

THE PROBLEM OF ESTABLISHING VALIDITY IN  
CROSS-CULTURAL MEASUREMENTS\*

DDH # 120

Marc Irwin, Robert E. Klein, Patricia L. Engle, and Charles Yarborough

INSTITUTO DE NUTRICION DE  
CENTRO AMERICA Y PANAMA  
GUATEMALA, A.

Division of Human Development  
Institute of Nutrition of Central America and Panama (INCAP)  
Guatemala City, Guatemala

BIBLIOTECA

Sara B. Nerlove

State University of New York  
Binghamton, New York 13901

Understanding the causes of behavior requires going beyond the behaviors observed. It requires operating in the realm of constructs, like intelligence and need for achievement, which cannot be measured directly. To be useful, such constructs must be operationalized or linked to behaviors that can be measured. Whether a set of operations can be described as measuring the construct it purports to measure is generally assessed through the process of construct validation.<sup>11</sup> This process involves working backwards from the behaviors measured, through conceptual networks, to predictions about other behaviors which constitute operationalizations of related constructs. Constructs and the operationalizations by which they are measured are together validated to the extent that predictions are successful and the logical structure through which they are generated is persuasive.

Understanding the causes of behavior in a strange culture also requires the use of inference, constructs, and operationalizations. Construct validity of behavioral measurements under these circumstances can have at least three distinct meanings. The three types of validity appropriate to measurement in a culture other than that of the researcher correspond to three approaches to cross-cultural research identified by Berry:<sup>3</sup> the imposed etic, emic, and derived etic. Berry's use of these terms to construct a typology of cross-cultural research techniques contributed a good deal to understanding of the dangers inherent in cross-cultural comparisons. The fact that Berry's typology has not generated a great deal of research can probably be traced to his failure to suggest adequate operational definitions for his terms. In the present paper, an attempt is made to operationally define imposed etic, emic, and derived etic validity. Efforts to establish each of these forms of validity in a cross-cultural investigation in progress in rural Eastern Guatemala are also described.

One of the approaches to cross-cultural research described by Berry is the imposed etic investigation. Imposed etic investigations employ (or impose) imported constructs in attempting to describe behavior in the host culture. Correctly noting the abuses that such an approach has generated in past comparative cross-cultural studies (particularly studies of intellectual ability), Berry dismisses the imposed etic approach as unable to produce more than anecdotal data of the "in culture X, they score Y on test Z" variety.<sup>3</sup> Yet, an

\* This research was supported by Contract No. N01-HD-5-0640 from the National Institute of Child Health and Human Development, National Institutes of Health, Bethesda.

imposed etic approach is not inherently fatal to attempts to understand behavior; for example, in child. rather than cross-cultural psychology, Piaget has employed it extensively to explain developmental phenomena among children, who are not likely to possess the concept of groupings.<sup>13</sup>

Imposed etic constructs may have within the setting where they are employed a species of validity which is independent of the host culture conceptual system. Establishing imposed etic validity involves the same procedure an investigator would use in his own culture; by reference to a conceptual network (probably, but not necessarily, imported from his own culture), the investigator attempts to predict various behaviors implied by that network, on the basis of his behavioral measurements. To the extent that this enterprise is successful, a case is made that the imposed etic constructs measured represent meaningful ways of describing the structure of behavior, that these constructs are related in the way the investigator has supposed, and that they have been adequately operationalized in that particular research setting. An example of successful imposed etic validation was Angelini's<sup>1</sup> demonstration of increased math test scores among Brazilian Ss, following experimentally manipulated failure experiences, a behavioral outcome predicted from research among Ss in the USA.

As noted, the establishing of imposed etic validity is independent of host-culture world view or categories of experience; Angelini's Brazilian Ss, for example, may or may not have possessed concepts equivalent to his notions of failure or need for achievement. However, successfully establishing imposed etic validity does imply that measurement techniques employed are relatively free of the measurement error introduced by unfamiliar stimulus materials, failure to understand instructions, and poor motivation to respond accurately or honestly.

Though ignoring the conceptual system of the host culture, valid imposed etic measurement is useful in explaining the causes of behavior in that culture. It may also yield important information concerning the universality of behavioral causality. In addition, imposed etic measurement is appropriate where the investigator is interested in predicting behavior on a specific criterion measure, such as performance in a Western-type school,<sup>4</sup> exported from his own culture.

Where an investigator wishes to understand the structure of behavior from the point of view of the host culture, a second approach to measurement, the emic approach, is appropriate. Berry described emic investigations as ones in which the investigator progressively modifies his imposed etic entry categories to the point where they accurately reflect indigenous categories of experience. Thus, the researcher's task in such an investigation involves not only transcending empirical observation but also one's conceptual networks as well. In a sense, the researcher in a strange culture must move like a child through the conceptual realm of the host culture, always doubting his own conceptions and heavily dependent upon the guidance of local informants.

The process of emic validation of behavioral measurements is in its general outlines similar to that of construct validation in one's own culture, and to that of imposed etic measurement validation. It consists of making predictions about behaviors on the basis of other measured behaviors, by reference to the investigator's understanding of the host culture's conceptual system. However, establishing emic measurement validity requires that the behaviors predicted be of a particular type: They must be judgments which are assumed to employ the construct dimensions that the investigator is attempting to measure. If in fact the investigator is validly measuring constructs that correspond to indigenous units of organizing the world, he should be able to predict the outcomes of these judgments on the basis of his behavioral measurements. Normally, these

judgments will not be made by the *S* whose behavior is being measured but will be judgments of *S* by peers, family members, etc. In the study to be described in this paper, an attempt is made to establish emic validity of psychometric measurements of child *Ss* by examination of correspondence between these measurements and judgments of the children by their parents and other village adults. We know of no previous attempt to objectively establish emic validity in the cross-cultural literature.

