Introduction

Purpose of the Review

In many Latin American countries creative efforts are being made to develop and/or adapt instruments designed to measure the psychosocial development of children from 0 to 6 years of age. The present review is an effort on the part of the Consultative Group on Early Childhood Care and Development.
Care and Development to contribute to the more optimal functioning of such instruments within the context of early intervention programs. The purpose of the review is to:

- focus attention on the importance of adequately developing and validating these instruments
- suggest specific criteria on which instruments could be selected
- examine existing instruments in the light of these criteria, and to make recommendations concerning future needs in regards to the development and implementation of such instruments.

Why the Topic is Important

Throughout the world, efforts are being made to improve the chances of survival and the quality of life of children, particularly during the early years. In order to implement strategies and to provide services in the most effective way, it is often necessary to identify those children who are at greatest risk of presenting developmental problems. Once programs are implemented it is also necessary to evaluate their impact on children's development and family functioning. In order to do this, measurement instruments are required, the characteristics of which should be adequate and appropriate for the task at hand.

Although the need for such measures has now become accepted in many places throughout Latin America, the importance of insuring their adequate design and validity has not received as much attention. The failure to use appropriate measures can cause considerable damage to the children and families who the services intend to help. Children and families in need of services may go unidentified and others may be needlessly alarmed. Limited resources will be used inefficiently to the detriment of the implementation and spread of early intervention services. It is therefore important to assure that the instruments being developed and in use today in Latin America possess the characteristics necessary to function adequately within the context in which they are being used.

This is not an easy task. The measurement of psychosocial development and of factors related to it, such as the quality of childrearing environments, is in and of itself quite complex. The identification of relevant risk factors and the detection of developmental difficulties is even more so. Added to this initial complexity are the requirements of simplicity and cultural relevance imposed by the programs in which the instruments are to be used. In addition, limited economic resources often represent important obstacles to the adequate development and validation of such instruments.

In spite of these problems, innovative efforts have been made throughout the region and many promising alternatives now exist. Frequently, however, important issues have remained unresolved which seriously limit the utility of the instruments. By not dealing adequately with these issues, considerable work and economic resources will be wasted, which in the long run will prove more costly than investing the additional time and effort in adequately resolving them from the beginning.
Which Instruments Were Reviewed

The present analysis is based on instruments which share the following characteristics:

- They are all being tested for use or are already in use in Latin America.
- They describe or are based on areas of psychosocial development of children between 0 and 6 years of age, the characteristics of the child's home environment and/or family interaction or risk factors related to developmental problems.
- They are relatively simple, designed for use in the context of service related programs which usually rely on limited resources and are directed toward the resolution of practical problems related to children's health and well-being. These measures require neither sophisticated training nor expensive materials or equipment.

Included as "instruments" are such diverse forms of measurement as developmental scales or tests, questionnaires and growth and developmental charts ("carnets" or "cartillas"). We are aware that not all the presently available alternatives have been included in this analysis. There are probably other options of which we are not aware. Furthermore, only those instruments for which some background information was available to the author have been described. Several instruments which are in the initial phases of development were also left out of this analysis.

Organization of the Paper

Section II of this review presents some basic conceptual issues which underlie the analysis. How one conceptualizes the process of child development and the factors which influence it will determine, to some extent, what one wishes to measure and how one will do it. We therefore consider it essential to establish a conceptual foundation for the analysis of the instruments. In Section III, specific criteria for the selection and evaluation of the instruments are described. For each criteria, a discussion of relevant issues and information has been included. Once these criteria are established, Section IV is devoted to the comparison of the instruments. First their characteristics are summarized, and then they are evaluated according to the proposed criteria. In the final part of the paper, Section V, general conclusions and recommendations are presented.

Conceptual Foundations

Underlying each instrument there must always be a conceptualization of child development and of the factors which influence it. Frequently such a conceptual foundation is not made explicit nor subject to critical analysis. We propose that the way one conceives of child development and of the factors which influence it should be the topic of serious consideration when developing or choosing any kind of instrument. Furthermore, because our knowledge of developmental processes has increased so much in recent years, it is imperative to consider an instrument's conceptual underpinnings in light of recent findings.

In this section we will briefly mention some basic principles which underlie our conceptualization of child development and of the factors which influence it. We hope this will facilitate the
readers' understanding of the sections which follow and contribute to a more critical analysis of the measurement instruments.

- **CHILD DEVELOPMENT IS A CONTINUAL, ON-GOING PROCESS WHICH PROCEEDS THROUGH QUALITATIVELY DIFFERENT STAGES IN AN INDIVIDUALLY UNIQUE AND IRREGULAR FASHION**

  - Early measures of performance may not be related to later measures since they reflect qualitatively different stages of development. As far as possible, early predictors should reflect precursors of the processes which one hopes to predict. An example of this comes from the prediction of I.Q. Traditional sensorimotor behaviors included in many infant tests seem to be poorer predictors than measures of early information-processing which appear to measure underlying cognitive abilities (Zelazo 1982; Kagan and Klein 1973; Lasky, Klein, Yarbough and Kallio 1981).

  - The rate at which different children proceed through the different stages may be quite different. Furthermore stages are not homogeneous but may reflect uneven performance on different skills or behaviors. Therefore, how well a child does at an early age does not necessarily have a 1:1 relationship with how well he or she will do when older.

- **CHILD DEVELOPMENT IS MULTIFACETED, INVOLVING DIVERSE PROCESSES WHICH ARE, HOWEVER, INTER-RELATED AND INTER-ACTING**

  - Child development reflects the integration of genetic, biological, psychological and social processes.

  - Development itself covers diverse areas of performance which can be conceptually separated but mutually influence each other to varying degrees.

  - Developmental problems are not unitary. Children can have difficulties in certain areas of development and be relatively problem-free in other areas. This is not to say that a deficit in one area does not affect other areas. Rather, it is important to remember that a child may have a deficit in a certain area while functioning well in others.

  - Developmental problems in certain areas will become manifest at earlier ages than others. This means that certain developmental problems can be detected at certain ages and others at other ages. Alterations of movement, posture and tone due to perinatal complications become apparent between 6 and 18 months, and congenital sensory deficits can be identified at birth and in the subsequent months. Moderate and mild, but important language and cognitive deficits, on the other hand, can not be accurately identified or predicted at such early ages.

- **DEVELOPMENT IS A DYNAMIC, INTERACTIVE PROCESS IN WHICH BOTH THE ENVIRONMENT AND THE CHILD'S CHARACTERISTICS RECIPROCALLY INFLUENCE EACH OTHER**

  - Developmental outcome is strongly influenced by the quality of the childrearing environment.

  - The quality of the post-natal environment will modify the effect of early biological risks, ameliorating or exacerbating their influence.
The child's characteristics influence the environment itself and modify the impact of the environment on development. The way the environment treats a child will be partially determined by the way the child is and is perceived (Samaroff and Chandler 1975). Furthermore, some children possess characteristics which allow them to thrive in very unfavorable surroundings while others suffer under relatively favorable circumstances (Werner 1987; Zeitlin and Mansour 1985).

The childrearing environment itself changes over time, both in response to the child and in reaction to other endogenous and exogenous influences.

The family environment also possesses characteristics that make it more or less resilient in the face of stress and crisis.

Certain self-righting tendencies can be observed throughout the developmental course.

- Early risks often attenuate over time.
- Many children seem to possess remarkable resiliency in face of stress and are able to overcome early disadvantages.

Having mentioned these basic underlying concepts, we will now turn to the criteria which we propose should be used to develop and/or evaluate the measurement instruments themselves.

Criteria for Selecting/Evaluating the Instruments

The purpose of the instrument should be specified and the instrument's characteristics should be congruent with the stated purpose. The instruments are designed and used for one or more of four basic purposes: description or diagnosis of the child's developmental level or of the family environment; detection or screening for developmental disabilities or environmental deficiencies, or risk thereof; periodic monitoring on an individual and/or community basis; and programme evaluation. Each of these purposes imposes specific requirements on the nature of the instrument itself.

- **Description or Diagnosis**

  These instruments should give a relatively complete description of a child's functioning in order to provide sufficient information concerning existing strengths and weaknesses so as to be able to plan an individualized therapeutic program when necessary. They should have the following basic characteristics:

  - Cover a wide variety of developmental areas or, if designed to describe the childrearing environment, of aspects of the quality of caregiver-child interaction.
  - Identify areas separately so as to be able to pinpoint strengths and weaknesses, providing a profile of functioning, not just a global score or quotient.
  - Include several items in each area so as to provide as complete a description as possible.
Be congruent with present-day knowledge of child development and environmental processes which foster child development.

Be based on data obtained on a population of children or families similar to those with whom they will be later used.

Be appropriate for use by the kind of personnel available in the area in which the instrument is to be used.

Report results in such a way so as to be easily combined with other evaluations to form an integrated, interdisciplinary diagnosis.

DETECTION OR SCREENING

The purpose of child development screening is to identify those children who are at risk of developmental problems or those who are showing some evidence already of a developmental problem. These instruments will NOT fully describe the child's characteristics. Rather, they are intended only to separate individuals or populations into two or, at most, three categories: at-risk vs. not at-risk: or abnormal, suspect and normal. The goal of such a broad classification is to identify children who appear relatively normal, i.e., who have not been previously identified as having a problem, and who might benefit from general preventive interventions or, if needed, could be referred for more complete assessment. Furthermore, such information at a community level could provide important data concerning the frequency and distribution of developmental risk and problems which would be impractical to obtain on a massive scale using more detailed instruments.

Instruments designed for detection or screening should have the following characteristics:

- Be short and inexpensive in terms of time required for application and simplicity of materials required.

- Be designed to be applied by personnel with only a minimal amount of training in their application. The characteristics of such personnel will vary in different countries and settings, according to the resources available.

- Include items which have clear implications for the kind of developmental problem the instrument is designed to detect. Redundant or irrelevant items should not be included. Neither should those which do not have a clear interpretation, i.e., can be considered normal if or if not present. Behaviors should be placed at ages which clearly indicate risk, that is, at which most or all healthy children would have been expected to have achieved the behavior.

- Include clear indications concerning the interpretation of results and procedures to follow in the case of a positive detection.

- Optimize their inherent educational value. Since the effectiveness of most interventions will necessarily depend on the parents, it is important to design the instruments in such a way that they facilitate the evaluator's communication with them. Ideally, the instruments should be easily understood by parents and health personnel. Their use should reinforce the parents' own efforts on behalf of their children, orient them when necessary toward improvements, and teach them to be the primary detectors of their children's developmental problems.
PERIODIC MONITORING

These instruments are designed to assess changes over time both for individual children or families as well as on a community, regional or national level. They may function as descriptive/diagnostic or detection/screening tools at each point in time. The primary difference is that these instruments also provide a way of comparing the status of children overtime. The characteristics of these instruments are the same as those of the others described earlier, with two additions:

- The instrument must provide continuity across ages in terms of the areas and developmental indicators included.
- Some provision should be made to facilitate comparison of results over time.

PROGRAM EVALUATION

These same instruments, and others, concerning psychosocial development and environmental/family interaction have been and will be used in the future as important aspects of program evaluation. The characteristics they should have will vary in accordance with the programme content and dimensions, which are even more diverse than the instruments used to assess them. The selection of appropriate instruments will be critical, since otherwise a fair evaluation of programme effectiveness will be impossible. The following general indications may help facilitate the proper selection of evaluation instruments.

- The nature of the evaluation instrument(s) should be congruent with the program's objectives and ideally should be decided upon during the project planning stages.
- The instruments should measure outcomes or processes which could feasibly have changed during the time period involved in the intervention.
- They should be aimed at measuring both direct as well as indirect effects of the intervention which could have continuing impact on future development, whether these be positive or negative changes. Many of these will involve behavioral changes of the family, mother, father and/or child which will vary according to the specific nature of the intervention. Among the more common possibilities for such measures are: changes in the family's utilization of health services and participation in the community; mother's self-esteem, confidence in the mothering role, continuation of schooling; the child's self-esteem, ability to cope with stress, social attractiveness, problem-solving motivation, adaptation to the school environment, etc. Considering the great diversity of possible dimensions to be evaluated here, it is not possible to include them in the instruments to be reviewed in this analysis. However, it is important to mention that measures of such processes are a very essential part of program evaluation and efforts should be made to include them wherever possible.

