Multicultural Psychoeducational Assessment

Elena L. Grigorenko, PhD, Editor

"[A] very useful resource, not only for psychologists and education professionals, but universally, for all those experts interested in multicultural assessment."

—**Heikki Lyytinen,** PhD
Professor of Developmental Neuropsychology
University of Jyväskylä, Finland

This volume provides a thorough and provocative examination of how different cultures measure intelligence and skill, why they use the tools they use, and how their assessment methods are changing in the globalizing world.

The contributors discuss the extent to which methods of assessment are limited and culture-bound. These methods must be revised and adapted to become relevant to foreign cultures. To this end, the book uses theoretical models and empirical studies to explore the use and validity of standardized tests, language and literacy tests, job interviews, and other methods of assessment across various cultures from both developed and developing countries.

KEY TOPICS INCLUDE:

- National and international standards and guidelines for test development and use
- Limitations of Western assessment tools for populations in the developing world
- The challenges of measuring abilities and competencies in Hispanics/Latinos
- Developing and adapting language and literacy assessments in Arabic-speaking countries
- Assessing competencies in reading and mathematics in Zambian children

ISBN 978-082610101-3

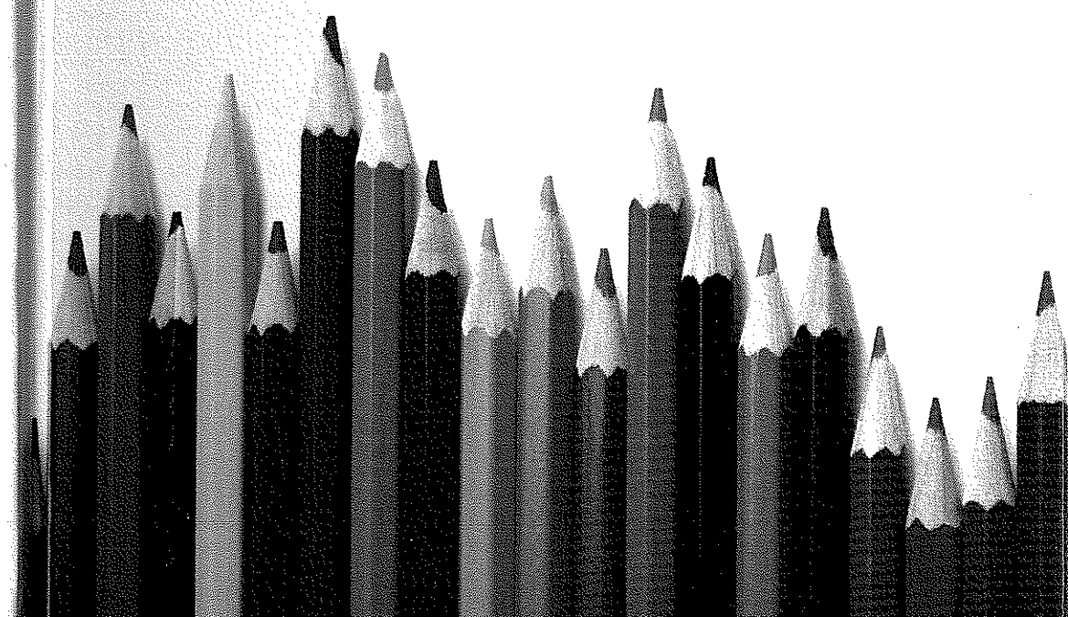
SPRINGER PUBLISHING COMPANY

GRIGORENKO

Multicultural Psychoeducational Assessment

ELENA L. GRIGORENKO
EDITOR

Multicultural Psychoeducational Assessment



6 Assessing Competencies in Reading and Mathematics in Zambian Children

STEVEN E. STEMLER, FLORENCE CHAMVU, HILARY CHART, LINDA JARVIN, JACKIE JERE, LESLEY HART, BESTERN KAANI, KALIMA KALIMA, JONNA KWIATKOWSKI, AIDAN MAMBWE, SOPHIE KASONDE-N'GANDU, TINA NEWMAN, ROBERT SERPELL, SARA SPARROW, ROBERT J. STERNBERG, AND ELENA L. GRIGORENKO

THE IMPORTANCE OF ASSESSMENT IN THE GLOBAL ERA

The practice of educational testing is so deeply woven into the fabric of modern Western schooling that it would be almost inconceivable for a child from a Western school to imagine what school would be like without testing. Indeed, by the time a typical American public school student graduates from high school, he or she will have taken several hundred tests. Yet, much has been written recently about the potential adverse impact of this increased emphasis on educational and psychological testing. Health care providers have reported that a myopic focus on the results of educational tests may lead students to experience greater levels of psychopathology associated with stress and anxiety (Kadison & DiGeronimo, 2004), and researchers have found that the enormous pressure to do well on these tests has also led to a degradation of moral behavior (e.g., increased cheating) not only on the part of students, but also on the part of teachers and administrators (Nichols & Berliner, 2007). Against this backdrop, an objective observer might reasonably ask what advantages, if any, are there to testing?

One perceived advantage of testing from a social perspective is that tests can provide a powerful and objective tool for decision making regarding the allocation of scarce resources. In many European nations, for example, tests are used to place individuals onto different educational tracks early in their lives, which some have argued represents an egalitarian approach to the allocation of resources within the public school system. In the United States as well as in many other countries, tests are used to select which students are eligible to pursue further educational opportunities (e.g., university study) at elite institutions where the demand by students to attend far exceeds the realistic supply of resources the school has to educate all such students well. However, despite the egalitarian intentions underlying the use of tests for placement and selection, the political justice of using tests for this purpose has been a subject of intense controversy in both the United States and the United Kingdom (Lemann, 1999). Tests also serve an important credentialing function in Western societies (Labaree, 1997). They certify the knowledge of particular individuals (e.g., medical doctors with university degrees) and help everyday citizens make informed decisions about the credibility of different individuals.

A second advantage to testing is that, when done well, tests can provide important diagnostic and formative feedback to individuals regarding areas of strengths and weaknesses. Indeed, it is in this spirit that the modern IQ testing movement was born. Alfred Binet was originally commissioned by the French government in 1904 to develop ability tests capable of identifying students struggling in school so that those students could be identified for remediation and extra assistance (Birney & Stemler, 2007). Even in the modern era, one of the primary purposes of testing in the school setting is to identify students who may have specific learning disabilities (LDs) that require accommodations in order for the students to reach their full potential. The research related to LD intervention has repeatedly demonstrated that the earlier such students are identified and treated, the more the potential adverse educational impact of the LD can be mitigated (Lyon et al., 2001). At the other end of the conceptual spectrum, tests can also be used to identify areas of great potential within students and are sometimes used to determine placement into gifted/accelerated school programs.

A third advantage of testing is that it can provide an objective benchmark against which to compare the diverse educational systems of the world. Indeed, an increasing number of countries have begun to participate in large-scale comparative studies of student achievement, such as the Trends in International Mathematics and Science Study (TIMSS) (Martin, Mullis, & Foy, 2008; Mullis, Martin, & Foy, 2008) and the Programme for International Student Assessment (PISA) studies (Organisation for Economic Co-operation and Development [OECD], 2000; Programme for International Student Assessment [PISA], 2001). Participating countries benefit because they are better able to gauge the quality of their educational systems relative to static international benchmarks in a variety of content areas. Such studies have also helped to shed light on the factors associated with effective schools nationally and internationally (Stemler, 2001; Teddlie & Reynolds, 2000).

The proliferation of testing throughout the globe should not be taken lightly, however. There are several issues that researchers must confront when thinking about assessment in the era of globalization. For example, consider the fact that the majority of the world's population (approximately 70% of the general population and almost 90% of its children) lives in the developing world, yet the majority of psychological research and assessment is carried out in the developed world, and mostly in the West (especially Europe and North America). Moreover, the growth of human population has been, is now, and in the future will be almost entirely determined by the world's less developed countries (Population Reference Bureau, 2008). Ninety-nine percent (99%) of global natural increase (the difference between birth and death rates) now occurs in the developing regions of Africa, Asia, and Latin America. U.S. Census Bureau (2003) projections indicate that, for the rest of this century, the number of deaths will exceed the number of births in the world's more developed countries, and *all* of the net annual gain in global population will, in effect, come from developing countries. Thus, in the near future, many more children, proportionally as well as in absolute numbers, will be living in the developing world.