The third class of cross-cultural measurement described by Berry, derived etic measurement, is appropriate where an investigator wishes to make cross-cultural comparisons that take into account the conceptual system of each culture compared. Though derived etic validity can theoretically be established for any number of cultures, this discussion will be confined to the prototypical case of two culture comparisons between the investigator's culture and the host culture.

According to Berry, comparisons are only meaningful for those constructs that constitute a derived etic, or lie in the intersection of the conceptual systems of the two cultures being compared. In fact, it is also meaningful to compare two cultures by employing constructs that have proven to have imposed etic validity when exported from the first culture to the second; the special contribution of derived etic validation is in uncovering cultural commonalities in how the world is viewed.

Establishing derived etic validity is, as its name suggests, a derivative procedure that cannot be undertaken until imposed etic and emic validity have been established. One method of identifying derived etics consists of identifying those aspects of measured constructs that have proven to possess both imposed etic and emic validity. These conceptual units represent an empirical convergence of home and host culture ways of organizing the world.

Establishing emic and imposed etic measurement validity of constructs involves the prediction of behaviors that operationalize related constructs. These related constructs may also possess derived etic validity. Those constructs employed in establishing emic validity originate in the investigator's understanding of the host-culture conceptual system; those constructs employed in establishing imposed etic validity are usually representative of the investigator's home culture conceptual system. The requirement of prior validation of each of these sets of constructs, necessary before derived etics can be sought in their intersection set, is satisfied where they have been involved in successful predictions of behavior. This is the case since construct validation is a reciprocal process in which measurement validity is established simultaneously for predictor and predicted constructs.

Identifying derived etics among constructs that have been predicted by emic measurements and those that have been predicted by imposed etic measurements proceeds by an examination of the interrelationships among operationalizations of the two. Where such interrelationships are found to exist, a derived etic convergence of constructs is empirically established.

The above description of establishing validity in cross-cultural measurement has been drawn in somewhat cleaner strokes than it could have been; for example, the distinction between predictor and predicted measures is in practice a rather arbitrary one. In practice too, the apparently separate activities of establishing imposed etic, emic, and derived etic validity of measurements tend to coalesce into a general exploration of the structure of behavior in the host culture. They do so because the testing of a construct's emic validity implies the exploration of the local conceptual realm, and that exploration is almost



invariably guided, at least at first, by the conceptual system which the researcher brings with him to the host culture. These activities also tend toward simultaneity because of the nature of all human efforts to understand; as Donald Campbell, who has probably contributed more to our understanding of the epistemology of social science and cross-cultural research than any other individual<sup>6-8</sup> has noted, it is virtually impossible to understand a cultural element without simultaneously understanding its larger cultural context.<sup>7</sup>

An important methodological implication of Campbell's statement is the advisability of testing many rather than few hypotheses about the nature of a target construct in the host culture's conceptual system. Both Campbell<sup>6</sup> in his statement about the need to generate many data points in cross-cultural research, and more recently Cole and his coworkers<sup>9,10</sup> have made this argument. By probing many dimensions of a construct, the researcher may be able to construct an interpretable picture of that construct's validity and meaning in the host culture, and its overlap with constructs emic to his own culture. The researcher's dilemma is analogous to that of a blind man with blind friends who wishes to describe an elephant; his best chance of doing so accurately lies in delegating as large a committee of his blind friends as possible to each explore one part of the elephant and to then assemble all of their information into a single description.

Unlike Berry, whose 1969 paper emphasized the epistemological importance of cultural overlaps, Campbell<sup>7</sup> has argued that it is possible to understand cultural differences, when these differences are viewed against a background of shared similarities. This view suggests that nonintersecting portions of the conceptual systems of the researcher's culture and the host culture can provide comparative information as valuable as that provided by the derived etic, or matching portions. The cross-cultural researcher with a pattern of derived etic conceptual similarities as well as conceptual differences characterizing his own and the host culture in hand is like a paleontologist (the former blind man, whose sight has been restored, and who, inspired by his success in describing the elephant, has gone on to study paleontology) attempting to reconstruct the skeleton of an unknown prehistoric beast; faced with sufficient similarity to a known culture or creature, the cross-cultural psychologist or paleontologist can each reconstruct the unique portion, and then venture comparisons of the old and the new.

We have thus far defined three forms of measurement validity that may characterize behavioral measurements made in a strange culture, and discussed rather generally their use in the field. The remainder of this paper will consist of a concrete example of the efforts of one research group to establish the emic, imposed etic, and derived etic validity of a set of psychometric measurements.

The study to be described is the Institute of Nutrition of Central America and Panama's (INCAP) longitudinal study of the relationship between moderate malnutrition and mental development, being carried out in rural eastern Guatemala. The existence of the study was occasioned by a problem of great importance; the need of third world policy makers for data concerning the effects of moderate malnutrition that would permit scarce resources to be allocated in a rational way. Though only preliminary results are yet available,<sup>18</sup> these results (not discussed in the present paper) suggest the first causal link yet established between malnutrition and deficient mental development, pinpoint gestation as a crucial period during which poor nutrition exerts its negative



effect on later mental test performances, and identify children of the lowest socioeconomic level of a population that is almost uniformly poor to begin with, as particularly vulnerable to nutritional insult to mental test performances.

In our attempt to measure the INCAP study dependent variable of mental development, establishing the validity of measurements of intellectual ability in children of various ages has been a continuing concern. Berry has argued that concepts of intelligence "can have no use in a study carried out within the imposed etic-emic-derived etic comparative framework".<sup>3</sup> His argument of why this should be the case is that behaviors assumed intelligent in culture X would probably not be considered so in culture Y, that only fragments of derived etic similarity would remain, and that these "scraps" of the emic wouldn't tell the researcher anything about the global term intelligence. Whether or not Berry is correct can, of course, only be determined empirically. We shall be arguing, through the data we present, that once the necessity of approaching intelligence as if it were a global entity is dispensed with, and the possibility of interpreting cultural differences as well as similarities is acknowledged, that Berry was unduly pessimistic.