The developmental problems which the instrument intends to detect or describe should be clearly specified. One of the first steps in developing or selecting an instrument is to clearly define what it is supposed to measure, i.e., what aspects or problems of development are of interest. Not only will the selection of the indicators which the instrument includes depend on
the clear definition of the target objectives but also the ability to adequately validate the instrument will depend on knowing what it should be measuring. This requirement is more complex than it seems at first glance and, in fact, is not always fulfilled. Especially if an attractive instrument is readily available, practitioners may adopt it without being aware of its objectives or conceptual underpinnings. The opposite problem also exists. Although, at times, appropriate instruments do exist and could be easily adapted for use in a new program, a new instrument is developed, often without adequate methodological foundations, due to the pressure to begin the program's action phases.

At this point it is important to clarify the difference between the way we are using the words "purpose" and "objectives." As described in the previous section, what we are calling "purpose" refers to whether the instrument is designed for complete diagnosis or description, detection or screening, monitoring, or program evaluation. However, the word "objectives" is being used in reference to the target of the instrument in terms of the developmental processes one wishes to describe or detect.

Different approaches exist to defining developmental target behaviors. Some authors define behavioral objectives in terms of areas of development. The most common divisions are based on Gesell's original separation of behaviors into language, large motor, social-emotional, and fine motor-adaptive areas. Other authors separate out cognitive functioning and perceptual abilities, as well as self-help skills and feeding behavior. Others establish content and divisions on theoretical bases, most notably following conceptual areas proposed by Jean Piaget.

Still others take a more clinical focus to defining developmental objectives and direct their attention to developmental problems or syndromes. Among the most common categories here are alterations of tone, posture and movement (sometimes called Cerebral Palsy); sensory problems, including visual and auditory deficits; cognitive deficits, including mental retardation; language disorders; specific learning disabilities, a term which groups problems in the areas of selective attention, activity regulation and academic performance in spite of normal I.Q.; and social-emotional problems.

Both alternatives, by areas and clinical problems, are actually quite similar. However, the way they are operationalized may be very different. In the first case the definition may be somewhat tautological: the problem the test is designed to measure is defined as a delay in the behaviors measured by the test. In the second case, there is a greater tendency to look for symptoms and signs which may cross over developmental areas and may be confirmed by observations made independently of the test instrument.

Still other approaches exist to defining the relevant developmental phenomena. Some authors (Ricciuti 1974) have proposed that clearly defined developmental processes should be measured systematically, both in order to predict developmental outcome and to evaluate the impact of specific interventions. In this vein, measures would be directed at underlying processes such as curiosity, attention, information-processing, goal attainment motivation, temperament, and sociability. Experimental evidence suggests that measures of these processes may be even more predictive of future development than the more commonly used behaviors taken from
standardized infant scales (Bornstein and Sigman 1986; Rose 1981; Caron and Caron 1981; Lewis and Goldberg 1969; Zelazo 1982).

In the case of detection/screening instruments, the severity of the developmental problems to be detected should also be defined. In some contexts very severe disabilities will be so obvious that measurement instruments will not be needed to identify them. However, moderate to mild— but functionally important—difficulties often go unidentified. If they were found, such conditions often can be ameliorated by appropriate interventions.

The severity of the conditions to be detected will, of course, determine to a large degree the number of children who will be identified as cases and who will be in need of services. Therefore, decisions concerning cut-off points will have to be made in light of the resources available for meeting the needs detected.

The indicators should be appropriate for the target objectives and for the population in which the instrument is to be used. Once the target objectives of an instrument have been clearly specified, the indicators should be selected to reflect them as closely as possible and be appropriate to the context in which the instrument will be used. A n indicator, in the sense it is being employed here, is any kind of measurement which is used to describe some aspect of the child's development, the quality of the environment or risk of developmental problems. Groups of indicators are combined in diverse ways in different kinds of instruments (such as scales, questionnaires and charts). For the purpose of our discussion we will divide the indicators into three categories according to their conceptual domain: behavioral manifestations in the child, quality of the environment and/or interaction; and risk factors in the child's background.

- **Behavioral Manifestations in the Child**

  The first category, behavioral manifestations in the child, is composed of the indicators most frequently included in developmental tests. This is not to say that they are necessarily the indicators which should always be used, but rather that they are what most people think about when considering such instruments.

  **Direct measures of the child's behavior.** A wide variety of possible indicators exist which reflect the diversity of development itself as well as different theoretical approaches to describing and predicting child development. Many instruments used in community-based programs have incorporated different but similar subsets of behavioral indicators, which in large part are based on the behaviors included in the most widely used baby tests, such as the Gesell, Bayley and Griffiths developmental scales. These latter scales describe the child's developmental level relatively well at the time of the evaluation in terms of the developmental indicators they include. They primarily cover sensorimotor skills, which depend largely on fine and gross motor abilities, as well as language production, and to some extent, comprehension, and the child's social interaction with the examiner (Zelazo 1982).
Over the years, widespread use of these original tests has demonstrated important limitations which should be considered before assuming that by taking indicators from these tests one necessarily can be sure the new instrument will be effective.

- In the first four years of life, the predictive, as opposed to concurrent, validity in relation to childhood I.Q. of these original baby tests is very poor when dealing with open populations (i.e., those which have not previously been screened) in which the baseline of developmental problems is low (McCall 1982).

- In high risk, pre-screened samples, where one would expect to find more developmental problems, the predictive validity is somewhat higher, although often the frequency of correct predictions is such that one would usually be wrong more often than right when trying to predict intelligence from infancy to childhood (McCall 1982).

- Developmental performance, on the kind of items included in the standardized infant scales, predicts future I.Q. when applied after a child reaches two or three years of age to a greater extent than when applied in the early months (Werner et al., 1987; Broman et al. 1975).

- Such indicators are more useful when combined with evaluations of the childrearing environment and pediatric/medical assessment of risk factors and neurological status. It appears that each of these elements contributes a piece of the prediction puzzle which the others do not necessarily provide (Werner, Honzik and Smith 1968).

- Behaviors not usually included in the infant tests may have more predictive validity for some developmental objectives, including I.Q. Among such indicators are a variety of alarm signs which clinicians have identified as precursors of later developmental pathologies, characteristics of the child's temperament or behavioral style, and a variety of information-processing abilities which infants possess but which are only beginning to be identified by researchers (Blackman 1986; Zelazo 1982.)

- The behavioral indicators should be appropriate for detecting the instrument's target behavioral objectives. For example, some indicators from traditional baby tests (such as visual and auditory responses to specific stimuli) are excellent indicators of certain developmental problems, such as sensory deficits, but inappropriate for others, such as cognitive abilities. Indicators should be chosen on the basis of their demonstrated relationship to the developmental outcome which they measure rather than on the frequency with which they are found in traditional developmental scales.

**Indirect measures of child behavior.** Although the most common indicators rely on direct observation of the child's behavior, some instruments evaluate these behaviors through indirect means, such as parents' or teachers' reports. Both alternatives, direct and indirect measures of behavior, have their pros and cons. It has been shown that instruments based on mother's report of child behavior (e.g., Ireton and Thwing 1972-1974) can be reliable across time and concurrently valid (Guerin and Gottfried 1987). However, since the validity of mother's report is associated with her level of education and the time she spends with her children, i.e., her opportunity to observe her child's development (Gottfried et al. 1984), such measures may be
appropriate only in certain socio-cultural contexts. Therefore, the validity of the indicators should be ascertained within the kind of setting in which the instrument will be used so as to avoid the adoption of measures which work well in other populations, but are inappropriate to the local setting.

QUALITY OF THE ENVIRONMENT AND/OR INTERACTION

In the second category, the quality of environment and/or interaction, we include indicators which characterize how well the child's needs are being met by his or her functional environment, and to what extent the interpersonal relations are facilitative of the child's psychosocial development. These indicators identify manifestations or consequences which have occurred in the child's environment, in contrast with the background risk factors which speak to an increased probability that a problem will occur and will be discussed in the next section. (Readers should be aware, however, that some classification systems will include these indicators as established risks in relation to developmental problems.)

While measures exist which characterize home environments and family functioning independent of possible impact on child development, the focus of our interest on environmental measures is their relationship to the child. It is therefore important that the environmental dimensions included be related to those aspects of child development which are of most interest to the project.

The definition of environmental characteristics that an instrument is designed to measure is as complex, if not more so, than the definition of the behavioral manifestations to be measured. A great deal has been learned only recently concerning the essential elements within the childrearing environment which foster healthy child development. However, differences of opinion as to which aspects are most important will be common. Nevertheless, it is important that the dimensions chosen be those which are most relevant to the behavioral objectives of interest.

Here too one can distinguish between indicators that are obtained from indirect sources vs. those which involve direct, "hands-on" measures based on observations of the childrearing environment and/or the child's interaction with his or her caregivers. Data based on indirect sources have the advantage of being relatively easier to obtain, since it is not necessary to directly evaluate the home environment itself. A variety of secondary sources can be tapped, including local health records, as well as interviews with the mother or father. The disadvantage of such indicators is, of course, that they are one or more steps farther away from "reality" and therefore may be less accurate in identifying the environment/interactional problems that they are supposed to identify. Caution should be taken to insure that their relationship to more direct indicators has been previously tested and that it is acceptably high.

RISK FACTORS IN THE CHILD'S BACKGROUND

The last category, which corresponds to risk factors in the child's background, includes biological as well as psycho-social factors which have been demonstrated to be associated with an increased probability of later developmental problems in children and/or with environmental disadvantage.
and/or problems of parent-child interaction. Our emphasis is on background risk factors which indicate a probability of some consequence occurring, including inadequate childrearing conditions and/or developmental difficulties.

Within this category, the risk can be separated into potential and established risk factors. While preterm birth, and other such potential risks, may or may not affect a child's development, other factors such as Down's Syndrome or sensory impairment will necessarily have deleterious consequences. Nevertheless, in the latter case, although their effects are clearly established within the child, they are still considered risk factors, in the sense that the final developmental outcome is not yet known and will be determined by a variety of influences.

Here again it is very important to specify what one wants to predict when choosing the appropriate risk factors. The key question becomes: risk of what. In order for the risk factors to identify cases efficiently, the relationship between them and the kind of environmental and/or developmental problem of interest should be clearly demonstrable in the kind of population where the indicators will be used. This is specially important when adapting an instrument developed in a different setting, in which some of the risk factors may have different implications, cut-off points and incidence.

The importance of developing and using appropriate and accurate risk indexes is considerable and has widespread implications for all kinds of community programs. They should be used as first stage screening instruments to identify families or children who should then be screened or evaluated more closely with instruments which identify problems in the child's environment and/or developmental performance. Such a system would optimize the available resources by reserving the use of more detailed instruments for groups of families and children in which higher baseline rates of problems are expected. By increasing the prevalence of problems in the population to be subsequently evaluated, it is possible to greatly increase the predictive validity of the measurement instruments used.