Compared with our knowledge of the development of children in the Western world, we know very little about the abilities and competencies of children in the developing world in general (Grigorenko & O'Keefe, 2004), and in Africa in particular (LeVine et al., 1994; Pritchett, 2001). There are many reasons for this asymmetry (Mpofu, Peltzer, Shumba, Serpell, & Mogaji, 2005; Nsamenang, 1997; Serpell, 1999; Sternberg, 2004; Super, 2005), but one is the lack of instruments suitable for evaluating competencies and describing relevant individual differences in children of the developing world. Many Western assessment methods do not travel well across oceans and continents, and the development of new, custom-tailored instruments is a challenging and time-consuming

task for which developing countries often do not have the necessary expertise or resources (Serpell & Haynes, 2004). Correspondingly, we see an urgent need for psychologists in developed countries to collaborate with their colleagues from developing nations to construct tools that will enable everyone to better study and understand children living in the developing world. Here we exemplify one such collaboration.

The main objective of this chapter is to present an example of a collaborative effort that brought together psychologists and special education professionals in the United States and Zambia and to discuss some of the major issues and challenges facing researchers who attempt to develop assessments in a global context.

BACKGROUND

Cultural Context of the Project

Zambia is a large sub-Saharan African nation (approximately 753,000 square kilometers of land) with a population of 11.3 million people, 46.5% of whom are younger than 14 years of age (The World Fact Book, 2006). It is a low-resource developing country with an estimated gross national annual per capita income of U.S. $900 (The World Fact Book, 2006); approximately 73% of the population live below the national poverty line and 87.4% live below the international poverty line (The World Bank, 2008).

The infant mortality rate in Zambia is high (88.3 deaths/1,000 live births) (The World Fact Book, 2006), and even if Zambian children survive through preschool age (child mortality rates drop dramatically for cohorts older than 7 years of age), their developmental trajectories are typically marked by high disease burdens derived from the poor environmental conditions in which they live. These conditions are characterized by high levels of exposure to biological, chemical, and physical hazards in the environment and an often-observed lack of resources essential for human health. Malnutrition, infection, and disease are highly prevalent among the school-age population in Zambia and most of sub-Saharan Africa.

To fully appreciate the conditions in which Zambian children develop, their schooling situation must be considered as well. Currently, a major characteristic of many sub-Saharan nations is their failing educational systems. Primary education, though much more widespread than

in the days before independence (the number of schools went from few to hundreds), is far from universal; the nation's literacy rate is approximately 81%, compared with approximately 99% for the United States (http://www.indexmundi.com/zambia/literacy.html). Moreover, being enrolled in school does not guarantee quality education: Classrooms are overcrowded and there are not enough teachers, textbooks, or instructional materials (Leithwood, 2000). Indeed, there is evidence that Zambian student performance is among the lowest when compared with students in other African countries. The fourth-grade Monitoring Learning Achievement (MLA) assessment (1999) and the sixth-grade Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) assessment (1998) placed Zambian student performance at or near the bottom of participating nations (Kelly & Kanyika, 2000). Only a small percentage of the Zambian children assessed with the MLA-99 assessment met the requirements for the minimum performance level. Only 37.8% were minimally proficient in literacy (the second lowest achievement level; the lowest percentage was Mali at 15.3%), and only 19.9% were minimally proficient in numeracy (the second lowest, with the lowest percentage Niger's 15.3%). An even smaller percentage of Zambian children performed at the desired level (7.3% and 4.4% in literacy and numeracy, respectively). Among the children from five African countries that participated in the SACMEQ study, Zambian children showed the poorest performance, with 28% of boys and 23.1% of girls performing at the minimum level and 5.6% of boys and 4.8% of girls performing at the desired level. In addition, drop-out rates in Zambian primary school are high, driven by the economic demands on children's time (UNICEF, 1999). Finally, Zambia lacks sufficient numbers of secondary schools, which in combination with poor overall achievement, results in low (<40%) acceptance rates to such schools (Serpell, 1993); thus, secondary schools are available for less than half of those children who graduate from primary schools.

The question of why Zambian children exhibit systematically lower achievement than their peers in other African nations has several potential explanations. From the biological perspective, one possibility is that the cognitive abilities of students in Zambia are deficient. As noted above, it is entirely possible that students in this African nation are exhibiting cognitive impairments that may be associated with specific environmental factors. For example, Bellinger and Adams (2001) have shown that exposure to environmental toxins could result in significant cognitive damage to individuals.

It is possible that the depressed student achievement may also be the result of a less than optimal educational system. In order to test these hypotheses, as well as many others, researchers must have at their disposal some objective measures of student abilities and competencies. In the absence of high quality, psychometrically sound instruments to measure abilities and competencies, we will be left to speculate on the causes of depressed achievement. However, armed with validated measures of abilities and competencies, we can begin to unravel the mystery and systematically examine issues such as the prevalence of learning disabilities (at least as defined using traditional Western criteria). There is, therefore, a significant need to develop psychometrically sound instruments suitable for the Zambian environment.

Assessment From the Ground Up: What to Measure and Why

Given good reasons for wanting to assess abilities and competencies (e.g., both for the benefit of individuals and for the benefit of schools as institutions), where does one begin within the context of a culture where formal assessment is still relatively unusual? One good place to begin is with theory. The theory guiding our assumptions related to the development of assessment measures globally is Sternberg's (1998) theory of abilities as forms of developing expertise. According to this theory, everyone is born with some raw abilities that, through training and development (typically in the form of schooling), can be turned into forms of competence. Through deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993), the competence can then be developed into expertise. To make use of this model, one must first be able to identify (through assessment) different abilities possessed by diverse individuals. In the United States and other Western countries, the identification of abilities is traditionally approached via IQ and related testing. Western IQ tests typically attempt to measure cognitive processes related to memory, reasoning, and adaptive behavior. By contrast, the identification of developing and developed competencies usually is carried out via tests of achievement in specific content domains (e.g., mathematics, reading, and writing).

The assessment of both abilities and competencies is particularly important for the identification of students with learning disabilities. Although there are many models for identifying students with specific and nonspecific learning disabilities, one classic model that has been used in

the West is the identification of an ability-achievement discrepancy (Individuals With Disabilities Education Act, 1997, 1999). Thus, students who have high ability but low achievement may be said to have certain types of learning disabilities (LDs). By contrast, students who have high achievement but low ability may be said to have developed coping mechanisms that they use to mask certain cognitive deficits that could potentially be remediated by enhanced education. In short, in order to help facilitate the development of expertise, we need to be able to assess both abilities and achievement (i.e., competencies). Studies in the United States also point to the fact that early intervention is critical for students with LDs (Lyon et al., 2001). Thus, it is critical to develop measures that can be administered early in a student's educational career.

Models of Assessment in the Era of Globalization

How one approaches assessment depends largely on the assumptions one makes about the nature of abilities and competencies cross-culturally. Sternberg (2004) has outlined four different models related to the assessment of abilities in different cultural contexts. In the first model, the assumption is that all human beings share a fundamental set of cognitive abilities and that cognitive ability measures developed in the Western context should be equally valid for any human population. If all individuals possess the same basic cognitive processes (e.g., problem definition, pattern recognition, reasoning, memory), any nonverbal test of these processes (e.g., Raven's Matrices, tests of reasoning) should be equally applicable to and interpretable in any culture.