#### THE INCAP STUDY

In 1969, a longitudinal study of the relationship between malnutrition and mental development was begun in rural Eastern Guatemala by INCAP. Though a number of studies had previously examined the question, none had been sufficiently well designed to yield more than suggestive evidence of an association between nutritional history and mental test performance, rather than compelling evidence of a causal link between the two. In addition, the INCAP study was unique in focusing on moderate malnutrition, which probably afflicts three-fourths of all children growing up in the world's developing countries,<sup>2</sup> rather than severe malnutrition, which is a relatively rare condition.

The design of the INCAP study is quasiexperimental. Dietary supplement beverages are made voluntarily available to all residents of four villages. In two of these villages, the beverage is a warm gruel, high in proteins and high in calories, and similar to an indigenous drink known as *atole*. In the remaining two villages, the beverage is a cold drink containing no protein and only one-third the calories of the first beverage. It is similar to a local cold drink called *fresco*. Both beverages contain the vitamins and minerals known to be limiting in the local diet. In addition to dietary supplementation, residents of all four villages have been provided with free outpatient medical care since the inception of the study. This care has had a substantial impact on villager health, eradicated some endemic diseases from the communities, and helped to cut infant mortality rates to one third what they were before INCAP's arrival.

The four communities participating in the study are Ladino, or Spanish-speaking and of mixed Spanish and Indian blood. Virtually all community families are engaged in subsistence agriculture. The main crops grown are corn and beans, and these crops are almost entirely consumed within the community in which they are grown. There is little permanent migration to or from the communities. Contact with the world outside the village is primarily limited to trips to nearby markets. Seasonal migration of some community men does occur once a year when these men travel to the coastal zone to harvest cash crops.

The pervasive poverty of these study communities is conveyed by the median family income, approximately \$200 per year. Families live in houses constructed of materials obtained locally, mainly adobe. These houses generally contain two

rooms, one of which serves as sleeping quarters for the whole family. Houses nearly all lack sanitary facilities. Both infectious diseases and moderate malnutrition are endemic. A brief ethnography of life in the villages can be found in Nerlove, *et al.*<sup>20</sup> A full ethnography is provided by Mejía Pivaral.<sup>19</sup>

The research team conducting the INCAP study is interdisciplinary. Epidemiology, pediatrics, developmental and cross-cultural psychology, physical and cultural anthropology, sociology, and mathematical statistics are all represented by one or more professionals and a number of paraprofessionals. In addition to data on nutritional status and mental development, a large number of anthropometric, morbidity, and social-environmental variables are also being measured longitudinally.

The basic sample of children for whom data are presented in the present paper consists of a subsample of 64 children (though, as will be noted, data are missing for three children for one of our measures, and for five others for another measure) drawn from the several hundred sample children residing in two of the four villages involved in the longitudinal study. These children participated in special studies of free behavior observation, in 1971, and of adult rankings of intelligence, in 1973. The special sample was constructed by persons having no knowledge of previous measurements of members of the longitudinal sample, employing as criteria for inclusion that both parents be living and residing in the same household, and age, sex, and family variables; a reasonably successful attempt was made to match same-sex children from one participating village with same-sex children from the other village for these variables. No more than one child in the sample came from the same family.

At the time of the free behavior and ranking of intelligence studies, sample children ranged in age from five to eight. TABLE 1 presents a breakdown of the sample by age, sex, and village.

TABLE 1\*  
SAMPLE DISTRIBUTION BY AGE, SEX, AND VILLAGE†

Age in Years	Boys			Girls		
	Village 1	Village 2	Total	Village 1	Village 2	Total
5	3	3	6	6	5	11
6	9	9	18	6	5	11
7	4	4	8	3	5	8
8	0	0	0	1	1	2
Total	16	16	32	16	16	32

\* After Nerlove, *et al.* (1974), p. 279.

† As of October, 1971.

#### PSYCHOMETRIC MEASUREMENT OF CHILDREN'S INTELLECTUAL ABILITY

Mental testing in the INCAP study begins within a few days after birth, with the administration of the Brazelton Neonatal Scale.<sup>5</sup> At 5-1/2, 7-1/2, 9-1/2, 11-1/2, and 13-1/2 months, an infant cognitive battery designed by Jerome Kagan to measure reactions to stimulus events, their repetitions, and discrepancies from expected stimuli<sup>21</sup> has been administered to infants in the sample.

Each infant has also been administered the Composite Infant Scale, composed of items selected from the Bayley, Cattell, Merrill-Palmer, and Gesell infant scales at 6, 15, and 24 months. Since the problem of cross-cultural validity is of a somewhat different nature, and probably much reduced, with infant tests, these tests will not be discussed in the present paper.

Beginning at 36 months of age, study children are annually administered the Preschool Battery, consisting of 24 tests<sup>21</sup> This battery was designed to tap a variety of cognitive skills, and to represent diverse psychological orientations (e.g., learning, psychometric, Piagetian). The multifactor approach to measurement of intellectual ability employed has been dictated both by INCAP's desire to identify as specifically as possible the effects of malnutrition on mental development, and by the research team's view that intelligence is best conceived of in terms of a number of discrete, though related, abilities.

Tests in the Preschool Battery were adapted to the research setting by a team consisting of American and Guatemalan psychologists, a Guatemalan cultural anthropologist, and Guatemalan testers and cultural informants.<sup>15</sup> Two years of pretesting, during which some tests went through as many as ten revisions, were devoted to developing test materials and instructions that both the intuitions of testers as well as the performances of local (pilot sample) children of various ages suggested were appropriate and meaningful.

Tests are administered in Spanish by Guatemalan female testers. Testing takes place in adobe testing-houses equivalent to village houses, but equipped with one-way viewing mirrors and battery-operated intercoms. Testers are extensively trained and highly competent; nevertheless, they are checked monthly for inter-tester reliability, and all data protocols are routinely doublechecked for irregularities.