### COMBINATION OF BEHAVIORAL, ENVIRONMENTAL AND RISK INDICATORS

Opportunities for developmental and environmental evaluation can be grossly divided into two time periods: those which are based around birth and include indicators which relate to pregnancy, birth and puerperium, and those which take place at later points in the child's development. The first kind will be geared to screening efforts designed to identify high risk infants who will require greater attention through follow-up evaluations or specific intervention programs. However, it is important to remember that many of these infants will improve over the first years and perform normally later on. Therefore, it is essential that detection and evaluation programs continue to identify children beyond the perinatal period. Post-natal evaluations will identify risk and/or problems in any or all of the three domains, employing indicators which identify more contemporary influences and effects. A perinatal risk factor such as low birth weight for gestational age will be less relevant at later ages, for example, 4 or 5 years of age, at which time the child's developmental performance will be of greater value. The earlier detection instruments should rely most heavily on risk factors, indicators of future environmental problems and measures which identify the infant's biological intactness, and potential vulnerability. Later on, indicators should include those which identify important changes in the post-natal
environment, but should center more closely on the quality of the childrearing environment and, especially after 2 years of age, on the child's developmental performance, using indirect or direct indicators.

**The instrument should be culturally appropriate.** In addition to the conceptual bases necessary for the selection of indicators, their cultural relevance presents another issue of extreme importance and complexity. This becomes most important when selecting instruments developed in other cultures or when basing an instrument on research in other countries. Often the same indicator will have different meaning in different socioeconomic and cultural settings.

**Behavioral measures**

While some milestones and developmental problems are relatively universal and equally valued among different populations, others are more or less emphasized in different cultures and some milestones may occur at somewhat different ages. The most important point here is that the behaviors chosen, which have already been selected for their relevance to the instrument’s objectives, should also be behaviors which children have had the opportunity to practice. They should be seen as important and understood by the parents so as to be useful both in terms of the education implicit in the evaluation itself and as increasing the instrument's appropriateness for intervention.

A typical example is that of "rides tricycle." Here the intention is to evaluate the child's motor coordination. However, if tricycles are so uncommon or can not be used due to uneven terrain, other motor behaviors of equivalent intention can be employed, such as "climbs trees well" or "walks on stone or adobe fences." Likewise, crawling and creeping may be important indicators in contexts where children are allowed to be on clean floors, but not favored in communities where floors are made of dirt and represent important health hazards if the child were left to crawl around on them. It is precisely due to the importance of such cultural influences that the WHO has undertaken a multicenter initiative to identify culturally-appropriate behavioral indicators (WHO 1985). Data from this project, which includes a site in Argentina, will provide important guidelines for identification and use of naturalistic indicators.

Furthermore, parents' goals for the socialization of their children differ in different countries, communities and between ethnic and social subgroups. While docile, obedient children may be highly valued and adaptive in one context, such behavior might cause concern on the part of parents and educators in another setting where assertiveness and independence are valued and/or required. Obviously, such differences should be taken into account when designing or adapting developmental measures. The relevance of the indicators should be analyzed so as to include items which have developmental significance both in universal and local terms.

The ages at which the behaviors are included should also be appropriate to the cultural context. It is well known that individual healthy, well developing children will present important milestones at somewhat different ages within an expected range. This expected age range may differ somewhat within and between different cultural groups, depending on the conditions in which children are brought up and on possible genetic variations. It is therefore ideal, though
often not practical, for instruments which were developed in another country to be standardized
with the kind of population in which they will be used. An instrument being developed
specifically for use in a certain population should also be standardized on a population which
includes the subgroups to which it will later be applied.

In general terms, this implies testing an instrument adequately on a sufficient number of children
of the relevant ages, as well as of the major socioeconomic and cultural groups within the
populations, such as rural vs. urban, affluent, middle class, poor and very poor, etc. Both sexes
should be equally represented in all of the resulting groups. Furthermore, the children should be
carefully selected so as not to include those who would be potentially deficient in their
performance, such as children with perinatal risk or concurrent health problems. High risk
children will be evaluated at a later stage or as a separate group in order to test the instrument’s
ability to differentiate them from the healthy children. However, the age norms should be based
only on those children who are not likely to present deficiencies or alterations in their
performance. The greater the number of children who can be included in each category, the
better will be the reliability of the age norms established. A s a minimum, it is usually
recommended that at least 30 children be included in each subgroup.

ENVIRONMENTAL MEASURES

In the case of instruments which include measures of the quality of the childrearing environment,
the issue of cultural relevance is very important. While it is possible to agree on certain broadly
defined, conceptually based functions the family or family substitute should perform and those
which describe facilitative caregiving, the specific content and their relative importance will
depend on the cultural context in which the evaluation will take place.

O nce one has decided on the conceptual content of the environmental indicators of, for
example, "opportunity for play", one still must develop an indicator which will discriminate the
differences that occur in the setting in which the instrument will be used. A n indicator such as
"number of toys available" or "number of meters of play space" will not be applicable in all
contexts. In some settings, such as rural farming communities in Latin America (in some
countries), toys may be so few and play space so unspecifiable and extensive that these indicators
will have to be seriously modified. The intention may remain the same, but the way to measure it
will have to change. Perhaps a measure of whether or not the child is even allowed time to play
would be more discriminating.

O nce specific items have been chosen that are relevant to the socio-cultural setting, the
appropriate cut-off points can only be decided on adequately by looking at the way they vary in
the target population. For example, such a universal indicator as maternal education does not
have universal significance or cut-off points.

W hat is considered low in one context, e.g., high school level, would be exceptionally high in
another. A lthough this example is quite obvious, others are less so, but still are equally important.
W hether one is measuring such "concrete" variables as "space available for play" or more
qualitative concepts such as "maternal sensitivity", the same challenge exists. The item must be
defined in culturally and socially relevant terms and the cut-off points established on the basis of empirical data.

The instrument should be reliable and internally consistent. From the beginning, when the instrument is being developed, items should be chosen which are able to be applied reliably, in other words, items which do not require personal interpretations in their form of presentation or in the way they should be scored. This should also be true with respect to the overall outcome of the test. The rules for determining the final result and the actions to be taken from it should be specified clearly so that each examiner would come to the same conclusion based on the same family’s or child’s responses or performance.

Furthermore, it is important for inter-observer reliability to be demonstrated with the kind of personnel who will be applying the instrument during the program. Ideally, a profile should be established that describes the characteristics which the person who will apply the instrument should have, based on initial studies which compare the reliability of the instrument when applied by different kinds of health personnel. What is essential is that the instrument be reliable in the hands of the people who will be applying it during the program.

It is also important to show that the instrument has good stability over time so as to produce identical or highly similar results in a period of time over which "real" changes in the family or child have not occurred. This is somewhat problematic in the case of developmental instruments given the inherently changing phenomenon they are measuring. Real changes occur extremely rapidly during the neonatal period and, to a lesser but still considerable degree, during the first two years of life. Nevertheless, it is important that test-retest reliability be demonstrated over a time period in which developmental changes will not have been expected to have occurred.

Many instruments establish conceptually meaningful areas into which they group the different indicators. Such divisions, if consistently and conceptually based, have the potential of facilitating the interpretation of results and the interface with programmatic components. However, before being able to assume that the items are related and correspond to a certain area or concept, the internal consistency of the scales must be demonstrated. This can usually be done through factor analytical or correlational techniques based on the data obtained with the normative sample. Items which do not relate to the areas posited, or to those derived from the analysis, will have to be eliminated or moved, according to their relationships with other indicators.

The instrument should have adequate concurrent and predictive validity. Even though all of the above criteria were to have been complied with, it is essential to know whether or not the instrument measures what one hopes it measures, i.e., whether it is or is not valid. Especially in the case of detection instruments, the question of their ability to really identify the families or children they are supposed to identify, i.e., those at risk or already with problems, is such a critical issue, and is so often ignored in community-based programs, that we will describe in some detail the kinds of data that should be obtained before an instrument is implemented as a valid detection tool. Although we will refer to the instrument to be validated as a detection or
screening instrument, the comparisons described are applicable to any kind of instrument one wishes to compare to an external standard.

The basic comparison which is needed is between the detection instrument's result and that of a more detailed, complete evaluation with known and accepted validity. This more complete evaluation is usually performed or supervised by professionals and should be sufficiently well documented that it can be considered the external criterion of "reality" or "truth". This is not an easy requirement to fulfill. However, to the extent it is possible, one should choose a well accepted external "truth" criterion when testing the accuracy of the detection instrument. In the case of developmental instruments, the use of reliable measures in the hands of well qualified clinicians or actual performance in school would be the best alternatives. Furthermore, the criterion evaluation should reflect the instrument's target objectives and not be based on some available but irrelevant test.

The validation of the environmental instruments may take either one of two approaches. If possible, one could use an external measure of the childrearing environment based on careful observations and interview techniques involving home visits. The other alternative is somewhat more indirect. It involves comparing the environmental instrument to developmental outcome. However, such a comparison confounds the validity of the instrument itself and the association between the environment and developmental level, but in some cases is the only feasible alternative.

The main idea of any validation effort using an external criteria is to obtain an independent evaluation which can reasonably adequately tell us if the family or child really has or has not the risk or problem that the detection instrument says exists. The following table (table not available) illustrates the different comparisons which can be made, based on the result of both evaluations. A "positive" result on the detection instrument means that the family or child has been identified as being at risk or having the problem the instrument is designed to detect. This may be confusing to some, since "positive" is often associated with better or healthy. However, it is common practice when talking about the results of screening to use "positive" to refer to cases which have been identified as having the risk or pathology under question. In a similar fashion, a "negative" result means that the detection instrument considered the family or child to be without risk or problems. In this sense "negative" means healthy or free of problems.

Through this comparison one can calculate a variety of important indexes, which will provide a good idea of how well an instrument is able to detect target subjects. True Positives (TP) are those who were detected by both instruments as being at risk or having problems and True Negatives (TN) as being healthy. The inaccurately identified cases are the False Positives (FP), those who were identified by the detection instrument as at risk but are really fine, and the False Negatives (FN), those who represent families or children who really have problems but were missed by the detection instrument.
Based on these categories, one can calculate the following proportions:

**Sensitivity** - This is the instrument's ability to adequately identify those families or children who really have risk or problems. Sensitivity is the proportion of True Positives over all the real cases in the populations, those that were detected (TP) and those that were not (FN). Therefore, sensitivity can be calculated by TP/(TP + FN). Obviously, if a detection instrument misses a lot of real cases, i.e., has a lot of FN's, it will have low sensitivity even if it detects some cases accurately.

**False Negative Ratio** - This is the complement of sensitivity: it tells us what proportion of real cases (TP + FN) were missed (FN). In other words, FN/(TP + FN). This index is an easy way of seeing what proportion of children or families who really need the intervention will not be identified as such.

**Specificity** - A detection instrument should also be good at identifying which families or children are really healthy and do not need special attention. The instrument's specificity is a measure of what proportion of truly normal subjects are correctly identified as such. An instrument's specificity will be the proportion of True Negatives (healthy subjects identified as such) over all healthy subjects (those which the instrument identified, as well as those which the instrument incorrectly called cases). This can be calculated as TN/(FP + TN). As can been seen through this formula, if an instrument over-refers a lot of cases, i.e., calls abnormal many healthy subjects (many FPs), the instrument's Specificity will decline, even though it accurately identifies some subjects as healthy (TN).

**False Positive Ratio** - This measure, the complement of specificity, describes the proportion of healthy individuals or families who have been incorrectly, and unjustly, labeled as abnormal. This can be calculated by the following formula: FP/(FP + TN). This proportion will tend to be unduly large in those instances in which the target problem or problems are infrequent in the population tested. It is an important index of the extent to which the instrument may unnecessarily alarm parents whose children are actually healthy.

It is important to note that an instrument's sensitivity and specificity should be considered together. An instrument can achieve very high sensitivity, i.e., identify all cases with risk or problems, by ignoring the questions of specificity. It would then over-refer many families or children. In other words, it would be so general that many healthy subjects would fall into the positive category. Such an instrument would be relatively wasteful of resources. Furthermore, it would require subsequent re-evaluations to reclassify healthy, non-risk subjects.