In the second model, the idea is that the same tests can be used cross-culturally, but the results cannot be interpreted in quite the same light. Indeed, individuals may administer tests that are used to measure higher order thinking skills in the West, but these tests may not adequately capture these same skills in another culture because the composition of the ability itself may be fundamentally different. For example, research by Cole, Gay, Glick, and Sharp (1971) showed that the Kpelle tripe in Liberia, Africa, had a completely different approach to solving the WISC similarities test than that taken by typical Western children. Specifically, when asked to sort common items into appropriate categories, the Kpelle tended to take a functional approach to sorting (e.g., putting a knife and a potato together because a knife is used to cut a potato) rather than taxonomical categories (e.g., putting a knife and spoon together because they represent the larger category of "silverware"). It was not that the

Kpelle were unaware of the possibility of taxonomical categorization, but rather that they saw functional sorting as more "intelligent." Indeed, when asked how a fool would sort the items, the Kpelle gave taxonomical solutions. Thus, under the second model, although the same cognitive test could be used for assessment, the results must be interpreted as "correct" differentially depending on cultural modes of thought.

A third model suggests that all humans share the same fundamental cognitive processes, such as the ability to define problems, brainstorm solutions, and monitor implementation; however, the way that we assess these abilities must be culturally sensitive and emic in their development. In other words, all human beings face problems and must come up with solutions to those problems, but the nature of the problems faced in different cultural contexts may be radically different.

Finally, a fourth perspective is that the fundamental structure of human ability is different in different cultural contexts. Furthermore, we cannot even use the same methods for assessing these abilities because the assessment instruments themselves must be emic in their development. Thus, one can only truly understand abilities and competencies within a single cultural context and any attempt at cultural comparisons of cognitive processes is considered futile.

The project that we are using as an example in this chapter involved developing and adapting Western-based measures of abilities and competencies within the context of the African country of Zambia. Specifically, we administered three types of assessment: (1) traditional Western-based measures of ability, (2) African measures of ability and competency, and (3) Western-based measures of competency developed for the Zambian context.

METHODS

Testing Cognitive Ability in Zambia

To assess cognitive abilities with a population of Zambian children by means of Western standardized tests, we used nonverbal measures of reasoning and memory. We chose three subtests of the *Universal Nonverbal Intelligence Test* (UNIT) (Bracken & McCallum, 1998) and one subtest of the *Kaufman Assessment Battery for Children* (KABC-II) (Kaufman & Kaufman, 2003). All tests could be presented nonverbally to students and required no verbal responses from them.

The Universal Nonverbal Intelligence Test is an individually administered ability test that measures general intelligence and cognitive abilities of children and adolescents between the ages of 5 and 17 years. We used the Cube Design (CD), Symbolic Memory (SyM), and Spatial Memory (SpM) subtests of the *UNIT* to assess participants' nonverbal reasoning ability (CD) and nonverbal memory (SyM, SpM). The Cube Design subtest consists of 15 items in which students construct three-dimensional designs using a set of cubes. The Symbolic Memory subtest consists of 30 items and asks students to re-create a sequence of symbols they have been shown. The symbols represent a girl, boy, woman, man, and baby, in each of two colors. Finally, the Spatial Memory subtest consists of 27 items in all, in which students are shown a series of grids with chips placed in some of the squares of each grid. Each pattern is shown for 5 seconds, and then the student must re-create the pattern of chips in his or her corresponding grid. The Spatial Memory and Symbolic Memory items are shown to students for 5 seconds, and then removed. There is no time limit for responses on these subtests. Instructions were typically delivered nonverbally in each subtest through the use of gestures and demonstration items. Scoring was completed as right/wrong and partial credit was awarded on the Cube Design subtests for completing some but not all parts of the design correctly. We applied the basal and ceiling rules established by the test developers to determine at which item difficulty level to start and end administration.

The Kaufman Assessment Battery for Children-II is an individually administered assessment of intelligence and achievement. We used the Pattern Reasoning subtest of the *KABC-II* to obtain a second measure of participants' nonverbal reasoning. The Pattern Reasoning subtest consists of 36 items in which the child must perceive a pattern in a series, generate and test hypotheses about the rule used to create the pattern, and apply that rule. The student sees a row of images with one image missing and selects an image that can be placed in the missing space to complete the pattern. This task is untimed and, as with the *UNIT* subtests, can be taught through the use of sample items. Scoring was completed as right/wrong. Basal and ceiling rules established by the test developers were used for this administration.

In addition to importing Western measures of ability, we worked with our colleagues in Zambia (particularly Robert Serpell, who has developed the *Panga Munthu* Test of African intelligence) in selecting culturally relevant measures of ability. The *Panga Munthu* Test (Make-A-Person Test) (Kathuria & Serpell, 1999) is a individually administered

ability test developed specifically for children in rural Zambia. The *Panga Munthu* is conceptually similar to the Goodenough-Harris Draw-A-Person (DAP) Test (Harris, 1963), which was developed in the West and was initially intended as a test of native intelligence in the United States (Goodenough, 1926). In the *Panga Munthu* Test, children are presented with wet clay and are asked to make a person. The produced clay figures are then scored by two independent raters for accuracy of representation of human physical characteristics. Wet clay is considered to be more familiar to rural Zambian children than are paper and pencil. Kathuria and Serpell (1999) cite an earlier cross-cultural study that found that Zambian children's performance on a pattern reproduction task was superior to that of English children in the medium of wire, inferior to that of English children in the medium of pencil and paper, and equal in the medium of clay. The *Panga Munthu* Test was untimed and instructions were delivered to children in English or Nyanja. Participants are instructed to "make the best model of a person you can" from a lump of modeling clay, without any example. Scoring for each physical characteristic criterion was completed as right/wrong (Kathuria & Serpell, 1999). Administration time for the *Panga Munthu* was approximately 10 to 20 minutes per student.

Testing Student Achievement in Zambia

The country of Zambia has one nationally standardized test available for Grade 5 students, the *Grade 5 National Assessment* (*NAG5*). The *NAG5* was developed by the Zambia Examinations Council (ZEC) to assess Grade 5 students' knowledge of the school curriculum in different subject areas. The number of subject areas assessed is increasing to include additional local languages, but we used two assessments that have been in use for several years. The *NAG5* in English consists of 30 multiple-choice items in which the student is asked to read three brief paragraphs and then answer 10 questions about each paragraph. A student's score is the total number of items (out of 30) answered correctly. The *NAG5* in Mathematics consists of 40 items assessing students' computation and problem-solving skills, and a student's total score is the total number of items (out of 40) answered correctly. Although the assessment was designed for Zambian students in Grade 5, children in Zambia have been underperforming on this assessment (Kelly & Kanyika, 2000), that is, less than 50% of all Grade 5 students score at or above 50% correct answers. The *NAG5* assessment items are not released by the ZEC, so

we cannot provide specific examples to illustrate the pool of items on the test.

If our goal is to use assessment as a means to facilitate educational growth, the administration of the *NAG5* test comes too late, particularly for those individuals who may possess learning disabilities. Consequently, we developed the *Zambian Achievement Test (ZAT)*. The *ZAT* is an individually administered test constructed to quantify academic achievement for the purpose of identifying academic difficulties in Zambian children in Grades 1 through 7. Two equivalent versions of the test were constructed to assess two groups of children: those who have received primarily English instruction at school and those who have received their instruction primarily in Nyanja, one of the Zambian local languages. The English version was developed first and reviewed for cultural sensitivity and familiarity of stimuli to Zambian school children (e.g., pictures of animals and types of vehicles commonly found in Zambia were used as items). The test was then translated into Nyanja by a professional translator. An independent back-translation was then conducted to assess the accuracy of the translation. Current national guidelines for Zambian education and Zambian group achievement testing were used to guide estimated item-level difficulty. This procedure for translation and cultural adaptation was inspired by the procedures used in recent large-scale, international comparative achievement studies (Chrostowski & Malak, 2004; Kelly & Malak, 2003).

The *ZAT* was constructed to evaluate performance in four core academic areas: (1) mathematics, (2) reading (letter and word) recognition, (3) pseudoword decoding, and (4) reading comprehension. The construction of the items was guided by Zambian school textbooks, proficiency examinations, and consultations with Zambian educators. The items were designed to cover a large range of difficulty levels, appropriate for children with 0–7 years of formal schooling in Zambia. Figure 6.1 (a–d) presents sample items from the four subtests of the *ZAT*.