Ten of the 24 tests in the Preschool Battery were instituted after the special studies described in this paper were completed. For the purposes of the present paper, data for five tests determined by earlier factor and content analysis to be representative of the 14 tests administered to the special study sample children will be presented and discussed. These five tests of vocabulary, verbal reasoning, discrimination learning, short-term memory, and perceptual analysis are the following:

*Vocabulary.* A picture recognition task similar to the Peabody Picture Vocabulary test. The child is shown a page of pictures and asked to point to the "horse," etc.

*Verbal Reasoning.* A sentence completion task in which *E* reads an analogy, omitting the last word in the sentence, which the child must supply (e.g. "the skin of the pineapple is rough, the skin of the banana is \_\_\_\_\_").

*Discrimination Learning.* A series of four discrimination learning problems. Each employs two pairs of stimuli differing along one dimension. In the first two problems, one value on the dimension (e.g. "big") is reinforced. In the second two problems, the opposite value is reinforced.

*Short-Term Memory.* A digit-span task in which the child repeats strings of digits of increasing length until three strings in a row are failed.

*Perceptual Analysis.* An embedded figures test adapted from one developed by Karp & Konstadt<sup>14</sup> in which the stimuli consist of 12 pictures with a triangle hidden in each. The child's task is to locate each embedded triangle.

Interobserver and test-retest (one-week interval) reliability have been computed for each of these tests. Interobserver reliability is above .99 for each of them. Test-retest reliabilities were .91 for Vocabulary, .87 for Verbal Reasoning,



46 for Discrimination Learning, .65 for Short-Term Memory, and .69 for Perceptual Analysis.

As previously noted, Preschool Battery tests are administered to each child yearly. The scores used in the analyses for the present paper were those for the testing closest to the time of the behavior observation study. Thus, Preschool Battery scores used here represent five-year-old performances for some children, and six-, seven-, or eight-year-old performances for others.

#### ESTABLISHING EMIC VALIDITY OF PSYCHOMETRIC MEASURES: MEASURES OF VILLAGE ADULTS' PERCEPTIONS OF CHILDREN'S INTELLECTUAL ABILITY

##### *Listura Rankings*

The most direct way of determining how local adults view a child is to ask them. We wished first to learn what place, if any, the notion of intelligence has in the conceptual universe of our study villages. Through systematic ethnographic interviews with rural Ladino adults probing indigenous conceptions of intelligence,<sup>17,16</sup> it was determined that the most descriptive indigenous term for intelligence is *listura*, which is most often translated back to English as "smartness." The concept of *listura* appeared to coincide closely with what is commonly referred to as "brightness" or "quickness" in the United States; the behavioral characteristics used by adults in the villages when describing children who are *listo* are "verbal facility," "good memory," "alertness," "independence," and a high level of physical activity.

*Listura* rankings for this study were generated in each village by photographing each child in a standing position against the same neutral background, and having a sample of women judges who were near childbearing age and did not have a close relative among the sample, use the photographs to make a series of random pairwise same-sex comparisons.<sup>20,21</sup> Each child was assigned a *listura* score on the basis of the rankings of all the judges who knew him or her. The score consists of the percentage of times the child was ranked as more *listo* divided by the total number of times ranked. (See Nerlove and Walters<sup>21</sup> for a method of composite rank-ordering based on a least squares solution used on this same body of data and resulting in different scores.)

##### *Chores*

A second measure of how village adults (in this case parents) perceive the intellectual ability of their children consisted of an index of the chores the children were assigned. Life is difficult in our study villages, and with few exceptions, families must struggle to survive. Under these circumstances, children are expected to contribute what labor they can to the family. Assuming that an important determinant of the extent to which parents assign tasks to their 5-8-year-old is their perception of the child's intellectual maturity, measures of children's chore participation have been correlated with psychological test performances and other indices of intellectual ability. The chore measure consists of the number of observations out of a random sample of 20

observations per child<sup>20</sup> in which the child was performing a chore. The average number of observations in which boys were performing chores was 6.78 (SD = 3.68). The average number for girls was also 6.78 (SD = 3.26).

As will be noted in the results section, this measure was highly correlated with test performance and other indices of intellectual ability for girls. It was not related to other measures for boys. However, the number of intellectually demanding chores performed by the child was related to test performances among boys, and this latter chore measure is used for boys. The classification of chores into intellectually demanding or not was done by the Guatemalan cultural anthropologist, Victor Mejía Fivarral, author of the definitive ethnography of the four villages,<sup>19</sup> who has worked in the villages continuously for the last six years. The average number of intellectually demanding chores engaged in by boys was 1.66 (SD = 1.77).

### *School Attendance*

Still another measure of parental judgments of their child's intellectual ability employed was the number of years of schooling completed by the child, as of January 1975. Since children do not normally begin school before eight years of age in the study villages, school attendance took place long after all of the other behaviors measured in the present study; our analyses involving this variable included predictions of events up to three years in the future.

Though each of the study villages has an elementary school, school attendance is sporadic and not universal among village children. Given the important economic contribution that the children increasingly come to make to their families at this time, we have hypothesized that an important determinant of the extent to which parents permit their child to attend school is their perception of the child's intellectual ability; where all resources including energy are in short supply, it would appear to make little sense to send a potentially productive family member to school if that individual were unlikely to benefit from the experience. The average number of years of schooling completed, as of January 1975, was 1.00 (SD = 1.11) for boys, and .88 (SD = 1.01) for girls, and 31 children had not completed a year of school by that date.

## ESTABLISHING IMPOSED ETIC VALIDITY OF PSYCHOMETRIC MEASURES: MEASURES OF INTELLECTUAL ABILITY BASED ON CHILDREN'S BEHAVIOR

### *Self-Managed Sequences*

One of the investigators' (imposed etic) assumptions about intellectually able children everywhere was that they are more likely than less able children to engage, without supervision, in complicated activities. To index the tendency to do so, a variable developed by Nerlove<sup>20</sup> in the observational study described previously is employed in the present analyses. This is the number of self-managed sequences in which each child was engaged out of 20 random observations. For this variable, the activities indexed are ones in which the child was at least to some degree in control of the situation, and thus giving indication of his or her level of cognitive development.