On the other hand, an instrument could achieve high specificity by ignoring the issue of sensitivity. Such a tool would be very good at identifying healthy individuals, but would be so stringent that it would let many cases go by. In this way it would do a disservice to those families and children who could have benefited from early intervention but will not receive attention since the detection instrument said they didn't need it.
It is therefore of utmost importance that both sensitivity and specificity be known and taken into consideration when choosing a detection instrument for a screening program. Some authors go so far as to recommend that both these levels should be around 80% (Meisels, in press). To achieve optimal levels of both sensitivity and specificity it may be necessary to test different cut-off points before establishing final interpretation criteria.

In some circumstances, several screening instruments are used sequentially in a coordinated effort to progressively identify children and families with problems. This is highly recommendable but requires that appropriate tools be used at each screening level. The first instrument to be applied to the general population should be very simple but highly sensitive, even if not very specific. At this point one wishes to "rule out" healthy cases. Over-referrals will be caught on the next phase of screening. Subsequent instruments should be somewhat more complex and highly specific in order to "rule in" cases who really have the target problems.

In addition to knowing the proportion of cases or of non-cases which are correctly identified, one wants to know what proportion of positive and of negative screen results are accurate. This can be done by calculating the next two indexes: positive and negative predictive value.

**Positive Predictive Value** This refers to the proportion of positive results on the detection instrument that really are cases, i.e., those who do have the problem or risk in question. This proportion, which can be calculated by $TP/(TP + FP)$, will tell us how likely it is that a family or child which has been identified by the screening tool as being at risk, is really at risk.

**Negative Predictive Value** In this case, the measure tells us what proportion of negative results ($FN + TN$) are really healthy ($TN$). It is calculated by $TN/(FN + TN)$.

In relation to both positive and negative predictive values, the prevalence of real cases, i.e., the baseline rate of alterations in the population studied, will dramatically influence the result. Even if the instrument has good sensitivity and specificity, it will have much lower predictive values in an open, unscreened population than in a group which has a higher proportion of real cases. This is extremely important to remember when evaluating the results presented for an instrument. Far too often, the instrument is tested on a clinic population which is easy to obtain but which has a relatively high proportion of real cases. In this context the predictive value is quite high. In light of the favorable results, the instrument is implemented as a primary detection tool in an open population where the baseline rate is low. In this setting the predictive value of the instrument drops to an unacceptably low level.

Considering all the difficulties inherent in developmental screening, it is highly recommended that attempts be made to coordinate the kind of screening instruments to be used according to the baseline of expected problems. At the level of the open population, risk indicators should be the first approach through which many normal children will be selected out of further screening. Next, indirect indicators could be employed to further refine the group in which hands-on evaluations will be used. If well planned and tested, such a system would most likely improve the chances of adequately identifying those children and families in need of attention and not over-referring too many who do not.
**Prediction Accuracy** - This is the proportion of all correctly identified children families, including True Positives and True Negatives, in relation to all the evaluations performed. However, in most instances it is of relatively little utility due to the fact that it combines the instrument's sensitivity and specificity into one global number.

These measures can be used to compare the results of the detection instrument and the external criterion at the same point in time (i.e., concurrent validity) or at some later point in time (i.e., predictive validity). The latter kind of comparison is more difficult to obtain since it implies waiting a reasonable time and following-up the families or children originally tested. This kind of comparison will provide invaluable data concerning the long-term implications of screening. Will those children considered at risk early on really present developmental problems if left without intervention? To what extent are the risk or problems originally observed self-limiting or transitory? A screening program should be directed at those kinds of problems which will not disappear on their own and that really represent enduring conditions that require and can benefit from appropriate interventions.

It may be useful to comment on the use of the word "predictive" since it is being employed in two different ways. In one usage, "predictive value" of a test, the comparison does not necessarily imply prediction over time. Rather it refers to prediction in the statistical sense of agreement between two measures. In the other usage, "predictive validity", we are emphasizing the ability of the instrument to foresee an outcome state over time. Both these usages will be found in reports concerning validity, and one should verify in what way the term "predictive" is being used.

All of these measures are relatively straight-forward and easy to calculate. The difficulty, of course, is obtaining adequate data on which to base them. Although validation studies imply the expense of considerable time and effort, it is essential that they be undertaken. A screening program will be of no use if the instrument used to distinguished between those at-risk and those not at-risk does not do its job well. Considerable damage, in terms of wasted resources and family suffering, will be produced by widespread use of untested, invalid screening instruments. All too often, proponents of an instrument, which they themselves developed or which they have used for a long time and like very much, consider that the instrument "works" without having done the necessary comparisons. The payoff for having tested the validity of the potential screening tool before implementing widespread programs will be well worth the effort.

**The characteristics of the instrument should adapt easily to program needs.** The instruments that are being analyzed here are intended to be useful in community-based programs which usually have limited resources and depend on the efforts of parents and personnel who have little formal training. It is therefore necessary that these instruments possess characteristics that help them fit into such contexts while optimizing their utility.

- The instrument should be appropriate for the program in which it will be used in terms of such considerations as training, materials and time requirements. A relatively complex instrument may function well in a setting which has adequate resources but be useless in a different setting which does not have the necessary infrastructure.
Indications for interpretation of the instrument's results should be clear and easy to calculate and understand for the kind of personnel who will be using the tool and, if possible, for the parents. Specific rules should be established as to what actions should be taken according to the result obtained.

The instrument should be designed so as to optimize its relation to parent education and other programmatic interventions. For example, developmental scales which describe a child's strengths and weakness in different areas can be more easily coordinated with case referrals and preventive orientation than would be the case if only a global score was provided.

Review of Instruments: a Summary

In this section, the characteristics of the instruments reviewed will be described and then they will be compared in terms of the criteria proposed in the previous section.

Measures of Psychosocial Development

In this section we will describe four original instruments, developed in Latin America: the Escala de Desarrollo Integral del Niño de 0 a 6 años (EDIN) from Costa Rica, the Heloise Marinho’s Developmental Scale for Brazilian Children (EDHM), the Escala de Evaluación del Desarrollo Psicomotor (E.E.D.P.) and the Test of Evolución Psicomotora (TEPSI), both from Chile. In the following discussion these same instruments are compared according to the seven criteria established in the previous section.

Comparison of the characteristics of the instruments:

**Age Range and Number of Indicators.** Only the EDIN and the EDHM cover the full 0 to 6 age span and in the latter case the instrument goes even farther. The Chilean instruments together cover the complete age span but the difference in their format makes it difficult to coordinate them as one instrument. The EDIN is by far the most complete of the four instruments, covering more developmental areas with the largest number of indicators. The number of indicators at each age is, however, highly irregular. It is the only scale that includes some reflexes at early ages. In contrast, the EDHM covers the widest age range but includes the next to the lowest number of indicators at each age.

**Age Placement of Indicators.** Two of the scales, EDIN and EEDP, place the behaviors at an age at which 70%-75% of the standardization sample had attained them, a practice more common in detection/screening purposes. On the other hand, the EDHM has them placed at a "typical" age. The TEPSI does not indicate ages for the items included but requires all children to be tested on all indicators. This last approach increases the application time of the instrument and runs the risk of boring the older children with the initial items and stressing the younger ones who repeatedly fail the more difficult behaviors.
**Areas.** All four scales are based exclusively on behavioral indicators, many of which are similar to those found on other infant tests, most notably in the case of the EDIN and the EEDP. The EDIN covers the broadest range of behavioral areas.

**Format, Instructions and Materials.** These four instruments all consist of written score sheets on which the behaviors are listed. They all require training support provided through instruction manuals, which are well constructed. The EDHM is said to require a weeklong training course in order for new personnel to acquire the fundamentals of the philosophy said to underlay the instrument. They also all use simple, easy to produce materials, which are carefully described in the respective instruction manuals.

**Scoring.** They all include specific instructions for calculating the child's developmental level, although the method each one uses is different. In the case of the EDIN, results are only reported as low, adequate and high, while the other instruments provide several alternatives. Among these the EEDP presents two attractive possibilities: a graph and a profile, which greatly facilitate visualization of the results.

**Comparison of the instruments according to the criteria proposed:**

**Specification of Purpose and Congruence of Characteristics.** Although all four instruments propose to provide a general description of child development, they also propose to fulfill additional purposes at the same time, including detection and education. This is especially true of the EDHM, the purpose of which is to "describe development within the context of programs to optimize development". The combination of purposes of the EEDP, description as well as screening, has resulted in the instrument being somewhat of a hybrid: it is simpler than a complete descriptive instrument should be but more complex than most primary screening tools.

With the exception of the EDIN, the low number of items per age reduces the likelihood that these scales would really provide a complete description of the child's functioning. They seem to share characteristics of screening instruments without having all the necessary elements for thorough description/diagnosis.

**Specification of Behavioral Objectives and Appropriateness of Indicators.** Some, but limited, specification of target objectives are found in the case of the EDIN, EDHM, and EEDP. All three mention "developmental delay". The EDIN further specifies central nervous system (CNS) problems and the EEDP makes reference to auditory deficits and motor problems. The TEPSI defines its objectives in terms of delays in the areas included in the test itself.

In the face of this ambiguity concerning the behavioral objectives of these scales it is difficult to evaluate the appropriateness of the indicators. In the case of the EDIN the inclusion of some reflexes may increase the instrument's sensitivity to CNS problems. However, some items have poor interpretability at very early ages. For example, "hands fisted" and "face unexpressive" at 1-2 months. If both have been adequately surpassed, the infant may be developing well or better than the infant who still presents such behavior. Still other items are poorly worded and could be confused with alarm signs if not well understood. For example, "when seated, arches back" could
be confused with opisthotonos; or "when held standing, supports weight on tip-toe" could be passed by a child who presents hypertonicity with scissoring. Such items will probably be poor discriminators in high risk populations, especially when used by para-professionals. Other items are not really behavioral indicators ("exclusive breast-feeding"). In this case the important developmental indicator is whether or not the infant sucks and swallows well.

**Cultural Appropriateness.** These four instruments are all based on evaluations done in Latin America. In three cases these were performed on well-selected standardization samples, while in the fourth instance, the EDHM, the data base was derived from 20 years of practical experience working with children. Nevertheless, many of the test items, especially in the EDIN and EEDP, are essentially the same as those found in traditional baby tests. The EDHM and, to a lesser degree, the TEPSI, have incorporated innovative items which appear to be naturally occurring developmental indicators.

**Reliability and Validity.** Careful research has been performed on all instruments concerning some aspects of reliability. All of them appear to possess adequate psychometric properties. Furthermore, all four instruments have data available concerning their construct validity in terms of age trends and SES comparisons. However, no data are available concerning the concurrent and predictive validity of the instruments. They were designed to provide better or more appropriate descriptions of child development than other available standardized tests, developed in other countries. However, at this point in time we do not know how they compare to other tests and clinical assessments performed at the same age or to subsequent developmental outcome.

**Programmatic Interface.** In all four instruments the programmatic interface seems quite good. They all seem to be designed to fit into practical programmatic settings. They all seem relatively easy to learn and can be applied by intermediate level personnel. The separation of areas in the EDIN is especially attractive for intervention possibilities. While the use of developmental areas is well taken, their internal consistency should be tested, especially since items were moved around freely during the final test construction and the location of some appears highly doubtful. The EEDP provides various useful alternatives for reporting the results, facilitating its coordination with other program components.

**Comparison of the characteristics of six simplified developmental scales:**

**Age Range and Number of Indicators.** With only two exceptions, both from Mexico, these instruments cover the entire age range of 0 to 6 years. The number of items varies, but fluctuates roughly from 1 to 10 per age. Although the Guatemalan version of the shortened EDIN has 12 behaviors per age, only 6 have to be passed by the child in order to receive credit.

**Age Placement of Indicators.** Since these are all detection instruments, the behaviors should be placed so that if a child does not perform them one can be certain he or she has a developmental problem. This criterion has been followed in most of these scales using 70% to 90% of the respective standardization samples as the criterion for age inclusion.
Areas of Development. The EDIN scales and the Graphic Sheet cover the widest range of developmental areas. With the exception of the Scale of Development (Mexico), these instruments have consistently included behaviors from each area at each age.