The Mathematics subtest (see Figure 6.1a) consists of 60 items that closely mirror the progression of the content of the mathematics curriculum in Zambia. Items are presented in a four-choice response format in which the student must simply point to the correct answer. Initial items instruct students to match a stimulus number with one of four numbers presented. Later items assess knowledge of time, money, calendar skills, numeracy, calculation, and problem-solving skills. The problem-solving questions involving text were read to the students so that math scores would not be biased by reading ability.

Point to the math problem at the top of the page.

Say: **What does 1 plus 5 equal? Find the answer to the problem—*down here.***

Point to the area with answers.

Say: **Point to it.**

Answer: **6**

$$\begin{array}{r} 1 \\ +\,5 \\ \hline \end{array}$$

A	B
C	D

4	6
10	16

Figure 6.1a Sample item from the ZAT-Mathematics subtest.

Point to the picture of the horse at the top of the page.
Say: **This is a picture of a horse.** while pointing to each of the four pictures.

Say: **Mouse, Pig, Fish, Hippopotamus.**

Say: **Which of these pictures begins with the same sound as this picture? Point to the word that begins with the same sound.**

Answer: Bottom Right

A	B
C	D

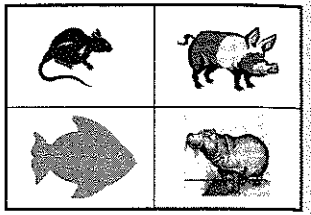


Figure 6.1b Sample item from the ZAT-Reading Recognition subtest.

Say: **I want you to read some words that are not real words. Tell me how they sound.**

Point to the first word: **How does this word sound?**

Point to the following words: **Read these words to me.**

Phonetically:
122. ig(as in p*ig*)
123. op (as in m*op*)
124. et (as in g*et*)
125. ak ('ack' as in bl*ack*)

ig
op
et
ak

ZAT-PW-English ZAT-PW-English

Figure 6.1c Sample item from the ZAT-Pseudoword Decoding subtest.

Student Instructions: **If you are standing, please sit down; if you are sitting, please stand up.**

Examiner Instructions; Say, **Do what this says.** (Use this cue as necessary.)

Correct response: Student should stand, if he/she is sitting, sit if he/she is standing.

If you are standing, please sit down; if you are sitting, please stand up.

Figure 6.1d Sample item from the ZAT-Reading Comprehension subtest.

The Reading Recognition subtest (see Figure 6.1b) consists of 120 items and was constructed with two response types. For the first 60 items, students simply point to one of four possible responses presented in the test booklet. These items begin with letter discrimination items, progress to sound matching and sound discrimination items, and then to letter–sound matching items. The first items are letter matching: For example, students are presented with a page in which a single letter, letter combination, or short two- or three-letter word appears at the top of the page. Four possible matched options presented in isolation or embedded in words are presented below. The examiner points these out to the student as she or he delivers the instructions "Find one like this, down here" or "Find one like this hidden down here." The sound matching items are presented with a single stimulus picture and four possible answer pictures; the sound discrimination items have just four pictures. The pictures are named for the student and the student is asked to point to the one picture that begins with the same sound as the stimulus item (sound matching) (see Figure 6.1b for an example), or to point to the one picture that begins with a different sound than the other pictures (sound discrimination). The letter–sound matching items present a single stimulus picture and four words. The student is asked to point to the one word that begins with the same sound as the stimulus picture. For the next 60 items, the student must read aloud single words presented in the test book. These words, theoretically, get progressively more difficult to decode.

The Pseudoword Decoding subtest (see Figure 6.1c) consists of 38 pseudowords with phonetically regular construction. Initial items consist of simple vowel–consonant combinations (e.g., *ig, ak*) and become progressively more challenging in their length and phonetic construction. The student is given the instructions "I want you to read some words that are not real words. Tell me how they sound." The student must simply read the pseudowords aloud.

The Reading Comprehension subtest (see Figure 6.1d) was constructed as a performance response assessment. Students were asked to read the word, phrase, or sentence presented on each page of the stimulus book and to perform the action directed. For example, the first item is "Jump" and students are instructed, "Do what this says." There are a total of 24 items in this subtest and items become theoretically more challenging through vocabulary and sentence construction. For example, the final item is, "Acknowledge your acquaintance's arrival by gesturing with your hand rather than your voice."

The *ZAT* is presented with colorful pictures for the students and clear instructions for both the student and examiner. Children were not given time limits for responding. Students' responses were recorded by the examiner onto answer sheets. For items with four response choices, the student's response was recorded. For items in which the student had to read a word or pseudoword aloud, the examiner recorded the correctness of the response. All items were eventually scored as right/wrong.

RESULTS

Validating Adapted and Newly Developed Measures in the Zambian Context

Evaluating the validity of adapted and newly developed measures in Zambia is an important and detail-oriented exercise. Given the space limitations of the current chapter, however, we will focus this section on the conceptual questions one should ask to determine whether assessments have traveled well and will report only the major findings related to the validation of our instruments. For readers seeking more detailed technical information, we refer them to a series of Web tables found at http://sstemler.web.wesleyan.edu/stemlerlab/webtables that provide substantial technical detail regarding the results of our validations' studies.

Determining if Western Ability Tests "Work" in the Zambian Context

How does one go about determining whether Western tests travel well to other cultures? One standard psychometric approach is to evaluate the reliability and validity statistics associated with these tests. In particular, it is important to look for evidence of variation in student scores. If all students are scoring at the bottom (or the top) of the exam, then this suggests that perhaps the exam is not a good measure within the given cultural context.

We administered three subtests of the *UNIT* (Symbolic Memory, Cube Design, and Spatial Memory) and one subtest of the *KABC-II* (Pattern Reasoning) to assess the cognitive abilities of Zambian children with conventional ability tests standardized in the West. As these are all nonverbal subtests with nonverbal administration procedures, there are no comparisons of the languages of administration.

Figure 6.2 presents a box and whiskers plot of the distribution of scores on each of these tests. In essence, these boxplots show that the Western tests exhibited excellent variability. Furthermore, the three subtests of the *UNIT* demonstrate acceptable internal-consistency reliability coefficients. The data were analyzed using the Rasch model (Rasch, 1960/1980; Wright & Stone, 1979). The lowest person reliability estimate (conceptually equivalent to, but computationally superior to, Cronbach's alpha—see Bond & Fox, 2001) was .70, found on the Cube Design subtest, while the Symbolic Memory subtest had person reliability estimate of .75 and the Spatial Memory subtest had a person reliability estimate of .81. The *KABC-II* Pattern Reasoning subtest also demonstrated acceptable person reliability (.81).

The item reliability estimates for each of the subscales are quite high, with no scale exhibiting reliability below .95. In addition, the item separation values, an indicator of the range of difficulty levels of items on the test, are well above the minimum requirement of 2.0, with values ranging from 4.45 to 6.69 (Bond & Fox, 2001). Taken together, these findings

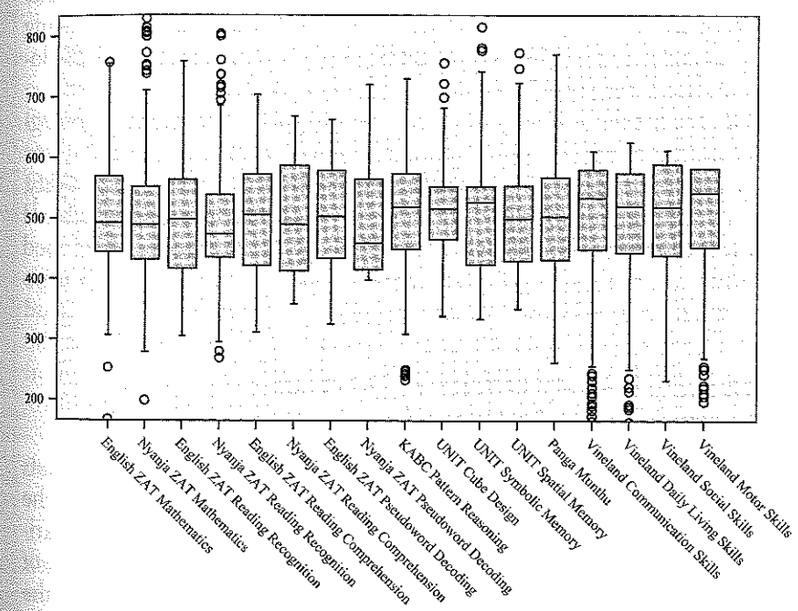


Figure 6.2 Comparison of distributions of each variable.