The self-managed sequence measure is an index that includes both work and play activities. Work activities included were those not easily supervised and involving transformations of objects and systematic movement through space, and those performed alone and requiring going outside the community or gathering, transforming, and relocating objects. Play activities included were those involving rule games, interactive role play, joint or systematic construction, or play that involved going outside the community. The average number of self-managed sequences engaged in out of 20 observations was 3.60 for boys ( $SD = 2.46$ ) and 2.13 for girls ( $SD = 1.41$ ). For a more complete description of this variable and observed behavioral examples, the reader is referred to Nerlove *et al.*<sup>20</sup>

### *School Performance*

School performance is commonly associated with the concept of intelligence in Western culture. This association may be explainable by the fact that the first intelligence test, developed by Binet in 1905,<sup>22</sup> was designed to predict school success, and school grades have continued to correlate with IQ test-scores ever since. In the present study, the measure of school performance employed was the ratio (for those 33 children who had completed a year or more of school), of the number of years passed to the number of years attended. Though teacher grades for the years 1972–1974 were not available, records of passing and failing are kept by the Guatemalan Ministry of Education, who made these data available to us. The mean ratio of years of school passed to years of school attended was .86 for boys in the sample ( $SD = .33$ ), and .78 for girls ( $SD = .36$ ).

## RESULTS

### CORRELATIONS BETWEEN PSYCHOMETRIC, ADULTS' PERCEPTIONS, AND CHILDRENS' BEHAVIOR MEASURES OF INTELLECTUAL ABILITY

The following section will present correlations obtained between psychometric measures of intellectual ability, measures of village adult's perceptions of children's intellectual ability, and measures of intellectual ability based on children's behavior. The issue of validity will be taken up in the discussion section to follow, since establishing emic and imposed etic validity involves interpreting the patterns of correlations found between measures, whereas establishing derived etic validity is a kind of third-order operation involving examining the interrelationships of operationalizations of constructs which have had emic and imposed etic validity established. Also, since scores on the Preschool Battery tests have easily communicable meaning only relative to other scores on the same tests, summary information for these scores will not be presented in the present paper. TABLE 2 presents the intercorrelations for the five Preschool Battery tests employed. It will be noted that Vocabulary correlates most strongly with the other tests included, and particularly with scores on the Perceptual Analysis test. Intercorrelations are generally higher for girls than for boys, though the differences must be taken only as suggestive, due to small  $ns$ . However, a consistent finding of the longitudinal study has been that of differential interrelationships by sex, and all results reported in the present paper will be presented separately by sex.



TABLE 2  
INTERCORRELATIONS AMONG PRESCHOOL BATTERY TESTS\*  
FOR BOYS AND GIRLS

	Vocabulary	Verbal reasoning	Discrimination learning	Short-term memory	Perceptual analysis
Vocabulary		.53†	.35‡	.31	.62§
Verbal reasoning	.57§		.08	.35‡	.40‡
Discrimination learning	.47§	.10		.37‡	.15
Short-term memory	.59§	.42‡	.26		.24
Perceptual analysis	.79§	.49§	.29	.50§	

Boys

Girls

\* For testing closest to time of behavior observation study performed in 1971, when sample children ranged in age from 5-8 (See Table 1).

† N = 32 for all correlations, excepting those involving girl's Verbal Reasoning, where n = 31.

‡ P < .05.

§ P < .01.

Correlations between Preschool Battery scores, perceived intellectual competence scores, and measures of intellectual competence based on children's behavior are shown in TABLES 3 and 4. As TABLE 3 indicates, for boys, various perceived competence and children's behavior measures correlated significantly with tests in the Preschool Battery. Of these, listura rankings and years of schooling correlated most highly with Preschool Battery test scores. Listura rankings correlated .60 ( $p < .01$ ) with Vocabulary, and .59 ( $p < .01$ ) with Perceptual Analysis. Number of years in which the child was kept in school, correlated .63 ( $p < .01$ ) with Verbal Reasoning, .45 ( $p < .01$ ) with Vocabulary, and .36 ( $p < .05$ ) with Perceptual Analysis. No measure of intellectual ability correlated consistently well with all of the psychological tests, and significant correlations between measures of intellectual ability and test scores were spread evenly across the five Preschool Battery Tests: in addition to the significant correlations seen between listura and Vocabulary and Perceptual Analysis and between number of years a child was kept in school and Verbal Reasoning, Vocabulary, and Perceptual Analysis, significant correlations were found between number of intellectually demanding chores assigned and Discrimination Learning and Short-Term Memory ( $r_{\text{learning}} = .35$ ,  $p < .05$ ;  $r_{\text{memory}} = .35$ ,  $p < .05$ ) and between self-managed sequences and Discrimination Learning ( $r = .40$ ,  $p < .05$ ) and Perceptual Analysis ( $r = .39$ ,  $p < .05$ ). Of all the nonpsychometric measures of intellectual ability, only school passing ratio failed to correlate significantly with any test. The highest correlation between school passing ratio and Preschool Battery tests was that with Perceptual Analysis ( $r = .25$ , n.s.).