Format, Instructions and Materials. The format, instructions and materials of all these scales are relatively simple, attractive and appropriate for community-based programs. However, within this simplicity some important variations exist. The Pautas Scale, the Individual Chart and the Graphic Sheet incorporate the use of pictures to help explicate the behaviors to be evaluated. The other instruments that do not use pictures have arranged the behaviors in simple, systematic ways which tend to facilitate a graphic view of the child's development. In the case of the DDST, the newer revised version published by the originators of the scale has greatly simplified the visual appearance of the test sheet by placing the behaviors in chronological order and thereby exemplifying the developmental progression. A further modification was produced in Montevideo, Uruguay by Miguel Matell and collaborators at the Latinamerican Center of Perinatology and Human Development (CLAP). This version, in form of a large laminated poster, presents the DDST behaviors (revised version) accompanied by simple line drawings which illustrate each item.

Materials and training are highly simplified in most of these instruments. The DDST and EDIN have the most complex instruction manuals, while the others tend to rely on general guidelines, pictures within the instrument itself and, in the case of the Individual Chart, a one-week long training course. In three cases, Pautas, Scale of Development and Graphic Sheet, no special materials are required. In the other three cases, the required materials are very simple and easy to produce.

Scoring. All the instruments clearly specify procedures through which the test results should be interpreted in terms of two or three basic categories which distinguish normal development from delayed development or from suspect and abnormal development. In the case of the Individual Chart, these categories are represented by arrows. The horizontal arrow is interpreted as "Your child may not be developing" and the downward arrow as "Your child may need better care". This latter interpretation may generate unnecessary guilt, especially in those cases in which endogenous factors are contributing to the child's developmental difficulty. Furthermore, it is important to note that these arrows do not reflect change over time, a fact which could be confusing to those who are accustomed to the arrows on the "Road to Health" growth chart. In some cases the interpretation rules have not been validated and therefore should be subject to further analysis and modification where necessary. In the case of the DDST, mild delays may be obscured by basing the outcome categories on the number of areas in which delays occur. Especially considering that no internal consistency data have been presented on the structure of the sectors themselves, this procedure may not be the most adequate one to follow.
Comparison of the instruments according to the criteria proposed:

**Specification of Purpose and Congruence of Characteristics.** Directly or indirectly all six instruments are designed as detection instruments. Their characteristics are congruent with this purpose although some are simpler than others.

**Specification of Behavioral Objectives and Appropriateness of Indicators.** All these instruments are designed to identify children with developmental delay or alterations. However, in most cases, little attention has been placed on specifying which developmental problems the instrument is designed to detect. Even in the case of the DDST which was proposed specifically to detect mental retardation (I.Q. < 70), it has been used to detect other problems not originally specified by the authors. The fact that the three shortened versions of the EDIN have chosen different behaviors for inclusion and have even included some of the same ones in different areas is symptomatic of the underlying problem: lack of definition of behavioral objectives.

**Cultural Appropriateness.** Five of the six instruments are based on behavioral observations of Latin American children. The DDST was normed on an unrepresentative sample of children from Denver, Colorado and has not been adapted to Latin America. Some of the behaviors included appear to be inappropriate to the cultural and socioeconomic conditions prevalent in the region. Furthermore, the widespread use of translations of the original score sheets implies the use of age norms which have not been replicated in Latin America. The fact that analyses in the U.S.A. have found differences in age of attainment of 39 of the 105 items in children from different socio-economic levels (Frankenburg, Dick and Carland, 1975), shows the importance of basing the age norms on children from socio-economic and cultural groups similar to those on whom the test will be used.

The other detection scales do not have this problem since they are based on age of attainment of groups of Latin American children from diverse SES groups. Nevertheless, the Individual Chart is the only one which has included behaviors markedly different from the standard infant tests which appear to be culturally appropriate, naturally occurring developmental indicators. This instrument is based on children from around the Rio area and so these behaviors are not necessarily appropriate for Brazilian children from other regions and cultural groups.

**Reliability and Validity.** With the exception of the Individual Chart, surprisingly little attention has been given to testing the reliability of these instruments, especially in the hands of program personnel with limited training in their application. Even less attention has been paid to the concurrent and predictive validity of these instruments. Practitioners seem to accept them at face value without adequately analyzing how well they do what they are supposed to do. In several cases detection instruments have been included in national health care norms without any previous evidence of their reliability in the hands of the health workers who will use them, and without evidence of their ability to detect children with developmental problems.

Experience with the DDST is very revealing concerning the need to adequately test the validity of screening instruments. It is easily the most widely used developmental screening instrument available today. It is also the instrument that has come under the closest scrutiny in terms of
A variety of concurrent and predictive validity studies have been carried out, using diverse external criteria, in different populations. Nine of these studies, including high-risk and unselected populations, from urban and rural settings, of infants and preschoolers, were reviewed and analyzed by Meisels (in press) in order to evaluate their sensitivity, specificity, false negative ratio and predictive value of positive results. Consistently, the DDST proved to be highly accurate in terms of specificity, i.e., its ability to identify healthy individuals. However, it had consistently low sensitivity. In other words, the test under-refers children at risk at a very high rate. Meisels pooled the data from these 3 concurrent and the 6 predictive validity studies and found that 59% and 82%, respectively, of the children with negative screen results, i.e., considered without developmental problems by the DDST, had positive results on the external criterion measures. In light of these results, Meisels concludes that “the DDST in its original and revised English language versions can no longer be recommended for use in mass developmental screening.” Furthermore, he proposes that other more recently developed instruments do exist which could more accurately screen preschool populations for educational problems.

The importance of these data is not so much in terms of whether or not one supports Meisels’ conclusion. Rather, it demonstrates once again the urgency of adequately validating any screening instrument within the context in which it is to be used and in relation to the target problems it is supposed to be able to identify. If the DDST, or any other screening instrument developed out of the region, is to be used in Latin America, these issues of detection accuracy are further compounded by the issues of cultural relevance, making adequate validation studies even more essential.

Programmatic Interface. Most of these instruments are simple and attractive for use in a variety of community-based programs. They are clearly interpretable within the context of primary health care services and preschool educational programs. Ironically, these same features may make them so appealing and apparently appropriate that the necessary validation studies are passed over.

The attractive and simple presentation of the Pautas scale makes it quite motivating for parents and relatively easy to learn and use by members of the health team. The inclusion of pictures to exemplify the behaviors seems to greatly simplify training and facilitate use of the instrument. The use of relatively broad age ranges will limit the usefulness of the scale in contexts where evaluations are performed at the intermediate ages. The Graphic Sheet also has the serious limitation of restricted ages at which the scale can be used (0, 4, 8, 12, 18 and 24 months) which do not correspond to the ages at which primary health care visits are recommended, at least in Mexico.

The original form of application of the DDST also proved unacceptable in many program contexts due to the relatively long time it took to apply the test: 20 to 30 minutes. The simplified version is far shorter, since only the key behaviors are evaluated. The other scales reviewed here are all quite short, taking approximately 5 to 15 minutes to apply.
It is also apparent that some of these scales are being incorporated directly into stimulation programs. To what extent they will be useful in this context is uncertain. While possibly useful for alerting and educating health care personnel to the importance of developmental evaluation, they are probably far less useful for guiding intervention or treatment programs on an individual basis. The separation of areas in some of these scales is useful for interfacing with more detailed educational material, but as it stands the areas are somewhat arbitrarily defined and need to be validated before relying too heavily on them in educational strategies.

The Individual Chart from Brazil is also expressly designed to have educational and motivational value for parents and health personnel. At present, it is individually distributed to each mother along with a separate growth chart and a separate risk index. This would be far too expensive for large scale use in many primary care programs. While the mothers appear to prize the chart highly, it is not clear to what extent it is really understood by them. IPH EM personnel consider that only the pictures and steps are intended to be understood by the mothers, while the fine print and calculations are intended for use by health care workers. Research would, therefore, be quite useful to determine to what extent the Chart is understood by mothers and which aspects could be modified to increase its educational value. Perhaps other elements, such as growth indicators and risk factors, could be substituted for those elements which are less useful for the parents. Such a simplified, more integrated version, once pilot tested and adequately validated, might be more economical and practical for widespread use in community programs.

One of the strengths of the DDST is the way it describes the range of age of acquisition of each behavior, by presenting the age at which 10%, 50%, 75% and 90% of the original sample achieved each behavior. This has great educational value for health and education workers who understand the bars displayed on the sheet. Nevertheless, the 90% age is the only one used for interpreting the test results. In community projects, which employ personnel with minimal formal education, the added complexity created by the bars seems to make the instrument difficult to apply. The revised version of the test sheet has arranged the behaviors in chronological order. This provides a less cluttered page on which it is easier to perceive the natural progression of development. Furthermore, only those behaviors which would qualify as delays need to be tested, greatly reducing the application time.

Similarly, the instruction manual, which is auto-didactic and well-designed, may be too complex for some community based projects. By relying on a detailed instruction manual, the DDST fits in best in settings which involve close supervision of moderately well-educated para-professionals or more highly trained members of the health team. Once again, depending on the context, instrument selection must take into account the ability of the test users to understand the material. The addition of pictures, as in the Uruguayan modification, may help to make the items easier to understand without the use of a lengthy instruction manual and therefore make the scale more appropriate for programs with more limited resources.

During the last few years, many attempts have been made to include developmental indicators on the widely used growth charts. Some early alternatives were reviewed in Atkin (et al. 1987). Here
we will analyze nine recent Growth and Developmental Charts which were available for comparison by the author. They represent innovative alternatives proposed by Argentina, Brazil, Colombia, Ecuador, Guatemala, Honduras, Mexico, Peru and Venezuela. Other alternatives probably exist and, in fact, competing versions may coexist within one country.

Although it is true that some of these charts are initially accompanied by supplementary materials, they will be analyzed here as they stand by themselves. This is probably the way they will be used in most contexts, both due to scarcity of resources for distributing additional printed material as well as to the common tendency on the part of many people of not reading such supplements. The idea of incorporating behavioral indicators into the growth charts implies that they should be able to be interpreted with as much ease as are the other health messages included, requiring only minimal training on the part of the health team personnel.

**Comparison of the characteristics of the instruments:**

**Age Range, Number and Area of Indicators.** All nine instruments cover the first 60 months of life, and in a few cases go up to 72 or 84 months. They are varied in the number of items, ranging from a low of 13 in the chart from Guatemala to a high of 62 in the one from Honduras. In all cases, behaviors have been included from several areas of development, most typically large and fine motor, language, cognitive and social. The Honduran chart includes the startle and sucking reflexes in the first three months and the Mexican version includes sucking (newborn).

**Format and Interpretation.** The behaviors appear in lists placed in different locations on the charts. In the cases of Ecuador and Guatemala, 1 or 2 behaviors appear at each age, while the others include more, ranging from 3 to 7 per age. With the exception of the Honduran chart, which specifies a precise age range, all the charts indicate one age at which the child is expected to have achieved the behavior(s) in question. Presumably, if the child achieves the behavior(s) by the age or before the next age, he or she will be given credit for it. All the charts provide a space for indicating whether or not the child performs the behavior. In all of them, except in the charts from Ecuador, Mexico and Honduras, a specific indication or space is included for registering the age of attainment. In the exceptions mentioned, the spaces available could be also used for this purpose although no specific indication as such is included.

Only in three cases, Argentina, Brazil and Peru, are pictures included to facilitate the comprehension of the developmental indicators. In the chart from Argentina, one drawing is included per age. The same drawing is placed beneath the age on the growth grid and also next to the corresponding list of behaviors along the right hand side of the same page. A similar arrangement is used in the Peruvian chart. Nine pictures are placed along the lower border of the growth chart. Their position roughly corresponds to the ages at which the behavior should be attained. On the other side of the chart, a more complete list of behaviors is found, with one picture per age. However, in this section only eight pictures are included, three of which are different from the ones on the growth chart side. Two different pictures are used to show the child sitting. The one on the growth chart side illustrates the six month behavior of "supports self in sitting position". On the other side of the chart, the picture for six months shows a behavior
listed at nine months “sits without help”. Furthermore, the picture placed at 3 to 4 years on the front side does not correspond to any of the items in the 3 or 4 year lists of behaviors.