*Note: Each scale has been standardized with Mean = 500 and *SD* = 100.

demonstrate that the various subtests of the *UNIT* and *KABC-II* possess items that represent a broad range of difficulty levels. Furthermore, the item reliability estimates indicate that we can be quite confident that the placement of the item difficulty values along the continuum of item difficulty would be replicated if these same items were administered to another sample of test-takers from the same population. It is interesting to note that across all four of these subtests, the order of items in terms of difficulty closely approximated the order of difficulty based on U.S. norms (Bracken & McCallum, 1998; Kaufman & Kaufman, 2003).

In order to evaluate the validity of the Western ability tests, we looked at the correlations among Western tests of ability, African tests of ability, and the newly developed Zambian tests of achievement. Table 6.1 presents the criterion-related validity results of the *ZAT*. Scores on each of the four subtests of the *ZAT* (Mathematics, Reading Recognition, Reading Comprehension, and Pseudoword Decoding) were positively correlated with tests of reasoning (the Pattern Reasoning subtest of the *KABC-II* and the Cube Design subtest of the *UNIT*), with tests of memory skills (the Symbolic Memory and Spatial Memory subtests of the *UNIT*), with an African test of intelligence (the *Panga Munthu*), and with teachers' ratings of students on a measure of adaptive behavior (the Daily Living Skills, Socialization Skills, Motor Skills, and Communication Skills domains of the *Vineland-II*).

The results demonstrate that the measures of ability (pattern recognition and memory) are significantly correlated with one another at levels ranging from .24 to .39. Thus, these tests showed convergent validity, but they also demonstrated discriminant validity in that they exhibited slightly lower correlations with measures of achievement (described more fully in the next section) and even lower correlations with measures of adaptive behavior.

Determining if Newly Developed Achievement Tests "Work" in the Zambian Context

In order to determine whether the newly developed achievement tests "work" in the Zambian context, we must also ask questions about the reliability and validity of these tests. The data revealed that the various subscales of the *ZAT* possess items that represent a broad range of difficulty levels, which was a desirable and sought after feature of the test. In addition, the results yielded variability among test-takers (see Figure 6.2) and high levels of internal consistency among the items.

The lowest person reliability estimate was .77, found on the Nyanja version of the Mathematics subtest; however, all other person reliability estimates were more than acceptable (> .80), with the Reading Recognition and Pseudoword decoding scales reaching person reliability estimates in the high .90s. In addition, the item separation values are well above the minimum requirement of 2.0 (Bond & Fox, 2001), with values ranging from 4.13 to 7.12. Furthermore, the item reliability estimates indicate that we can be quite confident that the placement of the item difficulty values along the continuum of item difficulty would be replicated if these same items were administered to another, similar, sample of test-takers.

Overall, the Rasch analyses revealed that 46 out of the 484 items (i.e., 9.5% of the total number of items) on both the English and Nyanja versions of the *ZAT* exhibited infit values greater than expected (i.e., greater than 1.3), and 40 out of 484 items (8%) exhibited infit values lower than expected (i.e., less than 0.70). This general pattern of results is extremely encouraging and suggests that each of the subtests is adequately and reliably measuring a distinct, unidimensional construct, and that each of the *ZAT* subscales—as well as the whole *ZAT*—is a psychometrically sound instrument.

Furthermore, a test–retest study was conducted and the correlations, calculated using the Spearman rank coefficient, were $\rho = .56$ ($p < .001$) for Mathematics, $\rho = .90$ ($p < .001$) for Reading Recognition, $\rho = .81$ ($p < .001$) for Pseudoword Decoding, and $\rho = .82$ ($p < .001$) for Reading Comprehension. These results indicate acceptable levels of stability of the *ZAT* indicators over time.

The four subtests of the *ZAT* show internal validity in that they are significantly correlated with one another, both in English and in Nyanja. Specifically, the Pearson correlation coefficients were all statistically significant at the $p < .001$ level, with Pearson correlation values ranging from .47 to .78, with the median correlation of .62 for the English and .67 for the Nyanja versions.

In addition, for the *ZAT-Nyanja*, the Mathematics subtest shows statistically significant correlations ($p < .001$) with tests of reasoning ability ($r = .20$ to .40), tests of memory ($r = .32$ to .37), the *Panga Munthu* Test ($r = .30$), and assessments of adaptive behavior ($r = .17$ to .44). Each of these correlations, although statistically significant, is moderate in magnitude, suggesting that the *ZAT* measures a construct that shares some variance with the above constructs, but also captures unique variance explaining individual differences in students' achievement.

Table 6.1

THE CRITERION-RELATED VALIDITY RESULTS OF THE *ZAT*: CORRELATION OF THE *ZAT* SUBTESTS WITH TESTS OF REASONING, MEMORY SKILLS, AFRICAN TEST OF INTELLIGENCE, AND ADAPTIVE BEHAVIOR

	SCALE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	ZAT_M_E	.73	0	107	0	94	0	99	0	100	107	94	98	56	71	71	71	63
2	ZAT_M_N	–	.77	0	497	0	162	0	359	491	501	463	440	292	373	372	368	365
3	ZAT_RR_E	.61***	–	.94	0	95	0	100	0	101	108	95	99	56	72	72	72	64
4	ZAT_RR_N	–	.57***	–	.86	0	164	0	359	491	501	463	440	292	372	371	367	364
5	ZAT_RC_E	.52***	–	.78***	–	.72	0	91	0	88	95	84	86	51	64	64	64	56
6	ZAT_RC_N	–	.47***	–	.75***	–	.70	0	159	161	165	135	143	98	134	134	134	132
7	ZAT_PW_E	.47***	–	.73***	–	.62***	–	.83	0	93	100	87	91	54	65	65	65	58
8	ZAT_PW_N	–	.60***	–	.73***	–	.75***	–	.87	354	362	329	322	218	268	267	264	261
9	KABC_PR	.41***	.40***	.36***	.41***	.25*	.33***	.26*	.34***	.53	623	571	553	359	459	458	454	442
10	UNIT_CD	.19	.20***	.20*	.21***	.15	.00	.10	.16**	.24***	.71	580	560	370	464	463	459	447
11	UNIT_SYM	.26*	.32***	.21*	.21***	.24*	.11	.08	.24***	.26***	.31***	.52	540	315	415	414	410	398
12	UNIT_SPM	.06	.37***	.06	.29***	.04	.10	–.06	.25***	.34***	.31***	.39***	.73	327	408	407	403	391
13	Panga_Mun	.04	.30***	.02	.32***	–.03	.24*	.02	.32***	.36***	.12*	.18***	.28***	.88	279	279	279	279
14	V_T_COM	.40***	.38***	.51***	.31***	.30*	.31***	.43***	.36***	.29***	.24***	.26***	.26***	.08	–	489	485	473
15	V_T_DL	.26*	.44***	.40***	.40***	.15	.39***	.40***	.42***	.34***	.18***	.27***	.33***	.28***	.67***	–	485	473
16	V_T_SOC	.17	.21***	.25*	.19***	.15	.21*	.28*	.24***	.16***	.02	.15**	.09	.00	.54***	.59***	–	473
17	V_T_MOT	.18	.17***	.16	.07	.07	.07	.12	.08	.09	.20***	.09	.07	–.07	.42***	.42***	.49***	–