Among the nonpsychometric measures, a fair amount of agreement was observed. Listura scores correlated significantly with years of schooling ( $r = .39$ ,  $p < .05$ ) and self-managed sequences ( $r = .41$ ,  $p < .05$ ). Number of intellectually

TABLE 3

CORRELATIONS BETWEEN PRESCHOOL BATTERY SCORES, MEASURES OF VILLAGE ADULTS' PERCEPTION OF CHILDREN'S INTELLECTUAL ABILITY, AND MEASURES OF INTELLECTUAL ABILITY BASED ON CHILDREN'S BEHAVIOR: BOYS

(n) =	Preschool Battery					Perceived Intelligence			Behavioral Intelligence	
	Vocabulary	Verbal reasoning	Disc. learning	Short-term memory	Perceptual analysis	Listura	Demanding chores	Years schooling	Self-managed sequence	School passing ratio (n = 17)
Perceived intelligence: Listura (29)	.66†	.35	.20	.69	.59†		.18	.39*	.41*	.26
Demanding chores (32)	.12	.12	.35*	.35*	.18			.39*	.64†	.10
Years of schooling (32)	.45†	.63†	.02	.25	.36*				.51†	.33
Behavioral intelligence: Self-managed sequences (30)	.37	.17	.40*	.20	.39*					.34
School passing ratio (17)	.00	-.11	-.08	.13	.25					

\*  $p < .05$

†  $p < .01$

TABLE 4

CORRELATIONS BETWEEN PRESCHOOL BATTERY SCORES, MEASURES OF VILLAGE ADULTS' PERCEPTION OF CHILDREN'S INTELLECTUAL ABILITY, AND MEASURES OF INTELLECTUAL ABILITY BASED ON CHILDREN'S BEHAVIOR: GIRLS

(n) =	Preschool Battery					Perceived Intelligence			Behavioral Intelligence	
	Vocabulary	Verbal reasoning (n = 31)	Disc. learning	Short-term memory	Perceptual Analysis	Listura	Chores (all)	Years Schooling	Self-Managed Sequence	School Passing Ratio (n = 16)
Perceived intelligence: Listura (30)	.75†	.39*	.61†	.38*	.55†		.51*	.53†	.23	.17
Chores (all) (32)	.53†	.40*	.44*	.38*	.33			.57†	.36	.09
Years of Schooling (32)	.56†	.22	.23*	.40*	.37*				.37*	.30
Behavioral intelligence: Self-Managed Sequences (31)	.44*	.19	.26	.13	.26					.18
School Passing Ratio (16)	.52*	.34	-.22	.14	.02					

\*  $p < .05$

†  $p < .01$



demanding chores also correlated significantly with both years of schooling ( $r = .39$ ,  $p < .05$ ) and self-managed sequences ( $r = .64$ ,  $p < .01$ ). It should be noted that the substantial relationship between Demanding Chores and Self-Managed Sequences is partially explainable by the fact that approximately a third of the behaviors scored as Self-Managed Sequences were also scored as Intellectually Demanding Chores. Years of schooling also correlated significantly with self-managed sequences ( $r = .51$ ,  $p < .01$ ). Interestingly, the number of years a child was allowed to remain in school was not significantly related to school-passing ratio ( $r = .33$ , n.s.).

As TABLE 4 indicates, a somewhat different picture emerged for girls. As compared to those for boys, Preschool Battery test correlations with adult perception and children's behavior were more often significant. Furthermore, all three measures of adults' perceptions of children's intellectual ability correlated roughly consistently with all five tests. Correlations with Listura rankings were particularly high, ranging from  $r = .38$  ( $p < .05$ ) for Short-Term Memory to  $r = .75$  ( $p < .01$ ) for Vocabulary. In further contradistinction to the pattern observed among boys, one Preschool Battery Test, Vocabulary, emerged as most consistently correlated with nonpsychometric measures of intellectual ability, correlating significantly with all five of these, with correlations ranging from  $r = .44$  ( $p < .05$ ) for self-managed sequences to  $r = .75$  ( $p < .01$ ) for Listura. In addition, a correlation of  $.52$  ( $p < .05$ ) was obtained between Vocabulary and School Passing Ratio.

In general, the picture that emerges for girls suggests that Preschool Battery tests are measuring a more unitary intellectual ability trait, of which Vocabulary is most representative, for girls than for boys. Though again, it must be noted that sex differences in intercorrelational patterns must be viewed only as suggestive in the present study, due to small *ns*, both the pattern of inter-test correlations seen in TABLE 1 and previous factor analyses (which technique was not feasible in the present analyses because of small sample sizes) support this contention.<sup>16</sup>

As was true for boys, some agreement was seen among measures of adults' perceptions of intellectual ability. Listura correlated significantly with chores assigned ( $r = .51$ ,  $p < .05$ ) and with years of schooling ( $r = .53$ ,  $p < .01$ ). Years of schooling also correlated significantly with chores assigned ( $r = .57$ ,  $p < .01$ ), and with self-managed sequence ( $r = .37$ ,  $p < .05$ ). As was also the case for boys, years of schooling was not significantly related to school passing ratio for girls ( $r = .30$ , n.s.), though this variable was significantly related to one children's behavior measure of intellectual ability, self-managed sequences ( $r = .37$ ,  $p < .05$ ).

## DISCUSSION

With the empirical evidence of presence or absence of interrelationships between our psychometric measures (five Preschool Battery tests) and measurements of village adults' perceptions and children's behavior presented in the previous section, we are now in a position to assess the emic, imposed etic, and derived etic validity of our psychometric measurements of sample children's intellectual ability. Since the five Preschool Battery scores obtained by each child yield a pattern of intercorrelations with each adult's perceptions and children's behavior measure, it is also possible for us to define each of our

nonpsychometric measures in terms of its psychometric test intercorrelation profile. These explorations of the empirical meaning of *listura* or self-managed sequences, for example, must be viewed as suggestive rather than definitive; *ns* for the present study are small, and a certain amount of subjectivity was necessarily involved in selection of Preschool Battery tests employed here. Finally, we will attempt to assess the derived etic validity of our measures, thus defining the shared meaning of intellectual ability for rural Guatemalan Ladinos and ourselves.

### *Emic Validation of Psychometric Measurements*

Our attempt to test the emic validity of our measurements of intellectual ability began with asking villagers about the ways children could be categorized. Out of that probing emerged our knowledge of the local concept of *listura*, which appeared to coincide closely with what is commonly referred in the U.S. as "brightness" or "quickness." Our first test of whether our measurements of intellectual ability had emic validity in the host culture was to correlate *listura* rankings with five Preschool Battery tests. The match between *listura* and all tests was relatively good for girls, and it was good for two tests, Vocabulary and Perceptual Analysis, for boys.