The Brazilian chart includes the behaviors in a separate list, three per age. A picture illustrates one of the behaviors included at the corresponding age. This chart is the only one which includes specific indications for the interpretation of results and for subsequent actions, in terms of further exploration, stimulation activities and referral to more specialized services. An interesting chart from Uruguay (produced by the Centro Latino Americano de Perinatologia y Desarrollo Humano-CLAP) came to the attention of the author after the completion of this paper. In this alternative the behaviors are presented through pictures superimposed on the growth chart itself. In addition, other pictures describe activities related to nutrition, health and prevention of accidents.

**Inclusion of Risk Indicators:** In addition to the intimate relation with growth indicators, these charts provide an excellent opportunity to combine the behavioral evaluation with the identification of risk factors. All the charts include registration of vaccinations and appointments, which provide a minimum of information concerning regularity of health care. However, only three include additional risk factors. The Guatemala chart includes a section concerning perinatal risk: pregnancy and birth abnormality, care provided at the birth, and Apgar and gestational age of the newborn, as well as other complications. The Mexican chart includes a space next to the developmental indicators in which the child's nutritional status is to be indicated and, during the first two years, whether or not the infant is being breast-fed.

The Colombian chart is exceptional in that it includes two lists of risk factors, one corresponding to the child and his or her siblings and the other to the environment. In both cases, squares are provided to indicate the presence of each risk factor and any additional one noted. In the first list, they have included birth weight, twin birth, lack of breast-feeding, malnutrition and/or death among siblings as well as the number and birth spacing. With respect to the environment, basic conditions are included: water supply, disposal of excreta and garbage, and presence of pests.

**Comparison of the instruments according to the proposed criteria:**

**Specification of Purpose and Congruence of Characteristics.** All of these instruments are ostensibly designed to be used in primary health care settings in order to facilitate early detection of children who are not developing well. They are necessarily far simpler than most of the scales already described since they must incorporate a few select indicators into a chart which contains a great deal of other information. It is precisely this integration with growth monitoring that makes these charts so appealing, but at the same time imposes severe restrictions on the number and kind of behavioral indicators which can be included.

It should be mentioned from the outset that some of these instruments were designed for the more general purpose of educating the public and health team in the importance of observing children's development. Nevertheless, this broad educational focus also implies the use of the chart as a detection instrument since such observations are presumably to be the basis of actions which foster better development if problems are found.
These "carnets" are designed to be given to the mother and to be presented by her each time her child is seen by the health team. The content of the "carnet" may therefore be directed at two audiences: both the mother and the health care worker. The information included should be carefully planned so as to be appropriate for the target audience. Some information will be directed at the mother while other data will be geared toward the health personnel.

These instruments are all simple and attractive, making them appropriate for the task at hand. However, their content is not optimal for early detection of high-risk children in open populations where they will presumably be used. For such a purpose they should be far more selective in terms of behavioral indicators and rely more heavily on risk factors in order to identify children in need of closer evaluation. Furthermore, clear indications should be provided concerning the interpretation of results and subsequent actions.

**Specification of Behavioral Objectives and Appropriateness of Indicators.** In most cases it is unclear what the possible developmental objectives are for these instruments. Considering the small number of indicators which can be included, it is of the utmost importance that indicators be well selected: they should be the most highly predictive behaviors in relation to the kind of developmental disabilities they are designed to detect. Without further information concerning the target objectives of each instrument, it is difficult to decide whether or not the behaviors included are the most adequate.

Taking all the behaviors together that are included in the different charts, it becomes apparent that there is considerable diversity in the specific indicators selected. According to an analysis of eleven growth charts done by Martha Llanos, UNICEF/Colombia, no single behavior was included in all the charts examined. "Smiles" was the single most common behavior, found in all except two of the charts. The second most frequent indicator, "recognizes people, reacts to strangers", was found in all except three charts. All the other behaviors were less consistently found in the charts examined. It is however important to consider the equivalency of similar indicators in terms of the kind of behavior to which they refer. Different authors may choose somewhat different steps within a developmental sequence or slightly different ways of evaluating a similar ability. A re-analysis was done using the nine charts available to the author, all of which were included in the Llanos analysis also. The charts were composed by grouping the behavioral indicators according to related abilities, in order of frequency of inclusion of the groups. In this way the chronological order is violated, but it is far easier to analyze the relative consistency or inconsistency of the indicators included.

All the charts include one or more indicator concerning the child's ability to stand up and walk, the most frequent one being "walks and runs" (7/9 charts). Early language production is also included in all charts, with 7/9 using some form of "says mama and papa". Finally, all the charts include one to three references to the child's ability to dress him or herself, most often "helps dress self" and "dresses alone" (6/9 charts each).

A rea included frequently, that is, in six to eight of the charts, include babbling and imitation of sounds (6/9), early phrases (7/9), names self (7/9), response to sounds (8/9), smiles (8/9), head
control (8/9), ability to sit up (7/9), jumping ability (8/9), reaction to strangers (6/9), ability to grasp objects (with one hand and with both hands) (8/9), and initial graphic skills (7/9).

In only half of the charts, i.e., four or five, indicators were included which correspond to ability to insert and take out objects from a receptacle (5/9), toilet control (5/9), play with other children (4/9), use of words to express wishes (5/9), recognition of body parts (5/9), ability to count using one's fingers (5/9) and identification of own sex (4/9). All other behaviors included were found in only one to three charts.

Summarizing then, one hundred and twenty four behaviors were included in the nine growth and development charts reviewed in this analysis. Only 52 were relatively idiosyncratic, i.e., included in three or fewer of the charts. All the others can be seen as variations of certain abilities or different steps in developmental sequences. In this light, the original heterogeneity is greatly reduced. It becomes apparent that the charts actually show considerable consistency in the kind of abilities they have included.

This is not to say, however, that the more frequently included items are necessarily the best indicators. Surprisingly, some of the clearest signs of developmental problems, for example, lack of visual fixation and following or head lag when pulled to sit, were included very infrequently. A gain, part of the problem arises from the lack of definition of the behavioral objects of these instruments. Especially when using such simplified screening tools, it is essential to define very clearly what kind of developmental problem can be detected and at what ages, as well as which ones are of particular interest to the health system in question, for example, which can be adequately dealt with if found.

Cultural Appropriateness, Reliability, Validity and Programmatic Interface. It appears that many of the behaviors included in these charts were taken from standardized infant tests without careful attention to their cultural relevance or normativity. At present, we are unaware of any data as to how well these charts are understood by parents and health personnel or in what way of developmental indicators are actually being used. (In one case, informal comments were made to the author that the behaviors on the chart were generally ignored!) Such information would be essential as first steps in evaluating how useful such instruments really are or could be. Of course, data are also needed concerning their concurrent and predictive validity, if in fact they are being used as screening instruments. One hopes that the initial effort to produce such attractive and innovative instruments will be translated into efforts to follow-up their use and analyze their strengths and weaknesses, in order to optimize their effectiveness within the context of primary health care.

DETECTION/SCREENING INSTRUMENTS-QUESTIONNAIRES USING INDIRECT BEHAVIORAL INDICATORS

In this section we will describe two simple questionnaires which are designed for screening purposes and use indirect means to find out if the child has developmental problems. Since these instruments are so simple their characteristics will be described directly rather than presented in table form.
The Ten Questions screening interview from the International Study of Severe Childhood Disability. This instrument was designed by Drs. Clarke, Bailment, Narayanan and Sell as part of a pilot study coordinated by personnel from Columbia University in New York. Its purpose is to detect children from three to nine years of age who have severe developmental disabilities, including severe mental retardation and psychiatric disorders, cerebral palsy and other motor disorders, blindness, deafness and epilepsy. The study was carried out in nine Third World countries, the only Latin American one being Brazil. This instrument was applied within the context of a validation study which included the use of another more detailed detection questionnaire and independent evaluations performed by local experts on all positive screens as well as on a random sample of negatives. The final report of the project (Belmont 1984) found that the Ten Question Interview (TQ) had better sensitivity than the more complete interview in most study sites. Although the target objective of the instrument is severe mental retardation, the TQ was found to also identify children with other developmental problems. In the Brazilian site, 132 of the 156 children with positive expert assessment were identified by the questionnaire screen (sensitivity = .84). The specificity was however low (= .32), showing a marked tendency to over-identification. In other sites the results were more favorable, suggesting that further trials should be done in Latin America.

This brief interview consists of four questions concerning vision, audition, movement and convulsions, and six questions concerning the cognitive development of the child. The questions are listed below.

1. Compared with other children, did the child have any serious delay in sitting, standing or walking?
2. Does the child have difficulty seeing?
3. Does the child appear to have difficulty hearing?
4. When you tell the child to do something, does he seem to understand what you are saying?
5. Does the child have weakness and/or stiffness in the limbs and/or difficulty in walking or moving his arms?
6. Does the child sometimes have fits, become rigid, or lose consciousness?
7. Does the child learn to do things like other children his age?
8. Does the child speak at all (can he make himself understood in words; can he say any recognizable words)?
9. Is the child's speech in any way different from normal (clear enough to be understood by people other than his immediate family)?
10. Compared with other children his age, does the child appear in any way backward, dull or slow?

A child is considered a "case" if any problem is detected during the interview. Any child with positive answers to any question other than numbers 4, 7 and 8, (in which negative answers are indicators) is considered as possibly having serious developmental problems.

This instrument is an interesting example of how very simple questions may be used to screen certain populations for specific developmental problems. The pilot study in which it was tested is also a model for instrument validation within the context of community programs. Hopefully, other instruments could be validated in similar studies.
Gensini-Gavito Questionnaire (Encuesta), Cali, Colombia. This brief twelve item questionnaire was developed by the authors of the "Pautas" scale previously described. It is applicable to children from 0 to 6 years of age and is designed as a first level screening instrument in open populations. Children who evidence risk on the questionnaire are then referred for evaluation by the "Pautas" scale.

The questions included in this instrument are as follows:

1. Were there any complications during the pregnancy of this child?
2. Were there any problems during the delivery?
3. Has the child ever had a fit or convulsion?
4. When the child was little, did he or she fall or suffer any hard blows to the head?
5. Are the weight and height of the child below normal for his or her age?
6. Does the child have problems eating and sleeping?
7. Does the child cry a lot and have a bad temper, or is he or she difficult to control?
8. Is the child highly distractable, very active, and doesn't obey anyone?
9. Does he or she fall frequently or do things drop from his or her hands frequently?
10. Does the child talk very little and/or use sign language to ask for things?
11. Is the child too calm, doesn't make noise and is very serious?
12. Does the child speak unclearly and is it difficult to understand him?

Only the first four questions are applied in the case of infants six months old or less, the first seven if the child is between six and twelve months old, and all the questions if older. If the child presents two or more affirmative answers, he or she is considered a positive case, who should be further evaluated. A study of children evaluated at public health clinics in Cali, Colombia (Gavito and Gensini 1979), found that above this cut-off point, all the children evaluated with more detailed instruments presented some form of developmental problem, including motor alterations, laterality problems, behavior and affective problems, and absence or alteration in language. Unfortunately, only about a third of the children with positive results and none with negative results were subsequently evaluated, thereby greatly limiting the usefulness of the results. Both this and the TQ are good examples of very simple first level screening instruments which, if adequately validated, could be very useful in a coordinated detection system.

Measures of the Childrearing and Family Environment

In comparison to the relatively large number of instruments now available which use behavioral indicators to measure child development, those which are based on environmental variables are quite scarce. Although considerable consensus exists as to the basic environmental dimensions which facilitate child development, the wide variations in the way these generic elements are manifested in diverse cultural and ecological settings make it difficult to create instruments which are appropriate in diverse settings.

