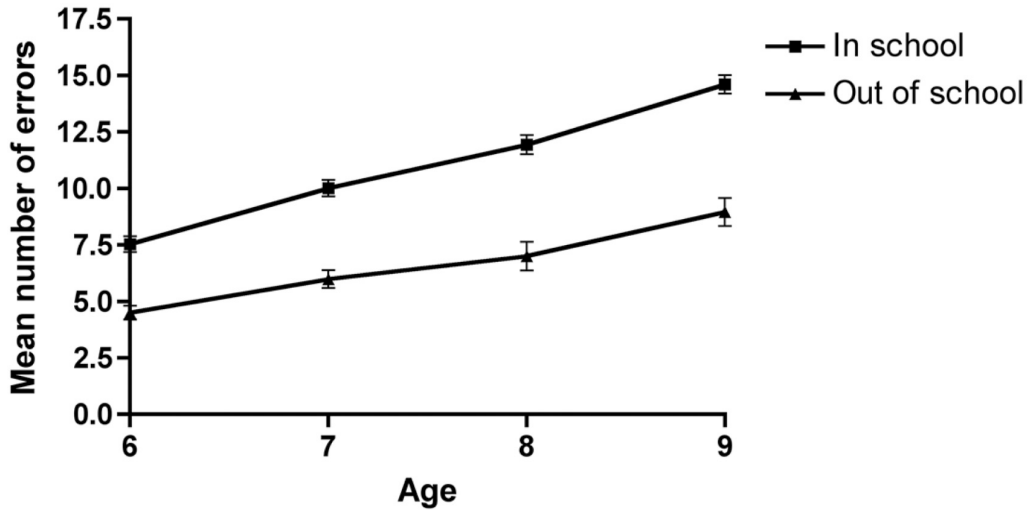
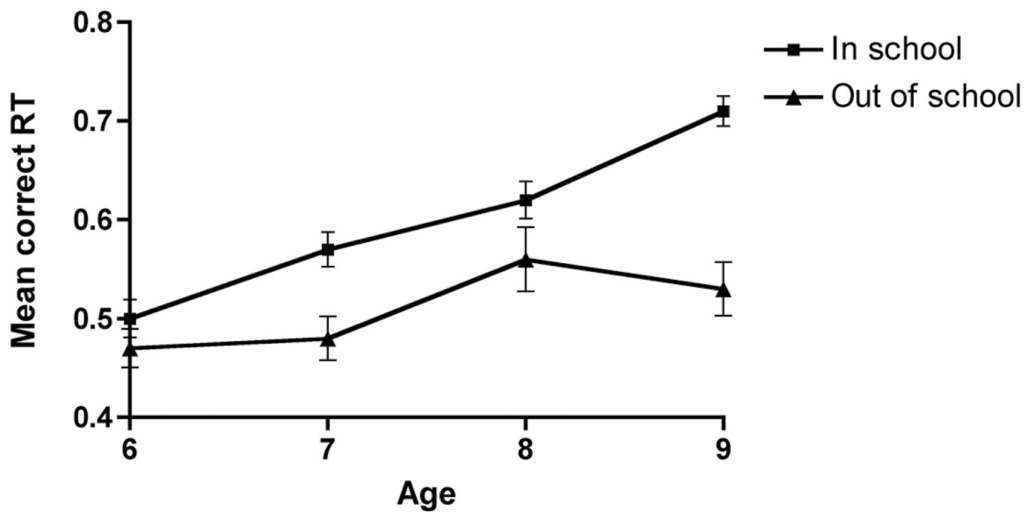


Figure 5
Age, Schooling, and Correct Reaction Time (RT) on the Matching Familiar Figures Test



Discussion

Testing Schooled and Unschooled Children

We sought to develop a battery of tests that would reliably discriminate different levels of performance in our target population, 35% of whom do not attend school despite being of an age to do so. The tests demonstrated acceptable measurement properties, with the majority of the test–retest reliabilities above 0.7 (Kline, 2000). Furthermore, test scores approached the normal distribution for children both in and out of school, showing sensitivity to within-population variance. When comparing test properties between schooling groups, important differences were observed, however.

