

We applaud the efforts of Gladstone *et al*[1] in their adaptation of developmental assessment tools for use in sub-Saharan Africa. We would like however to offer some advice for those attempting to follow in their footsteps with no background in cross-cultural psychology or psychometrics.

It is important when constructing a developmental test battery to have clear theoretical ideas of the domains of interest. When assessing the impact of disease on cognitive development, examining global intellectual ability is unlikely to show group differences [2]. Conversely, when examining children's overall developmental levels, items need to represent a variety of abilities that are known to be related to, for example, global cognitive ability.

Moving to item selection, clear reporting and rigorous implementation of selection criteria will assist those wishing to reproduce methodology or evaluate quality. No assumptions should be made about acceptability, even on the advice of local parents. We advocate a multi-method approach to item selection. This should include consultation with psychologists and fieldworkers, individual parent interviews and child assessments, followed by examining ceiling and floor effects. Children and parents can surprise researchers by their openness to sensitive or difficult items.

Replicability will be enhanced if administration procedures are also adequately described. Observation, parent report, or direct testing of the same skill can produce widely varying results. Confidence in assessments will likewise be increased if the usual statistical examinations of reliability and validity are made; table 1 describes some of these. Finally, returning to theoretical considerations, children's developmental abilities do not represent a unitary domain. Some abilities would be expected to be more closely related than others. Careful factor analyses can reveal whether, for example, parent reporting suffers from the "halo effect", observation or testing is influenced by overall behavioural cooperation, or whether genuine development in theoretically coherent domains is being assessed.

Amina Abubakar, PhD

Centre for Geographic Medicine Research – Coast, Kenya Medical Research Institute (KEMRI), Kenya/ Tilburg University, Netherlands

Katie Alcock, DPhil, CPsychol

Department of Psychology, Lancaster University, UK

Penny Holding, PhD

Centre for Geographic Medicine Research – Coast, Kenya Medical Research Institute (KEMRI), Kenya/ Case Western Reserve University, USA/ African Mental Health Foundation, Kenya

1. Gladstone MJ, Lancaster GA, Jones AP, *et al*. Can Western developmental screening tools be modified for use in a rural Malawian setting? *Arch Dis Child*. 2008; **93**(1):23-9.

2. Hughes D, Bryan J. The Assessment of Cognitive Performance in Children: Considerations for Detecting Nutritional Influences. *Nutr Rev.* 2003; **61(12)**:413-22.

Table 1

Type of statistical consideration	Summary	Usual method of evaluation	Notes
Item variability	Distribution of test scores	Floor, ceiling effects	Eliminate items with skewed distribution or very limited variability
Internal reliability	Intercorrelation of items within a test	Cronbach's alpha; split half reliabilities	Eliminate items with poor relationship to others
Other forms of reliability	Examples are test-retest, inter-tester, and inter-form reliability	Correlation of scores taken at two time points (intra-class correlations), by two assessors, or on two parallel forms	Two forms of a test useful for intervention studies. Reliability heavily influenced by wording of items, method of assessment (observation, direct testing, or parent report), warmup, familiarity of child with setting and assessor
Concurrent validity (including criterion validity)	Relationship between test under construction and simultaneous measures of same concept	Correspondence of new measure with existing standardised measures; with current best practice; correspondence of different methods of measuring same ability	Standardised measures not available in many settings. Validation still possible with clinically delayed children; between observation, parent report, and direct testing of same skill
Convergent validity	Relationship between abilities theorised to be closely related	Correlation between subscales of new measure	Motor scales likely to be related to each other, likewise symbolic play and language abilities
Divergent validity	Lack of relationship between abilities theorised not to be closely related	Lack of correlation or lower correlation between other subscales	Gross motor development shows little relationship to e.g. language development, beyond that due to maturational state