In light of the recognition of the great importance of the childrearing environment and of the difficulty of selecting relevant indicators on an a priori basis, the World Health Organization (1985) included a family interview within its multi-country initiative to develop culturally appropriate optimal and risk factors related to physical growth and psychosocial development.
The guidelines they propose, to be modified and supplemented according to local conditions, are useful as a tentative framework for developing appropriate instruments. Although the list is not structured as an instrument, we have included the categories proposed, hoping they may be useful in discussions of the other instruments we will describe in this section.

The following areas are included in the WHO interview:

- **Shelter**: structure of the dwelling and amenities available to the family, organization of the home.

- **Parental and family data**: household composition, parents' religion and ethnic group or other relevant groupings, parents' educational level and occupation, parents' and siblings' health history, child's health history including perinatal difficulties, resources available to the family, access to outside world.

- **Data on the child**: demographic information, position of the child in the family, birth history, health problems, history of breast-feeding, early mother-child contact, caretaking, provision of appropriate play opportunities, experiences outside the household, child's work, caretaker's involvement with child (including responsiveness and discipline), aspirations of the caretaker for self and the child, and the caretaker's perceptions of causality in relation to the child's health, well-being and happiness.

Data obtained by the family interviews adapted in each country will be analyzed in relation to the child's growth and development, measured by locally adapted indicators. It is hoped that in this way, culturally appropriate measures will be derived, some of which are related to optimal outcomes and others to sub-optimal growth and development.

In this section we will describe three instruments which are being used in Latin America in a variety of contexts. The first two are somewhat more complex: the HOME by Caldwell and Bradley and the Profile of Maternal Behavior by Bayley. The third instrument is somewhat simpler and intended specifically as a screening tool: the Questionnaire for Evaluating the Family Environment of the Child under Two by Lira.

### Characteristics of the Instruments

#### Age Span and Number of Items

The three environmental instruments are quite different from each other in structure and content. The two parts of the HOME (0 to 3 and 3 to 6) taken together and the Profile of Maternal Behavior are applicable over the entire age span of 0 to 6. However the Questionnaire is limited to the first two years of age. The number of items varies enormously: even the two versions of the HOME (U.S. and Mexico) differ considerably. In terms of number of outcome scores the Questionnaire is the simplest, providing three subtotals.

#### Areas

All three instruments overlap to some degree with respect to the conceptual areas included. They all consider some aspects of the affective interrelation and specific developmentally appropriate stimulation. However, in many other ways they differ considerably. The HOME covers the widest range of conceptual areas but is limited to the quantity and quality of stimulation provided in the child's home. The questionnaire covers socioemotional and...
cognitive stimulation and places considerable emphasis on the father's involvement with the child and with the mother. The Profile of Maternal Behavior is restricted to the behavior observed in the testing situation which reveals qualitative aspects of the mother-child relation but does not speak directly to the broader stimulation issues.

**Format, Instructions, and Scoring.** Both the HOME and the Profile of Maternal Behavior are based on direct observations of the mother interacting with the child. This is a very useful technique but requires some degree of specialized training, particularly in the case of the Profile which is relatively difficult to learn due to the lack of operational definitions. The complexity of the nine-point scales also makes this the most difficult instrument to learn. The Questionnaire is far simpler, using a straightforward question-answer approach with dichotomous responses which is very easy to apply.

**Specification of Purpose and Congruence of Characteristics.** The HOME and the Profile were designed primarily as descriptive instruments, although the HOME has been proposed in Mexico for use in primary detection. In their present formats, they are too complex for such a purpose, but could feasibly be simplified through appropriate research.

The Questionnaire was designed specifically to screen low SES families to identify those at greatest risk of having their children become developmentally delayed. It is far simpler than the other two and more appropriate for the task.

**Specification of Objectives and Appropriateness of Indicators.** All three instruments clearly define the conceptual areas they were designed to measure. However, during the construction of the HOME and the Questionnaire, factor analytic methods were used to determine the inclusion of items and the final groupings. These differ in some ways from the initial areas proposed, most notably in the case of the Questionnaire, which began by including 226 items in eleven areas and ended up with 50 questions grouped in three sub-scales.

**Cultural Appropriateness.** The HOME is based on well-defined conceptual areas which probably have broad application across cultures. However, the way each area is defined operationally reflects culturally biased content which seems to be inappropriate in many contexts in Latin America. Even in the U.S., the scale has come under attack as reflecting middle class child rearing practices to an unnecessary degree. As has been done in the Mexican version, adaptations to local conditions are essential for this instrument to be appropriate for use in Latin America. Even then, some authors have found it to be culturally inappropriate.

The Profile itself is actually relatively culture-free. The operational definitions of how each scale is to be defined can be adapted to the context in which it will be used. However, the situation in which the observations are made, i.e., a developmental test of the child, will have different meanings and implications for behavior in different cultural contexts.

The Questionnaire is the most culturally appropriate of these three instruments. However, the emphasis placed on father involvement may limit its usefulness in situations where single
motherhood is common and where one wishes to distinguish the quality of childrearing environments within groups of single mothers.

Reliability and Validity. The HOME (U.S. version) and the Questionnaire were developed using appropriate factor analytic procedures and have been demonstrated to have adequate psychometric properties. However, the resulting sub-scales of the Questionnaire are highly unbalanced in terms of number of negative and positive items.

All three instruments can be reliably applied, although the training required to achieve adequate levels for the Profile is far longer and more difficult than most community-based programs can afford. Some evidence exists to suggest that both the HOME and the Profile may predict risk of severe malnutrition. Further research is needed to test whether these instruments, or simplified versions of them, could actually have practical predictive value for important health and developmental problems in screened or unscreened populations. The author of the Questionnaire has proposed such studies using the instrument she has developed within the context of health and education programs in Chile.

Programmatic Interface. The Questionnaire is a simple tool which could easily be used in a wide variety of community-based programs. However, both the HOME and the Profile require a high level of resources which make them most appropriate for small, intensive secondary or tertiary level programs. Simplified versions might be useful in larger programs. However, the need for home visits and the training required for behavioral observations necessarily limit their use to settings with adequate resources. Research with these instruments could provide important information concerning culturally relevant environmental indicators which could be incorporated into simpler instruments.

Measures of Bio-psycho-social Risk Relevant to Child Behavior and Home Environments

In this section we will describe and comment on three kinds of simple instruments which could function, in certain contexts, as primary screening tools. They are, however, so different from each other that it would not be appropriate to make direct comparisons. Their unique characteristics have important implications for the kind of tools that might be useful to identify groups of children and/or families in need of further evaluation.

**CHECKLIST FOR SIGNS OF HIGH NUTRITIONAL RISK (ZEITLIN AND JOHNSON) MEXICO**

Based on the findings of a research project carried out in Mexico, near Cuernavaca, Zeitlin and Johnson (no date) proposed a list of ten psychosocial factors that could be developed into a detection instrument. Due to the fact so clearly reviewed by Zeitlin and Mansour (1986), that the same environmental risks that are associated with malnutrition are also associated with poor psychosocial development, such a risk checklist could also be relevant in relation to developmental problems. The initial checklist, proposed to pre-testing only, includes the following positive or neutral factors, from which risk indicators could be derived:

1. Absence of feeding disorder: Does your child have a good appetite? Do you think he eats more/the same/less than other children do?
2. Intimate support of primary caretaker: Who lives with you? Do you have someone who can help you with the baby?
3. Preferential bonding to infant (attachment): Do you have plans for your children’s future? What are they? (Ask for each child in the family and compare differences in response. Those for whom the mother has no specific plans could be non-preferred.)
4. Integrated personality: Do you find others generally friendly and helpful/unfriendly and unhelpful to you?
5. Maternal attentiveness: How does your baby show he is hungry? What do you do when your baby cries?
6. Consistent good health: Has your baby been ill within the last month? Two months?, etc.
7. Enterprising versus fatalistic attitudes (or upward mobility): What are your plans for the future? What do you want to be doing in the next five years?
8. Neighborhood support: Where do you live? Do you like your neighborhood? Do you like your neighbors? Are they friendly and helpful to you?
9. Maternal stores: (no questions specified)
10. Sanitation: Does your home have a latrine or bathroom? Does your family use one cup for dipping and drinking water? How do you wash your family's dishes and eating utensils? What type of flooring does your home have? Where does your family get its drinking water? Do animals come inside or live inside your home? Which ones?

Although this checklist cannot yet be considered an instrument, it is a very promising initiative based on research performed in Latin America. Its simplicity and integration of different domains are particularly appealing. One hopes that further research will identify the most relevant questions, develop an instrument, and test its validity appropriately.

TWO PSYCHIATRIC SCREENING QUESTIONNAIRES: THE GENERAL HEALTH QUESTIONNAIRE (GHQ) BY GOLDBERG AND THE SELF REPORT QUESTIONNAIRE (SRQ) BY HARDING

These scales are good examples of simple, well constructed and adequately tested primary detection instruments (McDowell and Newell, 1987). To the extent that a program has decided that a caregiver’s psychiatric problems would be an important risk factor to consider, instruments such as these might prove useful as preliminary screening tools. Some initial evidence concerning the relationship between mental health problems and the quality of childrearing and child health is suggested by a master's thesis supervised by the author, in which high scores on one of these scales, the GHQ-20, were associated with malnutrition and obesity in children living in a poor area of Mexico City (Gomez 1989).

These two instruments are quite similar. Both were designed to detect adults with non-psychotic psychiatric disturbances in primary care settings and in community programs. Both have been translated into Spanish and used in Latin America. Various versions of the GHQ have been tested and validated in Mexico by personnel from the Mexican Psychiatric Institute. The SRQ has been standardized in Brazil and is being used in conjunction with child developmental research and in primary health care services.

The GHQ was designed by Goldberg in England. Items were developed according to a clear conceptual framework and have been subjected to a wide variety of validation procedures, all of
which have shown favorable results. Five versions exist of the GHQ, each with a different number of items: the original GHQ-60, and shorter versions: GHQ-30, GHQ-28, GHQ-20, and GHQ-12. These were derived from an original pool of 140 items which were selected to cover the following conceptually defined areas: unhappiness, anxiety, social inadequacy, and hypochondriasis. The initial scale was applied to samples of normal people, severely ill and mildly ill psychiatric patients. Scaling techniques were applied to identify adequate items and various analyses were performed to evaluate the resulting scale's ability to discriminate between the groups of subjects. The 28 item version was developed through factor analytic techniques and is structured into four subsets of questions: somatic symptoms, anxiety and insomnia, social dysfunction and severe depression. The complete scale and the shorter versions have been used in a variety of countries, consistently showing good predictive validity. For example, Goldberg reports 96% sensitivity and 88% specificity for the GHQ-60 (McDowell and Newell, 1987). A study in Brazil, using the GHQ-12 in four different communities, found that sensitivity ranged from 82% to 89%; specificity, from 69% to 84%; with positive predictive values of 80% to 83% and negative predictive values of 71% to 90% (Mari and Williams 1985). In Mexico, studies have validated the scales and found that the GHQ-30 is the most adequate version in terms of its ability to discriminate between healthy and mentally ill subjects (Tarnopolovsky et al. 1979).

The SRQ was designed by Harding to study mental illness in developing countries. It is based on items from previous psychiatric screening instruments, including the GHQ. Two versions exist of the SRQ. One contains 24 items, 4 of which screen for psychotic symptoms and 20 for non-psychotic morbidity. The SRQ-20 corresponds to the latter items. The SRQ-24 has been translated into eight languages, including Portuguese, for use in the "Collaborative Study on Strategies for Extending Mental Health Care in Developing Countries". The SRQ covers psychosomatic symptoms.