Although Cronbach's alpha for children in school was acceptable on all tests in the battery, it was low for children out of school on the MFFT errors and the Information test. Because verbal knowledge tasks, either in the form of vocabulary tests or tests similar to the Information test, have been shown to have high correlations with IQ in in-school populations, they are, therefore, commonly used as proxy measures of IQ. Our results suggest that this type of test may be an unreliable and, therefore, inappropriate method for assessing IQ in children who have not had school experience. This finding supports the caution that we have expressed in assuming that the same methodology used for in-school children can be transferred without adaptation to those out of school. Whereas some tests appear to be operating in a similar manner across children both in and out of school, the content or measurement properties of others may preclude the reliable assessment of cognitive function.

Test scores showed the anticipated improvement in performance with age, overall, including reduced errors on the MFFT. For most of the measures in the battery, children both in and out of school do improve significantly as they get older. However, the schooling groups differed in the pattern of improvement observed.

On some tasks, children out of school made significant improvements in their performance. Although children out of school continue to score below those in school, they did not fall further behind throughout the age 6 to 9 range observed. For other tasks (MFFT, both errors and RT, Construction, and the combined Information/PVT), children who do not attend school fail to improve at the same rate as children who are in school. If children out of school fail to improve on tasks at the same rate as children in school, then where a test of cognitive ability is standardized on children in school, those children who do not attend school will appear to have standardized scores that become lower with age. This confirms our original hypothesis that the standard scores on a test battery obtained by children who are not in school will become lower as the children grow older.

It is also relevant to examine whether children in and out of school differ at the point when they enter school, as these may be due to intrinsic between-child differences, parental choices, or both and, hence, may affect schooled and unschooled children's performance on the test battery for reasons other than exposure to schooling. The official age for nursery enrollment in this area is 5 years, but parents frequently delay enrollment in nursery class (or do not enroll their children at all), so that children older than this can be either not in school or in the nursery class. If children who have only just entered school perform better on this battery, it is possible that there are initial differences between children whose families choose to enroll in school and those they do not. Although we did not test schooled children before they started school, we did test a large number of children in their 1st (nursery) year of formal schooling,

who would not currently be receiving intensive formal literacy instruction but who would have some familiarization with school procedures, culture, and school-based interaction with adults, as well as some informal instruction.

There is partial confirmation for the idea that children differ on enrollment: Whereas children in nursery class who are aged 6 and 7 years perform significantly better on almost all tests in the battery than children of the same age who are not enrolled in school, children aged 8 and 9 in nursery class only perform better than age-matched children who are out of school on the Construction task, and the PVT/Information and the MFFT reaction time were performed better by nursery children aged 9 only. A variety of possible conclusions can be drawn from this. Parents may only choose children who are relatively advanced to enroll in nursery at ages 6 and 7, believing it is better for less advanced children to remain at home longer; teachers likewise may suggest to parents that certain children are either not yet ready for school in any form or not ready to proceed beyond the nursery class. This would explain better test performance by 6- and 7-year-olds who are in nursery class than those who are not.

Alternatively, children may require a certain level of cognitive or social maturity to perform on tests in the battery. Children may be able to gain this maturity by having a small amount (around half a year) of nursery education. If they do not enroll in school, they may be able to gain this maturity by becoming a year or 2 older, chronologically. This maturity may include acquisition of specific test-related knowledge or test-taking skills or the ability to interact with unfamiliar adults in a tester/teacher role. The gain of this maturity would likewise explain equivalent test performance by 8- and 9-year-olds out of school or in nursery, whereas lacking this maturity would explain poorer test performance by 6- and 7-year-olds out of school when compared with those in nursery.

Effects of School Experience on Test Performance

For most of the measures in the battery, older unschooled children perform significantly better than younger unschooled children. Older unschooled children were observed to make fewer errors on the MFFT than younger unschooled children, although this decrease in errors was smaller than the equivalent comparison for schooled children of the same ages. On the MFFT reaction time measure, an increase in reaction time with age was seen in children in school. This may be indicative of more careful inspection of the stimuli, as previously observed on this type of task (Vurpillot, 1968). There was no such improvement in older unschooled children, however. These data suggest that the reduction in error scores observed in children out of school is a consequence of a different strategy from that used by those in school. Reduction in the out-of-school children may depend on improvement in processing capabilities without use of a new strategy, whereas older children in school deliberately use a more accurate (but slower) scanning strategy than younger schooled children.