We further explored the local emic validity of the concept of intellectual ability by postulating that if such a concept were part of the local world view, then logically it would be employed as a judgmental dimension in making the economically important decisions concerning the assignment of chores to children at an early age, and later sacrificing or *not* sacrificing an important labor source by sending and keeping a child in school. As we have reported, the match between the proportion of observations in which 5–8-year-old girls were engaged in chores, and tests scores was relatively good. For boys, the proportion of observations in which a child was engaged in intellectually demanding chores was significantly related to Learning and Memory scores. Furthermore, the number of years a child was allowed to remain in school was significantly related to four out of five test scores for girls, and to scores on three of the tests, Verbal Reasoning, Vocabulary, and Perceptual Analysis, for boys.

In summary, we believe that the evidence is substantial that the concept of intellectual ability, as operationalized in the five tests of our Preschool Battery, is emic to the rural Guatemalan Ladino villages where we are conducting our longitudinal study. As the culture of these villages incorporates rigid sex-typing it is not surprising that our data suggest that local concepts of intellectual ability differ for girls and boys. As it applies to girls, the concept of intellectual ability seems to be a relatively unitary one which is best reflected by the skills measured by our Vocabulary test. *Lista* girls tended to do well on all five of our tests and especially on Vocabulary. They were assigned many chores at a young age, and allowed to remain in school for a relatively long time, though they did not particularly excel in their schoolwork.

Our data suggest that the concept of intellectual ability as it applies to village boys may be a less unitary one. Being *listo* related significantly only to some of the tests in our battery (Vocabulary and Perceptual Analysis). *Listura* for boys was related to the ability to engage precociously in self-managed sequences, and it appears to connote a different kind of intelligence than that required to do intellectually demanding chores. Not surprisingly, village parents tend to assign such chores to boys who score well on tests of simple learning and memory, and

who are able to engage in self-managed sequences. As was true for *lista* girls, *listo* boys were likely to be allowed to remain in school for a relatively long time, though they did not tend to excel in their schoolwork.

### *Imposed Etic Validation of Psychometric Measurements*

In addition to assessing the emic validity of our psychometric measurements, we also wished to know whether our Preschool Battery measurements of village children possessed construct validity with reference to our own Western conceptions of intellectual ability. To assess the imposed etic validity of our psychometric measurements, two predictions were tested. The first was that high scorers on the Preschool Battery would be more likely to engage in self-managed complicated behavior sequences during random free-behavior observations. The use of this outcome measure was dictated by our imposed etic assumption that intellectual ability of children in any society will be reflected in the complexity of their activities. Although Preschool Battery scores showed less agreement with these behavioral predictions than had been found for villager judgments of intellectual ability, significant relationships were found between these psychometric measures and self-managed sequences. For boys, both Discrimination Learning and Perceptual Analysis scores were significantly related to the tendency to engage in self-managed complicated behavior sequences; for girls Vocabulary scores were significantly related to the tendency to engage in such sequences.

The second prediction employed to assess the imposed etic validity of our psychometric measurements concerned school performances. As previously noted, school performance is closely identified with intelligence in Western culture. Our measure of school performance predicted by Preschool Battery scores was ratio of years passed to years of school attended, and only those (33) children who had finished a year or more of schooling were involved in these analyses.

School performances were found to correlate relatively poorly with Preschool Battery scores. For boys, no test was significantly related to passing ratio; for girls, only Vocabulary was significantly related. Though the sample of school attenders in these analyses was very small, and the pass/fail measure from which the passing ratios were constructed was obviously incapable of very precise discriminations between children, we are inclined to accept the probability that school success is less related to intellectual ability in rural Guatemala than in the USA, for example. A study of antecedents and consequences of school attendance currently in progress in our study villages,<sup>12</sup> which employs much larger *ns*, has yielded Preschool Battery : school passing ratio correlations similar to those in the present study. It appears not unlikely that school success among our sample children is as much dependent upon nonintellectual factors, such as family modernity, or press for success, as upon intellectual ones. We are presently exploring this proposition in our study communities.

### *Derived Etic Validity of Psychometric and Nonpsychometric Measurements*

We have presented empirical evidence that our psychometric measurements of intellectual ability have both emic and imposed etic validity within our rural Guatemalan research setting. Thus, the concept of intellectual ability, as operationalized in our five Preschool Battery tests, constitutes a derived etic; by the logic of construct validation, the concept of intellectual ability has meaning



both with reference to our own and our host culture's conceptual systems and we have measured the relative amounts of it possessed by sample children.

Defining the shared meaning of intellectual ability for rural Guatemalans and ourselves is more difficult to do than establishing that the global concept is shared. We have had only the limited data of this exploratory study to assist us in defining the indigenous concept, and despite decades of research, there exists no commonly accepted definition of intelligence for Western society. However, it is possible to assess the derived etic validity of our psychometric measurements of intellectual ability not only as a whole, as we have done, but also for each of our five component tests, and it is through an assessment of derived etic validity of these measures and of our nonpsychometric measures that we will seek shared meanings.

For boys, perceptual analysis skill constituted a derived etic component of intellectual ability; our measurements of this skill achieved both emic and imposed etic validity. Perceptual analysis tasks have frequently been employed as an operational measure of field independence-dependence.<sup>23</sup> Our findings support Berry's contention<sup>3</sup> that measures of perceptual analysis skill represent a promising cross-cultural comparative tool.

In addition to perceptual analysis skill, our data suggest that discrimination learning may also constitute a derived etic component of intellectual ability, for boys. However, the evidence supporting derived etic status of this component of intellectual ability is less strong.

For girls in our study sample, our measurement of vocabulary knowledge demonstrated both emic and imposed etic validity, correlating significantly with all of our adult perception and child behavior measures. Thus, vocabulary knowledge constituted a derived etic component of intellectual ability, for girls.