The GHQ is phrased in such a way that it is evaluating recent psychiatric problems and is not sensitive to long-term chronic disorders. The SRQ, on the other hand, does not specify a timeframe for the symptoms evaluated. Other than this difference, the shorter versions of the GHQ and the SRQ are quite similar. The GHQ uses a four category response format (0 = not at all, 1 = same as usual, 2 = rather more than usual, and 3 = much more than usual). However, an alternative scoring system which essentially results in a "yes" or "no" format has been tested and shown to be just as good as the more complex four-category alternative. The SRQ uses a "yes" or "no" format. Cut-off points have been established for each version of both scales: between 3 and 4 symptomatic responses on the GHQ-12 and between 7 and 8 on the SRQ-20.

We have listed the questions included in the shortest versions of each scale.

**GHQ-12**

Have you recently:
- been able to concentrate on whatever you are doing?
- lost much sleep over worry?
- felt that you are playing a useful part of things?
felt capable of making decisions about things?
- felt constantly under strain?
- felt that you could not overcome your difficulties?
- been able to enjoy your normal day-to-day activities?
- been able to face up to your problems?
- been feeling unhappy and depressed?
- been losing confidence in your self?
- been thinking of yourself as a worthless person?
- been feeling reasonably happy, all things considered?

SRQ-20

- Do you often have headaches?
- Is your appetite poor?
- Do you sleep badly?
- Are you easily frightened?
- Do your hands shake?
- Do you feel nervous, tense or worried?
- Is your digestion poor?
- Do you have trouble thinking clearly?
- Do you feel unhappy?
- Do you cry more than usual?
- Do you find it difficult to enjoy your daily activities?
- Do you find it difficult to make decisions?
- Is your daily life suffering?
- Are you unable to play a useful part in life?
- Have you lost interest in things?
- Do you feel you are a worthless person?
- Has the thought of ending your life been in your mind?
- Do you feel tired all the time?
- Do you have uncomfortable feelings in your stomach?
- Are you easily tired?

RISK INDEXES, PROGRAM OF MATERNAL-INFANT ATTENTION (PAMI), BOGOTA, COLOMBIA

As part of a longitudinal follow-up program of high-risk infants, personnel from the PAMI program ("Programa de Atencion Materno-Infantil") at the Javeriana University in Bogota developed a series of risk indexes for use during pregnancy, at 48 hours post-partum, at the end of the puerperium, at 4 weeks, 28 weeks, 1 and 2 years, and 3 and 4 years of age. Biological and socio-demographic risks are combined with indicators of possible problems in the childrearing environment. Relative weights have been assigned to different responses and the total number of points is registered on the same sheet on which the indicators are listed. Cut-off points are indicated, distinguishing between low and high total scores. Unfortunately, data that were collected using these indexes in the context of the longitudinal study, and which could provide some validation of the instruments, have not been analyzed. However, we have listed the
environmental risk indicators included in the post natal evaluations, with the number of points assigned in parentheses, since they appear to be interesting alternatives, based on research in Latin America.

- Acceptance of the child: unwanted pregnancy (7); attempt to abort (12).
- Family relations: disharmonic (3); dissolution due to separation (3); dissolution due to death (2).
- Child care: by family members other than the mother (2); by non-family members (3); day care (3).
- Maternal-child relationship: indifferent (8); aggressive (10) Family data: paternal consanguinity (2); father (5), mother (10) or both parents (20) drug addicts or alcoholics; mental illness in father (2) or mother (3); sibling death under one year of age (2); sibling malnutrition (2); four or more persons sleep with the child (2).

This series of risk indexes is an interesting example of the way biological, social and psychological factors can be combined into primary screening instruments. It is unfortunate that research with this instrument could not continue far enough to evaluate its predictive validity, for which data were potentially available through the longitudinal study in which it was applied. This is one of several cases in which unanalyzed data exist which could ideally be used to help develop the culturally relevant and valid measures so urgently needed in Latin America.

Conclusions and Recommendations

Interest has spread throughout Latin America in creating and using culturally appropriate instruments for measuring psychosocial development of children from 0 to 6 years of age. Several innovative and creative alternatives have been developed in the region which provide possibilities both for the purpose of description/diagnosis and for detection/screening and monitoring. Some of these may also be useful for programme evaluation, depending, of course, on the objectives of each specific project. To a lesser degree, awareness has also grown concerning the need for instruments to measure developmentally significant environmental characteristics as well as bio-psycho-social risks associated with developmental disadvantage.

Such interest needs to be translated into effective actions on behalf of children who are in need of preventive interventions. It is therefore important that a concerted effort be made to optimize the characteristics of the instruments and their insertion into well-designed programs.

The comparison we have made of some of the available instruments generates the following conclusions and recommendations, which we have organized along the lines of the criteria proposed for evaluating instruments.

Specification of Purpose and Congruence of Characteristics

- Conclusions
Too little attention has been given to specifying the instrument's purpose and to insuring congruence with the instrument's characteristics. In many cases, lack of specification, or the ambitious combination of multiple purposes has created instruments which do not exactly fit the needs of any particular purpose. Some descriptive diagnostic instruments are too simple; detection/screening tools too complex. Time-consuming hands-on evaluations may be proposed for use in large populations where more indirect approaches might be more effective.

**RECOMMENDATIONS**

- Careful attention should be paid to specifying the purpose of proposed instruments in relation to the kind of population in which will be used, and to assuring that their characteristics are developed or chosen accordingly.

- It is preferable that one clear purpose be proposed for a particular instrument and fulfilled adequately rather than trying to accomplish too many things with one instrument.

- Greater attention should be given to the incorporation of well-selected risk factors and environmental measures within the instruments.

- Increased efforts should be made to coordinate instruments with different purposes in order to create an integrated system of detection, diagnosis and intervention. Biological, psychological and social risk factors, as well as indirect environmental and behavioral indicators could be used in open populations to reduce the population on which direct behavioral "hands-on" assessments would be performed. Subsequent from there, identified cases would be evaluated with more detailed assessment using appropriate descriptive/diagnostic tools. These successive phases need to be coordinated and tested within the context in which they will be implemented.

**Specification of Objectives and Appropriateness of Indicators**

**CONCLUSIONS**

Far too frequently it is unclear what the target of the instrument is in terms of the behavioral objectives. Partially as a consequence of this imprecision, the behavioral indicators are inappropriate for the task for which the instrument is being used. Even in those cases in which objectives have been clearly defined, practitioners may use the instrument with other objectives in mind. In still other cases, instruments appear to be directed at behavioral difficulties which cannot be identified at the ages which the instrument encompasses nor with the kind of indicators included.

**RECOMMENDATIONS**

- Instruments should clearly specify which aspects of development they are designed to describe and/or detect, both during their development and when implemented.

- The developmental areas or problems should be chosen according to the program's priorities and to the feasibility of describing and/or detecting specific developmental difficulties at the ages involved and with the available resources (both in terms of capacity to detect problems and in terms of capacity to provide intervention).
Efforts to identify the most salient and frequent objectives which should be the target of specific instruments would be useful. Perhaps a coordinated effort could be made to identify such priorities throughout the region and to develop, select and test instruments directed at such objectives. Some possible common objectives (in general terms) might be the identification of neonates and/or infants at risk of certain kinds of developmental deficits; pre-school screening for potential learning and academic difficulties; identification of developmental problems in preschool and school age children within the context of large epidemiological studies (in order to assess which developmental problems are prevalent in a given population); and the description of conceptually defined developmental processes (for example, self-esteem, attention, problem solving, motivation, etc.) in such a way as to be applicable in community-based projects (especially for use in program evaluation).

Behavioral indicators not included in the traditional infant tests should be considered where relevant. Attempts should be made to incorporate indicators derived from recent research findings on resiliency/vulnerability factors within the child and the family environment, as well as on precursors of later cognitive development.

More attention should be paid to developing instruments which deal with important environmental processes related to developmental outcome. These should be based on local childrearing practices and relevant factors which place families under stress. It is especially important to develop measures which capture environmental characteristics or processes which increase or decrease a family's vulnerability to stress and risk of presenting childrearing deficiencies.

Cultural Appropriateness

CONCLUSIONS

Wide spread concern for the cultural relevance of measurement instruments has stimulated a remarkable amount of creative, innovative work in the region. In most cases this has been translated into original and attractive formats for the instruments and the obtaining of appropriate age norms for behaviors derived from other infant tests. Less attention has been given to the identification of new more culturally appropriate indicators along the lines of the WHO initiative. To what extent such "natural indicators" could facilitate the educational impact of the instruments and at the same time increase their concurrent and predictive validity remains to be seen and should be the topic of careful analysis.

RECOMMENDATIONS

- More attention should be given to the identification and validation of culturally appropriate indicators and their appropriate incorporation into specific instruments.
- The words and/or pictures used to describe behaviors should be carefully planned and tested to optimize comprehension on the part of parents and health care personnel. Test jargon should be avoided wherever possible. Pictures should be used instead of instruction manuals where feasible and appropriate.
Practical research should be undertaken to evaluate how well the different components of the instrument are understood by the population with which it will be used. We should not take for granted that formats we find attractive are necessarily understood or relevant to population groups with different cultural backgrounds.

**Reliability and Validity**

**CONCLUSIONS**

This is the area in which the major limitations and urgent needs are found. Although more attention has been given to inter-observer reliability issues, continued efforts are needed to assure adequate reliability in the hands of the people who actually are using the instruments.

In some cases the internal consistency of developmental areas proposed as separate in some scales has not been sufficiently evaluated.

Very little attention has been paid to concurrent and/or predictive validity of the instruments. This is particularly important in the case of the detectional screening instruments. Even with some of the most attractive instruments, we do not know whether or not they identify children or families who really have problems, or to what extent they pass over children or families in need of help.

**RECOMMENDATIONS**

- Reliability and validity studies should be carried out before instruments are implemented on a large scale.
- Modifications should be made according to the results of such studies in order to optimize the reliability and validity of the instruments.
- A coordinated effort could be made by knowledgeable researchers in the field to develop practical models for validation studies which optimize resources. Would it be possible to compare several instruments which have similar purposes and objectives to the results of a single set of independent expert assessments of developmental status? To what extent are data already available from clinical settings which could facilitate the necessary validation comparisons?

**Programmatic Interface**

**CONCLUSIONS**

The remarkable efforts which have been made to develop simple measurement alternatives have shown that it is feasible to incorporate such tools into a variety of community-based programs. Instruments have been created which have simple, attractive formats, easy to produce materials and practical, simple ways of reporting results. In some cases however too little attention has been given to evaluate how well the instruments are really being used and understood within the programmatic context.
RECOMMENDATIONS

- Continued efforts should be made to design instruments in the simplest most culturally relevant ways to optimize their utility in community based programs.

- Pictures seem to add to the educational value and simplicity of some instruments, but their acceptance and meaning should be evaluated along with other elements within the programmatic context.

- A concerted effort among practitioners and researchers should be made to propose practical ways of evaluating the usefulness of instruments within the programmatic context. To what extent the instruments are used and understood could be evaluated through focused discussion groups with clients and systematic observations within the applied setting.

Final Comments

The need for appropriate instruments to measure aspects of psychosocial development of children from 0 to 6 and of developmentally relevant aspects of the childrearing environment and biological, psychological and social risk factors has become apparent to many and has stimulated a great deal of constructive and highly creative work.

Multiple alternatives now exist which can and should be critically evaluated in order to optimize their usefulness. By taking time to pause and evaluate the advances already made, it is hoped that future efforts will become even more effective in benefiting children and families throughout the region.
References


Meisels, S. J. “Can Developmental Screening Test Identify Children Who are Developmentally At-Risk?” Pediatrics in press.


Riccuitti, H. N. “Major Areas of Behavioral/Psychological Assessment During First Two Years of Life,” unpublished manuscript, 1974.


Werner, A., Honzik, M.P. and Smith, R.S. “Prediction of Intelligence and Achievement at 10 years from 20 months pediatric and psychological examinations.” Child Development, 1063-1075, 1968.


Copyright © 1989 Lucille C. Atkin, Ph.D.