As discussed above, some suggestions have been made that parents actively choose children to send to school who they think may benefit more from schooling (Jukes et al., 2008) and, likewise, that teachers may provide an ability- or potential-based gateway at the entry point to school, either by suggesting to parents that their child is not ready for school or by requiring children to spend more time in the nursery class. The differences between children in and out of school even where the children in school have only had an average of half a year's semi-formal nursery education (for children aged 6 or 7 years), in fact, point in this direction.

However, if this were the case, the children out of school might be expected to fall further behind their school-going peers on all tests in the battery, and for all tests, there would, therefore, be a significant interaction between schooling and age, which was not found. The lack of an increasing gap in performance on some tests, combined with the improvement in performance observed in the out-of-school children on the majority of tests in the battery, suggests, in fact, that those children who do not attend school have similar potential for learning to those who do attend. Likewise, for children aged 8 and 9 in nursery class and out of school, differences were, by and large, nonsignificant, again suggesting that children who have just been selected for schooling and those who have not been selected for schooling are very similar at these ages.

Anecdotal evidence ("No Swots, Please," 2002) has suggested that some Kenyan families, rather than choosing cognitively more able children to send to school, may in fact choose less intelligent children to send to school. Our data do not support this suggestion, however. Rather, they lend weight to the idea that children attending school maintain their standardized scores on intelligence tests (in other words, improving with age at the same rate as peers) because they are learning appropriate skills and information in school. Children who are not attending school are still learning something, although, in the case of some areas of cognitive skill, not at the same rate as their peers in school, so that their standardized scores decline if these scores are based on performance by schooled children.

The difference in the performance levels between the two school attendance groups indicates that school exposure needs to be taken into account in evaluating both absolute performance levels and the rate of change in test performance that a child can be expected to demonstrate. Calculating test scores standardized on age alone will give the impression that the IQ of children out of school lowers year by year (Ceci, 1991; Ceci & Williams, 1997).

Methods of Assessing Cognitive Impairments

In many developing countries, the only means of assessing learning impairments in children is either clinical assessments by physicians or classroom-based assessments by teachers. Examining one of the validation methods, clinical assessments, it is clear that these, as used in the region, are extremely insensitive. The vast majority of children who were assessed as having a learning impairment by a clinician were also assessed as such on the cognitive battery. However, there are many children whose clinical assessment did not reveal a learning impairment but whose cognitive test performance was in the moderate to severe learning impairment range.

Implications and Conclusions

It is clear that differences between the experiences that children have at home and school can have a significant effect on cognitive test performance, and on change in test performance with age, although this effect depends on the cognitive skill being measured. Our findings are important in a practical sense: It is certainly possible to create standardized psychometric tests of cognitive abilities that can be used with children who have had little or no schooling experience, and these can be the same measures that are useful with children who are in school. However, it is clear from the effect of schooling on each and every

measure that standardizing a test on children who are experienced school-goers will not give a realistic, unbiased measure of ability in children who have not yet been to school. It is also clear that some subtests (notably, Information and the MFFT) would be hard to develop to the point where measurement properties were acceptable in both in- and out-of-school children. On these tests, the properties of individual items were examined, the poorest performing items removed, and reliability measures recalculated, but these were still poor. It is likely that the Information test has poor measurement properties in out-of-school children due to more variable acquisition of this type of knowledge in these children. For the MFFT, differences in application of strategies seem to be the reason for the poor properties of this test in this group.

In the arena of test construction, where a standardized test of cognitive abilities is being designed for a population that includes large numbers of children who do not attend school, the ideal approach would clearly be to standardize the test separately for children in school and those out of school, given the differing age profiles of scores for children in each group. A compromise for such a population could be to include an appropriate proportion of children in each group, based on the proportion in the population in general, but this might have the drawback of overestimating standard scores of in-school children while underestimating scores of out-of-school children.

However, for those assessing the performance of individual children with variable school experience in a Western setting where most children attend school, it is acknowledged that such a restandardization will almost certainly not be practical. If the differences we have found are not taken into account in analyzing test performances of children out of school, or those with limited school experience, misdiagnoses of cognitive impairments are highly likely. From our data (compare, for example, the performance of the two groups on Information/PVT and on Construction in Figures 1 and 2), this effect translates into a performance at the age of 9 years that is similar to that of a child approximately 2 to 3 years younger. On some tests, this is a straightforward delay, which may increase in magnitude with age, whereas on others (such as the MFFT), the pattern of change with age will differ for the two groups.

These data are also interesting theoretically, because they bring added knowledge of the effect of a cultural influence, namely, attending school, on cognitive performance and some hints about how this might arise. Although some differences are seen immediately upon entry to the nursery class between schooled and unschooled 6- and 7-year-olds, there are few significant differences between schooled and unschooled 8- and 9-year-olds in the same groups. This suggests that exposure to school, rather than (or possibly in addition to) preexisting differences between children, alters their performance on the test battery. Likewise, children appear to continue to improve on the battery as they grow chronologically older, even if they are not enrolled in school.

Our study has, hence, not just achieved the primary aim of constructing an appropriate battery of cognitive tests for use with both schooled and unschooled children. The investigation of the relative contributions of formal school-based education and chronological maturation is almost impossible to carry out in a Western setting. The importance of this and further studies of school exposure in populations such as ours is that it enables a detailed investigation of the influence of formal education where years of schooling and years of age do not correspond on a one-to-one basis.

Appendix Information Questions

1. What kind of water is good to drink?
 2. Tell me two kinds of things that have wheels.
 3. Tell me two kinds of money.
 4. What is our president called?
 5. Where does the sun set?
 6. What is your father's sister called?
 7. What should you do to have good health?
 8. Do you know what day today is? What day is it?
 9. What month is it?
 10. What year is it?
-

Note

1. *Specificity* refers to the probability that the test under examination (the battery of cognitive tests) allocates an individual to an unimpaired diagnosis (no learning impairment) given that he or she has already been diagnosed as unimpaired by the reference criterion (in this case, clinician assessment). *Negative predictive value* refers to the probability of a child being unimpaired on the reference criterion, given that he or she has been diagnosed as unimpaired with the test battery. *Sensitivity* refers to the probability of a child being allocated to an impaired diagnosis, given that he or she had already been diagnosed as impaired by the clinician assessment, and *positive predictive value* refers to the probability that a child who has been diagnosed as impaired on the reference criterion also receives a diagnosis of impairment on the test battery.

References

- Alcock, K. J. (1998). *Battery of cognitive and educational achievement tests in use in Tanzania 1997 to 1998*. Oxford: Partnership for Child Development.
- Alcock, K. J., & Bundy, D.A.P. (2001). The impact of infectious disease on cognitive development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 221-253). Mahwah, NJ: Lawrence Erlbaum.
- Alcock, K. J., Jukes, M. C., Ngorosho, D., & Deus, C. (2008). *We don't have language at our house: Disentangling the relationship between phonological awareness and literacy*. Manuscript submitted for publication.
- Alcock, K. J., Nokes, K., Ngowi, F., Musabi, C., McGregor, S., Mbise, A., et al. (2000). The development of reading tests for use in a regularly spelled language. *Applied Psycholinguistics*, 21(4), 525-555.
- Baddeley, A., Gardner, J. M., & Grantham-McGregor, S. (1995). Cross-cultural cognition: Developing tests for developing countries. *Applied Cognitive Psychology*, 9(S), S173-S195.
- Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Bellamy, C., & UNICEF. (2004). *The state of the world's children 2005: Childhood under threat*. New York: UNICEF.
- Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 60(5), 1239-1249.
- Carter, J. A. (2003). *Cognitive and linguistic sequelae of severe malaria in Kenyan children*. Unpublished doctoral thesis, University College, London.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27(5), 703-722.
- Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, 52(10), 1051-1058.
- Cole, M., Gay, J., Glick, J. A., & Sharp, D. W. (1971). *The cultural context of learning and thinking: An exploration in experimental anthropology*. New York: Basic Books.
- Conant, L. L., Fastenau, P. S., Giordani, B., Boivin, M. J., Chounramany, C., Xaisida, S., et al. (2003). Environmental influences on primary memory development: A cross-cultural study of memory span in Lao and American children. *Journal of Clinical and Experimental Neuropsychology*, 25(8), 1102-1116.

- Conant, L. L., Fastenau, P. S., Giordani, B., Boivin, M. J., Opel, B., & Nseyila, D. D. (1999). Modality specificity of memory span tasks among Zairian children: A developmental perspective. *Journal of Clinical and Experimental Neuropsychology*, *21*(3), 375-384.
- Connolly, K. J., & Grantham-McGregor, S. M. (1993). Key issues in generating a psychological-testing protocol. *American Journal of Clinical Nutrition*, *57*(2), S317-S318.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, UK: Multilingual Matters Ltd.
- Deary, I. J., & Crawford, J. R. (1998). A triarchic theory of Jensenism: Persistent, conservative, reductionism. *Intelligence*, *26*(3), 273-282.
- Durkin, M. S., Hasan, Z. M., & Hasan, K. Z. (1995). The Ten Questions Screen for childhood disabilities: Its uses and limitations in Pakistan. *Journal of Epidemiology & Community Health*, *49*(4), 431-436.
- Ellis, N. C., & Hennelly, R. A. (1980). Bilingual word-length effect—Implications for intelligence-testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, *71*(1), 43-51.
- Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*, 259-272.
- Gay, C. L., Armstrong, F. D., Cohen, D., Lai, S., Hardy, M. D., Swales, T. P., et al. (1995). The effects of HIV on cognitive and motor development in children born to HIV-seropositive women with no reported drug use: Birth to 24 months. *Pediatrics*, *96*, 1078-1082.
- Gordon, H. (1923). *Mental and scholastic tests among retarded children, physically defective, canal boat and gipsy children and backward children in ordinary elementary schools. An enquiry into the effects of schooling on the various tests*. London: Board of Education.
- Gustafsson, J.-E. (2001). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *International Education Journal*, *2*(4), 166-186.
- Harris, D. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace & World.
- Holding, P. A., Stevenson, J., Peshu, N., & Marsh, K. (1999). Cognitive sequelae of severe malaria with impaired consciousness. *Transactions of the Royal Society for Tropical Medicine and Hygiene*, *93*(5), 529-534.
- Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., et al. (2004). Assessing cognitive outcomes in a rural African population: Development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society*, *10*(2), 246-260.
- Jacobson, J. L., Jacobson, S. W., Padgett, R. J., Brumitt, G. A., & Billings, R. L. (1992). Effects of prenatal PCB exposure on cognitive processing efficiency and sustained attention. *Developmental Psychology*, *28*(2), 297-306.
- Jensen, A. (1980). *Bias in mental testing*. New York: Free Press.
- Jukes, M. C., Grigorenko, E. L., Alcock, K. J., Sternberg, R. J., & Bundy, D. A. (2008). *Development of adaptive behavior in school and community: Evidence from Gambia and Tanzania*. Manuscript under revision.
- Jukes, M. C., Nokes, C. A., Alcock, K. J., Lambo, J. K., Kihamia, C., Ngorosho, D., et al. (2002). Heavy schistosomiasis associated with poor short-term memory and slower reaction times in Tanzanian schoolchildren. *Tropical Medicine and International Health*, *7*(2), 104-117.
- Kagan, J. (1964). *Matching Familiar Figures Test*. Cambridge: Harvard University Press.
- Kathuria, R., & Serpell, R. (1998). Standardization of the Panga Munthu test: A nonverbal cognitive test developed in Zambia. *Journal of Negro Education*, *67*(3), 228-241.
- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test* (2nd ed.). Circle Pines, MN: AGS Publishing.
- Kenya Institute of Special Education. (1984). *Screening test for children: 6 months–6 years*. Nairobi: Author.
- Kline, P. (2000). *The handbook of psychological testing*. London: Routledge.
- Losen, D. J., Orfield, G., & Civil Rights Project. (2002). *Racial inequity in special education*. Cambridge, MA: Civil Rights Project at Harvard University, Harvard Education Press.
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*(1-2), 45-64.
- Morais, J., Content, A., Cary, L., & Mehler, J. (1989). Syllabic segmentation and literacy. *Language & Cognitive Processes*, *4*(1), 57-67.
- Mulenga, K., Ahonen, T., & Aro, M. (2001). Performance of Zambian children on the NEPSY: A pilot study. *Developmental Neuropsychology*, *20*(1), 375-383.

- Mung'ala-Odera, V., Meehan, R., Njuguna, P., Mturi, N., Alcock, K., Carter, J. A., et al. (2004). Validity and reliability of the "Ten Questions" questionnaire for detecting moderate to severe neurological impairment in children aged 6-9 years in rural Kenya. *Neuroepidemiology*, *23*(1-2), 67-72.
- Mung'ala-Odera, V., Meehan, R., Njuguna, P., Mturi, N., Alcock, K. J., & Newton, C. R. (2006). Prevalence and risk factors of neurological disability and impairment in children living in rural Kenya. *American Journal of Epidemiology*, *35*, 683-688.
- Njuguna, P., Mung'ala-Odera, V., Chong, W. K., Meehan, R., & Newton, C.R.J.C. (in press). Computed tomography of the brain in a community study of neurological impairment in Kenya. *Journal of Child Neurology*. No swots, please, we're Maasai. (2002, March 21). *The Economist*.
- Nokes, C., Grantham-McGregor, S. M., Sawyer, A. W., Cooper, E. S., Robinson, B. A., & Bundy, D.A.P. (1992). Moderate to heavy infections of *Trichuris trichiura* affect cognitive function in Jamaican school children. *Parasitology*, *104*, 539-547.
- Olness, K. (2003). Effects on brain development leading to cognitive impairment: A worldwide epidemic. *Journal of Developmental and Behavioral Pediatrics*, *24*(2), 120-130.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255-1264.
- Pollitt, E., Saco-Pollitt, C., Jahari, A., Husaini, M. A., & Huang, J. (2000). Effects of an energy and micronutrient supplement on mental development and behavior under natural conditions in undernourished children in Indonesia. *European Journal of Clinical Nutrition*, *54*(Suppl. 2), S80-S90.
- Raven, J. C., Court, J. H., & Raven, J. (1992). *Manual for Raven's progressive matrices and Mill Hill vocabulary scales*. Oxford: Psychologists Press.
- Rutter, M., Graham, P., & Yule, W. (1970). *A neuropsychiatric study of childhood*. London: Heineman.
- Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology*, *70*(3), 365-380.
- Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology*, *6*(4), 632-650.
- Waterman, A. H., Blades, M., & Spencer, C. (2000). Do children try to answer nonsensical questions? *British Journal of Developmental Psychology*, *18*, 211-225.
- Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence-Revised* (2nd ed.). San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- Willatts, P., Forsyth, S., Agostini, C., Bissenden, J., Casaer, P., & Boehm, G. (2004). *Effects of long-chain polyunsaturated fatty acid supplementation in infant formula on cognitive function in later childhood*. Paper presented at the World Association for Infant Mental Health, Melbourne, Australia.

K. J. Alcock, DPhil, is a lecturer in psychology at the University of Lancaster. Her research areas include the relationships between health and poverty and cognitive and language development, as well as interrelationships between cognitive, motor, and language development.

P. A. Holding, PhD, is a pediatric psychologist with a position as lecturer in clinical psychology at Nairobi University and is also a research associate jointly based at the African Mental Health Foundation in Mombasa, Kenya, and Case Western Reserve University. She currently works on projects investigating the effect on child development of early exposure to HIV and other infections, and on the psychosocial well-being of orphans and other vulnerable children in deprived urban areas.

V. Mung'ala-Odera holds a PhD in epidemiology from the University of Amsterdam and has research interests in the areas of mortality and disability studies.

C. R. J. C. Newton is a professor of tropical neurosciences and pediatrics at the Institute of Child Health, University College London, United Kingdom, and Centre for Geographical Research (Coast), Kenya Medical Research Institute, Kilifi, Kenya. His research interests include infections of the central nervous system in children living in tropical areas and epidemiological studies of neurological impairment and disability in resource poor countries.