Note. Values in the lower diagonal are Pearson correlations (*r*). Values on the diagonals are internal consistency reliability estimates. Values in upper diagonal are sample sizes. ZAT_M_E = *ZAT* Mathematics Subtest, English; ZAT_M_N = *ZAT* Mathematics Subtest, Nyanja; ZAT_RR_E = *ZAT* Reading Recognition Subtest, English; ZAT_RR_N = *ZAT* Reading Recognition Subtest, Nyanja; ZAT_PW_E = *ZAT* Pseudoword Decoding Subtest, English; ZAT_PW_N = *ZAT* Pseudoword Decoding Subtest, Nyanja; ZAT_RC_E = *ZAT* Reading Comprehension Subtest, English; ZAT_RC_N = *ZAT* Reading Comprehension Subtest, Nyanja; KABC_PR = *KABC-II* Pattern Recognition Subtest; UNIT_SYM = *UNIT* Symbolic Memory Subtest; UNIT_SPM = *UNIT* Spatial Memory Subtest; UNIT_CD = *UNIT* Cube Design Subtest; NAG5 = *Zambian National Achievement Test,* Grade 5; PANG_MUN = *Panga Munthu* (Make-A-Person) Test; V_T_COM = *Vineland-II* Teacher Version, Communication Subscale; V_T_DL = *Vineland-II* Teacher Version, Daily Living Skills Subscale; V_T_MOT = *Vineland-II* Teacher Version, Motor Skills Subscale; V_T_SOC = *Vineland-II* Teacher Version, Social Skills Subscale; V_T_MAL = *Vineland*-II Teacher Version, Maladaptive Behaviors Subscale.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

For Reading Recognition, the correlations of the scores on the *ZAT-Nyanja* were statistically significant with a variety of external criteria, with Pearson correlation values ranging from .21 to .41 ($p < .001$) with tests of reasoning ability, from .21 to .29 ($p < .001$) with tests of memory, at .32 ($p < .001$) with the African test of intelligence, and ranging from .07 (*ns*) to .40 ($p < .001$) with assessments of adaptive behavior.

The Reading Comprehension subtest showed lower correlations with the external criteria than did the Mathematics and Reading Comprehension subtests. Specifically, the Reading Comprehension subtest was significantly correlated with the Pattern Reasoning subtest of the *KABC-II* ($r = .33$, $p < .001$); however, it was not significantly related to other tests of reasoning or memory ($r = .00$ to .11). A moderate but statistically significant correlation was found with the *Panga Munthu* Test of intelligence ($r = .24$, $p < .05$), and correlations ranged from .07 (*ns*, Motor Skills subtest) to .39 ($p < .001$, Daily Living Skills domain) on indicators of adaptive behavior.

Finally, the Pseudoword Decoding subtest showed statistically significant but moderate correlations with a variety of external criteria. Correlations with reasoning tasks ranged from .16 ($p < .01$) to .34 ($p < .001$), from .24 to .25 ($p < .001$) with tests of memory, .32 ($p < .001$) with the *Panga Munthu* Test, and from .08 (*ns*) to .42 ($p < .001$) with tests of adaptive behavior.

The *ZAT-English*, taken by a much smaller sample of test-takers than those who took the *ZAT-Nyanja*, shows similar trends to the *ZAT-Nyanja*. Specifically, the Mathematics subtest of the *ZAT* shows similar correlation values with tests of reasoning ability ($r = .19$, *ns* to .41, $p < .001$), slightly lower correlations with tests of memory ($r = .06$, *ns* to .26, $p < .05$) and the *Panga Munthu* Test ($r = .04$, *ns*), and similar correlation values with tests of adaptive behavior ($r = .17$, *ns* to .40, $p < .001$—Communication domain). Again, each of the correlation values is moderate in magnitude, suggesting that the *ZAT* measures a construct that shares some variance with the above constructs, but also captures unique variance explaining individual differences in students' achievement.

In addition to the data collected during the course of the main study, we also collected data from a smaller sample of 115 students. The purpose of this sample was to evaluate the correlation between the *ZAT* and the *NAG5* in English and Mathematics. We expected to find convergent validity evidence between the relevant subtests of the *ZAT* and the two tests of the *NAG5*. Table 6.2 presents the results of this analysis.

The data reveal that the *NAG5* tests had sufficient internal-consistency reliability (Cronbach's $\alpha = .84$ for Mathematics and Cronbach's

Table 6.2

CORRELATIONS BETWEEN THE *ZAMBIAN NATIONAL ACHIEVEMENT* TEST AND THE *ZAMBIAN ACHIEVEMENT TEST*

	TEST	1	2	3	4	5	6
1	NAG5_M	–	115	92	92	92	90
2	NAG5_E	0.80***	–	91	91	91	89
3	ZAT_M_E	0.71***	0.67***	–	93	93	91
4	ZAT_RR_E	0.80***	0.79***	0.64***	–	93	91
5	ZAT_PW_E	0.63***	0.66***	0.53***	0.78***	–	91
6	ZAT_RC_E	0.69***	0.74***	0.66***	0.73***	0.68***	–

Note. NAG5_M = *Zambian National Achievement Test*, Grade 5 Mathematics; NAG5_E = *Zambian National Achievement Test*, Grade 5 English; ZAT_M_E = ZAT Mathematics Subtest, English; ZAT_RR_E = ZAT Reading Recognition Subtest, English; ZAT_PW_E = ZAT Pseudoword Decoding Subtest, English; ZAT_RC_E = ZAT Reading Comprehension Subtest, English.

*$p < .05$. **$p < .01$. ***$p < .001$.

$\alpha = .89$ for English) and that student achievement on all subtests of the *ZAT* and both tests of the *NAG5* are statistically significantly correlated. In particular, the Pearson correlation between the Mathematics subtest of the *ZAT* and the Mathematics portion of the *NAG5* was $r = .71$ ($p < .001$). In addition, the correlations were statistically significant between the *NAG5* English test and the Reading Recognition subtest of the *ZAT* ($r = .79$, $p < .001$), the Pseudoword Decoding subtest of the *ZAT* ($r = .66$, $p < .001$), and the Reading Comprehension subtest of the *ZAT* ($r = .74$, $p < .001$).

Thus, the ZAT exhibits excellent internal validity statistics (e.g., reliability) as well as promising evidence supporting the external validity of the measure.

DISCUSSION

Challenges in the Assessment of Abilities and Competencies in Zambia

There are a number of challenges associated with the development and adaptation of instruments in a new cultural context. Theses challenges

include both technical issues and cultural issues, including but not limited to (1) ensuring accurate translation of the test items, (2) equating test difficulty when fundamental aspects of different languages vary in their complexity, (3) choosing the relevant basis for comparing students to one another (e.g., age vs. grade level), (4) differences in cultural norms that interact with test behavior, and (5) differences in student motivation that are culturally related.

With regard to the issue of translation, the first question one faces is which language(s) to use for the test. As is characteristic of the majority of African countries, Zambia has 10 languages spoken by more than 1% of the population, 7 of which are officially recognized. Current educational policy assumes that teaching in early primary grades is to be carried out in the mother tongue, with a gradual introduction of English in children's school careers. Although a given mother tongue can be predominant in a given geographical location, it is never universal because Zambia has many intertribal settlers and marriages. In such situations, English is often preferred as the communicational and instructional medium. Correspondingly, all assessments that required a student to use language were developed and administered in one of two languages, English or Nyanja, a predominant language of the Lusaka and Eastern Provinces of Zambia where this study was conducted. However, we must note right away that the diversity of tribal languages suggests that test-takers may possess higher levels of abilities and competencies than they are capable of expressing in English or even in Nyanja. Although this issue is an inherent limitation of all efforts at assessment, the issue is much more salient in the developing world and should temper our interpretation of the test results. Within the context of the current dataset, it seems that the dimension of cognition tapped by the *Panga Munthu* Test is closer to that tapped by tests in the children's indigenous language of everyday discourse than to that tapped by tests in the medium of English, which most of them only use at school.

Once the decision was made to develop the assessment instrument in two different languages (English and Nyanja), we faced the technical challenge of demonstrating that the difficulty of the test items was psychometrically equivalent across different languages. One useful approach to evaluating equivalence involves the construct plots of item difficulties across test forms (see Figure 6.3).

The Reading Recognition plot shows that students found the Nyanja items significantly easier than the English items on this subtest. This is not entirely surprising because Nyanja is a highly transparent language.