As we have earlier noted, our measures of adults' perceptions of children's intellectual ability and our measures of intellectual ability revealed in children's behavior represent an additional source of possible derived etics. These measures originated in our understanding of the host culture conceptual system and our own culturally conditioned preconceptions about behavioral causality, respectively. The emic validity of the notion that children differ in intelligence and can be evaluated with respect to it was demonstrated by the significant associations seen between adults' perceptions measures and Preschool Battery scores; as noted, construct validation is a reciprocal process. The imposed etic validity of our self-managed sequences measure was also established by its significant associations with Preschool Battery scores.

Both the emic notion that children's intelligence can be evaluated and the imposed etic one that intellectually able children are likely to demonstrate that ability in their ordinary behavior, proved to constitute derived etics; self-managed sequences was significantly related to the number of years girls were allowed to remain in school. It was significantly related to all three of the adults' perceptions of children's intelligence measures for boys. These correlations signal a final derived etic convergence of the way in which we and the rural Guatemalan villagers who have suffered us to study them, view the world.

#### REFERENCES

1. ANGELINI, A. L. 1966. Measuring the achievement motive in Brazil. *J. Soc. Psychol.* 68: 35-40.
2. BEHAR, M. 1968. Prevalence of malnutrition among preschool children of developing countries. In *Malnutrition, Learning and Behavior*. N. S. Scrimshaw & J. E. Gordon, Eds. M.I.T. Press, Cambridge, Mass.
3. BERRY, J. W. 1969. On cross-cultural comparability. *Inter. J. Psychol.* 4: 119-128.

4. BIESHEUVEL, S. 1969. Psychological tests and their application to non-European peoples. *In* Cross-Cultural Studies. D. R. Price-Williams, Ed. Penguin. Middlesex, England.
5. BRAZELTON, T. B. 1973. Neonatal Behavioral Assessment. William Heineman Medical Books. London, England.
6. CAMPBELL, D. T. 1961. The mutual methodological relevance of anthropology and psychology. *In* Psychological Anthropology. F. L. K. Hsu, Ed. Dorsey Press. Homewood, Ill.
7. CAMPBELL, D. T. 1964. Distinguishing differences of perception from failures of communication in cross-cultural studies. *In* Cross-Cultural Understanding: Epistemology in Anthropology. F. C. S. Northrop & H. H. Livingston, Eds. Harper & Row. New York, N.Y.
8. CAMPBELL, D. T. & D. W. FISKE. 1959. Convergent and discriminant validation by a multitrait-multimethod matrix. *Psychol. Bull.* 56: 81-105.
9. COLE, M., J. GAY, J. W. GLICK, & D. W. SHARP. 1971. The Cultural Context of Learning and Thinking. Basic Books. New York, N.Y.
10. COLE, M. & S. SCRIBNER. 1974. Culture and Thought: A Psychological Introduction. John Wiley & Sons, Inc. New York, N.Y.
11. CRONBACH, L. J. & P. E. MEEHL. 1955. Construct validity in psychological tests. *Psychol. Bull.* 52: 281-302.
12. IRWIN, M., P. L. ENGLE, R. E. KLEIN, C. YARBROUGH, & S. ROSENHOUSE. 1975. Ability, sex, and schooling in rural Eastern Guatemala. Manuscript in preparation. DHD/INCAP. Guatemala.
13. KAGAN, J. 1975. Personal communication.
14. KARP, S. A. & N. L. KONSTADT. 1963. Children's Embedded Figures Tests. Cognitive Tests. New York, N.Y.
15. KLEIN, R. E., O. GILBERT, C. A. CANOSA & R. DE LEON. 1969. Performance of malnourished in comparison with adequately nourished children on selected cognitive tasks (Guatemala). Annual Meeting of the Association for the Advancement of Science. Boston, Mass.
16. KLEIN, R. E. & H. E. FREEMAN. 1975. Indigenous conceptions of intelligence. A cross-cultural replication. Manuscript in preparation. DHD/INCAP. Guatemala.
17. KLEIN, R. E., H. E. FREEMAN, & R. MILLET. 1973. Psychological test performance and indigenous conceptions of intelligence. *J. Psychol.* 84: 219-222.
18. KLEIN, R. E., M. IRWIN, P. L. ENGLE & C. YARBROUGH. 1975. Malnutrition and mental development in rural Guatemala: An applied cross-cultural research study. *In* Advances in Cross-Cultural Psychology. N. Warren, Ed. Academic Press. New York, N.Y. (In press).
19. MEJÍA PIVARAL, V. 1972. Características económicas y socio-culturales de cuatro aldeas Ladinas de Guatemala. Guatemala Indígena (Monografía), Vol. VIII(3).
20. NERLOVE, S. B., J. M. ROBERTS, & R. E. KLEIN. 1974. The smart child: Women's judgments of *listura* in two Guatemalan communities. American Anthropological Association Meetings, Symposium on Comparative Field Studies in Child Socialization. Mexico City, Mexico.
21. NERLOVE, S. B. & A. S. WALTERS. 1975. Pooling intracultural variation: A composite rank-ordering of community judgments. Mathematical Social Science Board Conference on Standardization and Measurement in Social and Cultural Anthropology. Coloma, California.
22. NERLOVE, S. B., J. M. ROBERTS, R. E. KLEIN, C. YARBROUGH, & J-P. HABICHT. 1974. Natural indicators of cognitive development: An observational study of rural Guatemalan children. *Ethos* 2: 265-295.
23. DHD-INCAP. 1975. Progress Report 1974/75. Contract PH-43-65-640. Division of Human Development, Institute of Nutrition of Central America and Panamá. Guatemala.
24. TUDDENHAM, R. D. 1963. The nature and measurement of intelligence. *In* Psychology in the Making. L. Postman, Ed. Knopf. New York, N.Y.
25. WILKIN, H. A., R. B. DYK, H. F. FATERSON, D. R. GOONEOUGH, & S. A. KARP. 1962. Psychological Differentiation. John Wiley & Sons, Inc. New York. (Republished in 1974. Lawrence Erlbaum Associates. Potomac, Md.)