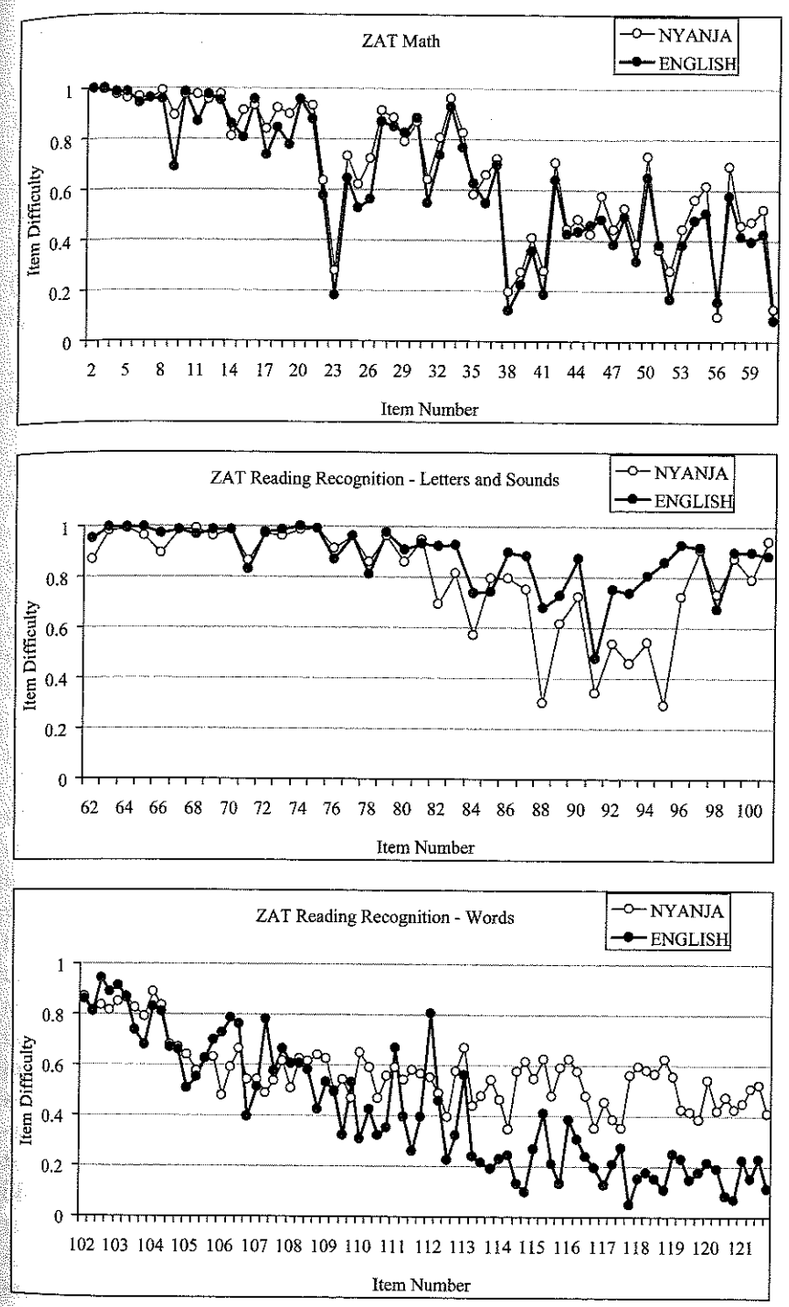
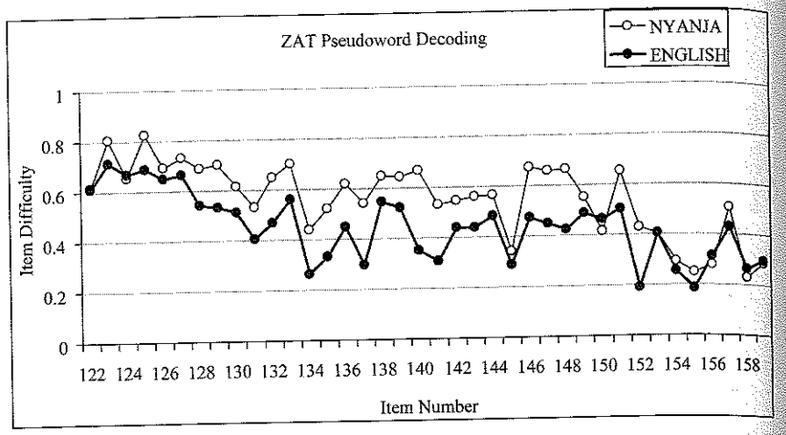


Figure 6.3 Comparison of item difficulty statistics in English v. Nyanja for the ZAT's subscales.



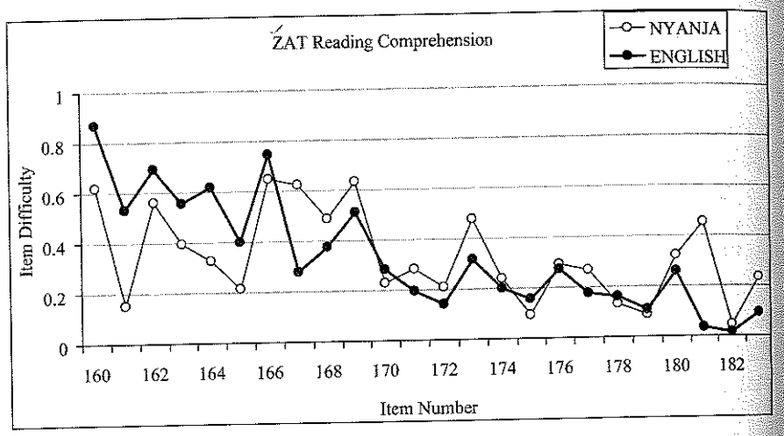


Figure 6.3 *(Continued)*

Once an individual masters the basic structure, there are few exceptions and reading/decoding becomes a fairly easy task. This explains the discrepancy in item difficulty across the tests. The encouraging news, however, is that the pattern of item difficulties remained constant across test versions. Thus, the items that were relatively more difficult in English also tended to be relatively more difficult in Nyanja (compared to easier items in the same language).

A third challenge we faced in the development of new tests in Zambia was whether to norm them on the basis of age or grade level. In

contrast to Westernized countries such as the United States and United Kingdom, where age and grade level are almost perfectly correlated, age does not correlate as highly with grade level in Zambia. For our sample of 206 participants, for whom we had both age- and grade-level information, the Spearman rank correlation between age and grade was $\rho = .52$. Because the *ZAT* was designed as a measure of achievement that should predict school performance, we decided to norm the test on the basis of grade level.

In addition to the technical challenges associated with the development of culturally appropriate tests, there are also many cultural factors to consider that present challenges to the interpretation of test results. For example, cultures differ markedly with regard to norms of appropriate behavior. This was particularly salient to us in the development of the Reading Comprehension subtest of the *ZAT*. Recall that this subtest was designed as a performance assessment. The first item on the exam is "Jump" and students are instructed, "Do what this says." We realized quickly that there were certain words or commands that students in the Zambian context were simply not comfortable performing because they were outside of social norms. Furthermore, particular translation issues also became salient on this test. During the pilot testing, we discovered instances of words that could be interpreted in multiple ways, and these items had to be revised or discarded.

Finally, we noticed a marked difference in terms of the level of motivation students brought to the test. As gratitude for participation in the exams, students were offered snacks and drinks after completing the assessments. Our research team noted a few instances in which a student would attempt to take the battery of tests a second time, maybe in order to receive more snacks, or maybe because they enjoyed the novelty of the situation and the individualized attention. Because testing is relatively novel in this cultural context, most students exhibited higher levels of intrinsic motivation for the task than we have observed in some studies involving Western students, though some students may also have been made more nervous by the novelty of individual assessment.

CONCLUSIONS

Assessment is already woven into the fabric of Western schooling and is quickly gaining in importance across the globe. Yet, developing assessments of abilities and competencies cross-culturally is rarely as

simple as importing currently existing Western-based tests of ability. As van de Vijver (2002) has pointed out, although Western measures may appear to measure a single invariant ability, such as memory or reasoning, differences in stimulus familiarity may call into question the precision of the cross-cultural equivalence of these constructs. Thus, the development and adaptation of tests of abilities and competencies should be solidly grounded in one's theoretical perspective about the nature of abilities and assessment cross-culturally (Sternberg, 2004).

This chapter presents the results of a collaborative study between researchers in the United States and Zambia seeking to develop culturally relevant measures of abilities and competencies. A series of Western-based tests of abilities were adapted and administered in the Zambian context and the results indicate that these measures seemed to travel reasonably well. The tests exhibited sufficient variability, strong reliability, and reasonable validity statistics when measures correlated with each other and with native tests of African intelligence.

In addition, we developed a new culturally based test of student achievement (*ZAT*) for use in assessing the achievement of students in Grades 1–7. Converging evidence indicates that the newly developed *ZAT* allows users to reliably and validly assess competencies of Zambian children in academic-related domains across a wide range of grades, ages, and number of years spent at school.

The development of psychometrically sound measures of abilities and competencies in the country of Zambia carries with it important implications for future research. Most notably, armed with psychometrically validated and culturally sensitive measures of abilities and competencies, future researchers will now be in a position to estimate the etiology of specific and nonspecific learning disabilities using the ability-achievement discrepancy. Furthermore, future researchers will be in a much better position to explore questions related to school effectiveness and student achievement.

Developing measures of abilities and competencies in the era of globalization brings with it several important challenges. Both technical and cultural factors play an important role in the quest to develop culturally sensitive measures; however, the development of psychometrically strong measures brings with it tremendous advances in our capacity to help both individuals and organizations understand how best to improve upon their strengths and remediate areas of weakness.

AUTHOR NOTE

This research was supported by a grant from the U.S. National Institutes of Health (NIH) under R21 TW006764 (PI: Grigorenko)

REFERENCES

Bellinger, D. C., & Adams, H. F. (2001). Environmental pollutant exposures and children's cognitive ability. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 157–188). Mahwah, NJ: Erlbaum.

Birney, D., & Stemler, S. E. (2007). Intelligence Quotient. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 2, pp. 473–476). Thousand Oaks, CA: Sage.

Bond, T., & Fox, C. (2001). *Applying the Rasch model.* Mahwah, NJ: Erlbaum.

Bracken, B. A., & McCallum, R. S. (1998). *The Universal Nonverbal Intelligence Test (UNIT).* Itasca, IL: Riverside Publishing.

Chrostowski, S. J., & Malak, B. (2004). Translation and cultural adaptation of the TIMSS 2003 instruments. In M. O. Martin, I. V. S. Mullis, & S. J. Chrostowski (Eds.), *TIMSS 2003 Technical Report.* Chestnut Hill, MA: Boston College.

Cole, M., Gay, J., Glick, J. A., & Sharp, D. W. (1971). *The cultural context of learning and thinking.* New York: Basic Books.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Goodenough, F. (1926). *Measurement of intelligence by drawings.* New York: World Book.

Grigorenko, E. L., & O'Keefe, P. (2004). What do children do when they cannot go to school? In R. J. Sternberg & E. L. Grigorenko (Eds.), *Culture and competence* (pp. 23–53). Washington, DC: American Psychological Association.

Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity.* New York: Harcourt, Brace & World.

Individuals With Disabilities Education Act. (1997, 1999). 20 U.S.C. 1400 et seq. (Statute). 34 C.F.R. 300 (Regulations).

Kadison, R., & DiGeronimo, T. F. (2004). *College of the overwhelmed: The campus mental health crisis and what to do about it.* San Francisco: Jossey-Bass.

Kathuria, R., & Serpell, R. (1999). Standardization of the *Panga Munthu* test: A nonverbal cognitive test developed in Zambia. *Journal of Negro Education, 67,* 228–241.

Kaufman, A. S., & Kaufman, N. L. (2003). *Kaufman Assessment Battery for Children* (2nd ed.) *(KABC-II).* Circle Pines, MN: American Guidance Service.

Kelly, D. L., & Malak, B. (2003). Translating the PIRLS reading assessment and questionnaires. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 Technical Report.* Chestnut Hill, MA: Boston College.

Kelly, M. J., & Kanyika, J. (2000). *Learning achievement at the middle basic level: Summary report on Zambia's National Assessment Project 1999.* Lusaka, Zambia: Zambian Ministry of Education.

Labaree, D. F. (1997). *How to succeed in school without really learning.* New Haven, CT: Yale University Press.

Leithwood, K. (Ed.). (2000). *Understanding schools as intelligent systems.* Stamford, CT: JAI Press.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy.* New York: Farrar, Straus, & Giroux.

LeVine, R. A., Dixon, S., LeVine, S., Richman, A., Leiderman, P. H., Keefer, C. H., et al. (1994). *Child care and culture: Lessons from Africa.* New York: Cambridge University Press.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., et al. (2001). Rethinking learning disabilities. In C. E. Finn, Jr., R. J. Rotherham, & C. R. Hokanson, Jr. (Eds.), *Rethinking special education for a new century* (pp. 259–287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.

Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international science report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Mpofu, E., Peltzer, K., Shumba, A., Serpell, R., & Mogaji, A. (2005). School psychology in sub-Saharan Africa: Results and implications of a six-country survey. In C. L. Frisby & C. R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 1128–1150). Hoboken, NJ: Wiley.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts American schools.* Cambridge, MA: Harvard Education Press.

Nsamenang, A. B. (1997). Towards an Afrocentric perspective in developmental psychology. *IFE Psychologia: An International Journal, 5,* 127–139.

Organisation for Economic Co-operation and Development (OECD). (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical, and scientific literacy.* Paris: Author.

Population Reference Bureau. (2008). *2008 world population data sheet.* Retrieved October 14, 2008, from http://www.prb.org/Publications/Datasheets/2008/2008wpds.aspx

Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review, 15,* 367–391.

Programme for International Student Assessment (PISA). (2001). *Knowledge and skills for life: First results from PISA 2000.* Paris: OECD.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press.

Serpell, R. (1993). *The significance of schooling: Life-journeys in an African society.* New York: Cambridge University Press.

Serpell, R. (1999). Opportunities and constraints for research on education and human development in Africa: Focus on assessment and special education. *Prospects, XXIX,* 349–363.

Serpell, R., & Haynes, B. P. (2004). The cultural practice of intelligence testing: Problems of international export. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Culture and competence: Contexts of life success* (pp. 163–185). Washington, DC: American Psychological Association.

Stemler, S. E. (2001). Examining school effectiveness at the fourth grade: A hierarchical analysis of the Third International Mathematics and Science Study (TIMSS). *Dissertation Abstracts International, 62*(03A), 919.

Sternberg, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher, 27*(3), 11–20.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*(5), 325–338.

Super, C. M. (2005). The globalization of developmental psychology. In D. B. Pillemer & S. H. White (Eds.), *Developmental psychology and social change: Research, history and policy* (pp. 11–33). New York: Cambridge University Press.

Teddlie, C., & Reynolds, D. (Eds.). (2000). *The international handbook of school effectiveness research.* New York: Falmer Press.

UNICEF. (1999). *Multiple indicator cluster survey: Zambia.* Retrieved January 16, 2005, from http://www.childinfo.org/MICS2/newreports/zambia/zambiatables.pdf

U.S. Census Bureau. (2003). *International population reports.* Retrieved January 4, 2005, from http://www.census.gov/prod/www/abs/popula.html

van de Vijver, F. J. R. (2002). Inductive reasoning in Zambia, Turkey, and the Netherlands establishing cross-cultural equivalence. *Intelligence, 30,* 313–351.

The World Bank. (2008). *Key developmental data and statistics.* Retrieved from http://www.worldbank.org

The World Fact Book. (2006). *Guide to country profiles.* Retrieved January 10, 2006, from https://www.cia.gov/library/publications/the-world-factbook

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA.