CHAPTER 3

THEORETICAL ISSUES RELATED TO CROSS-CULTURAL EVALUATION AND DEVELOPMENT OF 'NON-BIASED' TEST MATERIAL

3.1 INTRODUCTION

Cross-cultural and multi-lingual assessment in African populations is a daily challenge faced by speech-language pathologists throughout South Africa. When assessing an area such as problem solving, language and cognition, the challenge is to ensure that the true potential of the individual, irrespective of culture or language, is reliably measured. There are, to date, no such tests, developed and standardized specifically for African children. Testing is in it self, a particularly western middle class phenomenon, as the manner, content and criteria for evaluation are firmly embedded within western middle class culture and standards.

An attempt to address and deal with the difficulties we face when evaluating language and thinking in a cross- cultural, multilingual context must be made. It will serve to highlight the gap that exists between the demands of a western education system on the one extreme, and rural Africa on the other. The goal is to understand thinking skills in this extreme context, so that true and best potential may be identified when assessing rural African children.

This chapter begins with a perspective on 'non-biased' assessment (Vaughn-Cooke, 1986), and a discussion on the criteria necessary for a 'non-biased' test. A discussion on theoretical constructs and models in the development of a non-biased test follows. The different components of a test which must be considered to make it culturally and linguistically reliable, viz. test translation, adaptation of content, adaptation of stimuli and presentation of the stimuli, are then reviewed.

This is followed by a review of the different ways verbal problem solving may be evaluated, leading towards the development of a protocol for the development of a 'non-biased' test of verbal problem solving. The Test of Problem Solving, (TOPS) (Zachman, Jorgensen, Huisingh and Barrett, 1984) as a theoretical model for testing verbal problem solving skills is

critically reviewed, and a conceptual framework for a 'non-biased' test is presented. A discussion on considerations for the scoring of a 'non-biased' test of verbal problem solving and ability to explain, concludes this chapter.

3.2 PERSPECTIVES ON 'NON-BIASED' ASSESSMENT

Numerous attempts have been made to meet the challenges of testing 'non-mainstream' speakers or speakers outside of the English-speaking middle-class group, on whom most tests are standardized (Vaughn-Cooke, 1986).

On the one extreme, at worst, use has been made of existing inappropriate tools such as the Illinois Test of Psycholinguistic Abilities (ITPA) (Kirk, Mc Carthy and Kirk, 1968), and the Peabody Picture Vocabulary Test (PPVT) (Dunn and Dunn, 1981). These tests were frequently used in the 60's and 70's as research tools, and consistently demonstrated inferior performance amongst 'non-mainstream' children (Vaughn-Cooke, 1986). Recently attempts have been made to translate the PPVT into Zulu and Northern Sotho to make it more appropriate for use with African children (Naidoo, 1994; Pakendorf, 1996), but merely translating tests ignores the significant influences of culture and oral tradition.

At the other extreme, groups such as the Association of Black Psychologists in the United States, have called for a moratorium on the use of all tests that would automatically indicate inferior performance of Black American children (Vaughn-Cooke, 1986). This serves to highlight the need for the development of new 'non-biased' modes of assessment.

Two approaches for improving test instruments have been adopted in the area of test standardization, but neither has proved to be effective in producing a truly 'non-biased' test.

Firstly, following the recommendation that children should be compared to the standard of 'own community', clinicians have standardized existing tests on the target community and used that as the norm standard. However this has led to the development of a lower set of standardized norms for culturally, or socially different children, leading to the assumption that low norms equals low potential. The reason for the presence of low norms is due to the fact

that the very content of the test is biased towards what normal middle-class English-speaking children know and say. This has been shown to be true for the Grammatic Closure Sub-test of the ITPA (Wolfram, 1983).

Secondly, the inclusion of a small percentage of e.g., African children in the sample for standardization was attempted. However, this failed to produce 'non-biased' norms, as the percentage included would have had to reflect the demography of the whole population to be reliable, and this was not the intention. This approach also ignored the test content itself.

Using spontaneous sample analysis when evaluating for language delay, as opposed to standardized tests has also been attempted. This approach has its main problem in the norm referencing of performance. At present there are few norms for language development for African languages in South Africa, even if one did use a mother-tongue speaker to analyse the sample. Even in countries where non-mainstream speakers do speak English, there is still inadequate information on language development of the non-mainstream English used to validate this approach (Taylor, 1986b).

Standardization, which uses criterion-based referencing as the standardized measure, has not provided a solution, as identification of the criteria becomes a problem. In addition, without age norms, performance of the child cannot be compared with other children of that age, but only with pre-determined criteria. However, it may provide the first step in an explorative study of an area not previously investigated, in which the understanding of the issue, and not norm referencing, is the goal.

Modification or revision of existing tests was rejected as a solution, as this resulted in the establishing of two scales of performance, with the 'non-mainstream' one being considered as inferior (Vaughn-Cooke, 1986). This further entrenches the bias that culturally or socially different children are inferior in their potential to achieve.

A number of researchers have proposed that the only solution to this problem lies in the development of new tests specifically for the group to be tested. This does imply that tests constructed in this way would, then, not be suitable for any other cultural group.

3.2.1 Criteria for a 'non-biased' Test

If the intention is to create a 'non-biased' test or research instrument, an attempt should be made to comply with the type of criteria stipulated by Vaughn-Cooke, (1986) for the creation of a 'non-biased' developmental language test for non-mainstream cultures. The following discussion takes place within the context of creating a 'non-biased' test for evaluating thinking skills and their application to problem solving in everyday life.

- The test should be based on *valid assumptions* about problem solving, language and cognition. Thus there must be clarity about what problem solving is, as well as which language and cognitive skills must be demonstrated in order to lead to the conclusion that this individual is capable of problem solving. Verbal explanations and the underlying thinking skills have been put forward as one manner of arriving at such valid assumptions (chapter 2, 2.7).

- The test should account for a variation in *language and cognitive style* of the cultural group. In cross-cultural testing this variation may be of a dialectal and articulatory nature or may involve the use of a totally different language. The issues of test translation become significant here (chapter 3, 3.4.1). There is recognition that cognitive style and its relationship to the oral or literary tradition must inform the mode of testing, the design of the test, its elicitation procedures as well as the framework for eliciting responses unique to that culture (chapter 2, 2.3). Although production tests add a particular level of complexity in the process of translation of the test material as well as the accurate translation of responses, it is the most effective way of ensuring the cultural validity of the test (chapter 2, 2.7.5).

- Vaughn-Cooke, (1986) stresses that valid assessment tools should have a universal perspective in that they should assess universal concepts of language. This has direct application to problem solving and explaining and their underlying skills, which are universal to all languages and cultures.

- The test should be based on a *developmental model* . This would inform us of the order in which skills can be expected to appear as well as identifying the child who is not performing according to the expected age norm. This is a particularly difficult area in

cross-cultural testing of skills related to schooling and academic success. On the one hand there is emphasis on identifying disorder or failure in relation to the child's own culture (Taylor and Clarke, 1994) and on the other, there are particular skills which must be developed and achieved through schooling in order for children to be academically successful. Two situations may arise. One, is that, through the test performance, a child is identified as falling below the norm of other children of his/her age and culture. This is a clear presentation of disorder. The other is when, through measuring test performance, there is a mismatch between the skills present in the whole group of children of a particular culture and the identified skills required for effective problem solving at a particular stage of schooling. This information may be constructively used to devise broad-based intervention programmes, which could be introduced at the school rather than the individual level.

- Results of the test should provide some *principled guidelines for intervention.* This is directly related to the validity of the assumptions underlying the assessment tool stated above. The PPVT has been criticized on the basis, that whereas it identifies the child's receptive vocabulary age, and even extrapolated IQ, there are no indications within the test itself to act as a guideline in the remediation of a child falling below his/her age norm.

- In assuming that problem solving ability depends on identified thinking skills which themselves are tested, failure in any particular skill would form the basis of the guidelines for intervention. The TOPS (Zachman, Jorgensen, Huisingh and Barrett, 1984) is such a test, in that it identifies five thinking skills integrating language and cognition, which form the basis of verbal problem solving skills. Should any one of these five skills be identified as below a given norm, that skill may be used as a guideline for intervention.

- The test should provide an *adequate description of some aspects* of the child's cognitive ability when applied to problem solving and explaining. This criterion acknowledges that one tool cannot test every aspect of, for example, problem solving. But it does require of the test that it has a clearly defined focus, and that the information gained about the child will enable the tester to adequately describe that aspect of problem solving behaviour.

- The test should reflect the *latest developments* in problem solving and explanation. The point that is made here by Vaughn-Cooke (1986) is that, in the difficult process of test development cross-culturally, researchers may tend to focus on areas that are easy to access, e.g. syntactic development in developmental language. However semantics and pragmatics, which are essential to a complete understanding of language development, are ignored due to their complexity. Similarly, acclaimed tests, such as the Bracken Basic Concept Scale (BBCS) (Bracken, 1984) may also be seen to be failing in terms of this criterion. This test uses comprehension as the mode of assessment for problem solving skills, and places value on the fact that minimal expression is used. This contradicts the value of leaving as much control for performance in the test in the hands of the child, to decrease the effects of cultural difference (chapter 2, 2.7.5).

Adhering to the above criteria when developing cross-cultural test material will help to ensure the test is as 'non-biased' as possible. Referring to theoretical constructs and models will give a further perspective on the issue of culture-fair testing.

3.3 THEORETICAL CONSTRUCTS AND MODELS IN THE DEVELOPMENT OF A 'NON-BIASED' TEST

Individual ability, learning capacity and individual performance all have some bearing on how an individual is ultimately judged in terms of a particular skill. The attempt to measure a cognitive construct such as verbal thinking skills, particularly in a cross-cultural context, must be based on a sound understanding of these three inter-related factors.

The distinction made between two types of intelligence differentiate emic, culture specific, from etic, universal aspects of intelligence. Cattell (1971) refers to fluid intelligence as a component which is a basic inherited capacity, developed through the interaction with the environment and which is common to all societies and is therefore etic. He refers to crystallized intelligence as a component, which is mediated by culture, and also by schooling, in which specialized skills promoted by and required by a specific culture develop, and is therefore emic. It is the use of tests which measure crystallized ability within a cross-cultural setting, that has come under criticism (Berry, 1984).

The development of language and thinking skills is undoubtedly culture-bound (chapter 2, 2.2) and therefore part of crystallized intelligence. Solving problems of everyday life, which develops through interaction with the environment, is part of fluid intelligence. Developing a 'non-biased' test of thinking skills related to familiar daily problems could exclude language, but it would then focus on conceptual relationships between events. It would be measuring fluid intelligence in the same way as Piaget's tests, or other apparently universal tasks such as seriation, classification, pattern completion, identification of the conceptual odd-man-out, do. Yet even these tests have been shown to be culture-bound in relation to children from Africa (Mwmwenda and Mwamwenda, 1989).

Within each culture, different patterns of abilities emerge. Although they are culture-bound, they emerge as process skills developed through interaction with the environment, and are hence fluid intelligence. This occurs through a process of transfer in which abilities in one area are transferred to another context, facilitating the development of a new ability (Ferguson, 1956). In the example of verbal problem solving of everyday problems, the intrinsic conceptual problem solving abilities, must be transferred into the area of language, and expressed not through action, but through words. Hunt (1980) proposes that it is the cognitive strategies themselves, and not the abilities related to the actual problems, which enable this process of transfer to take place. It is this aspect that is considered to be fluid intelligence. The ability to verbalize the processes is crystallized intelligence, and brings cognitive process into interaction with linguistic development.

A model presenting the inter-relationship among ability, learning and performance (Carlson and Wiedl, 1992), provides a construct for understanding verbal problem solving in a cultural context, and distinguishes three types of intelligence to this end. A representation of this model (Figure 3.1) attempts to integrate the relevant concepts included in this model.

Whereas intelligence A is genetic potential, and B is actual intelligence derived from interaction with the environment, intelligence C is what is measured in intelligence tests or in a test of verbal thinking skills. It is this aspect that is affected by cultural content of the test, familiarity with test material and testing conditions (Figure 3.1).

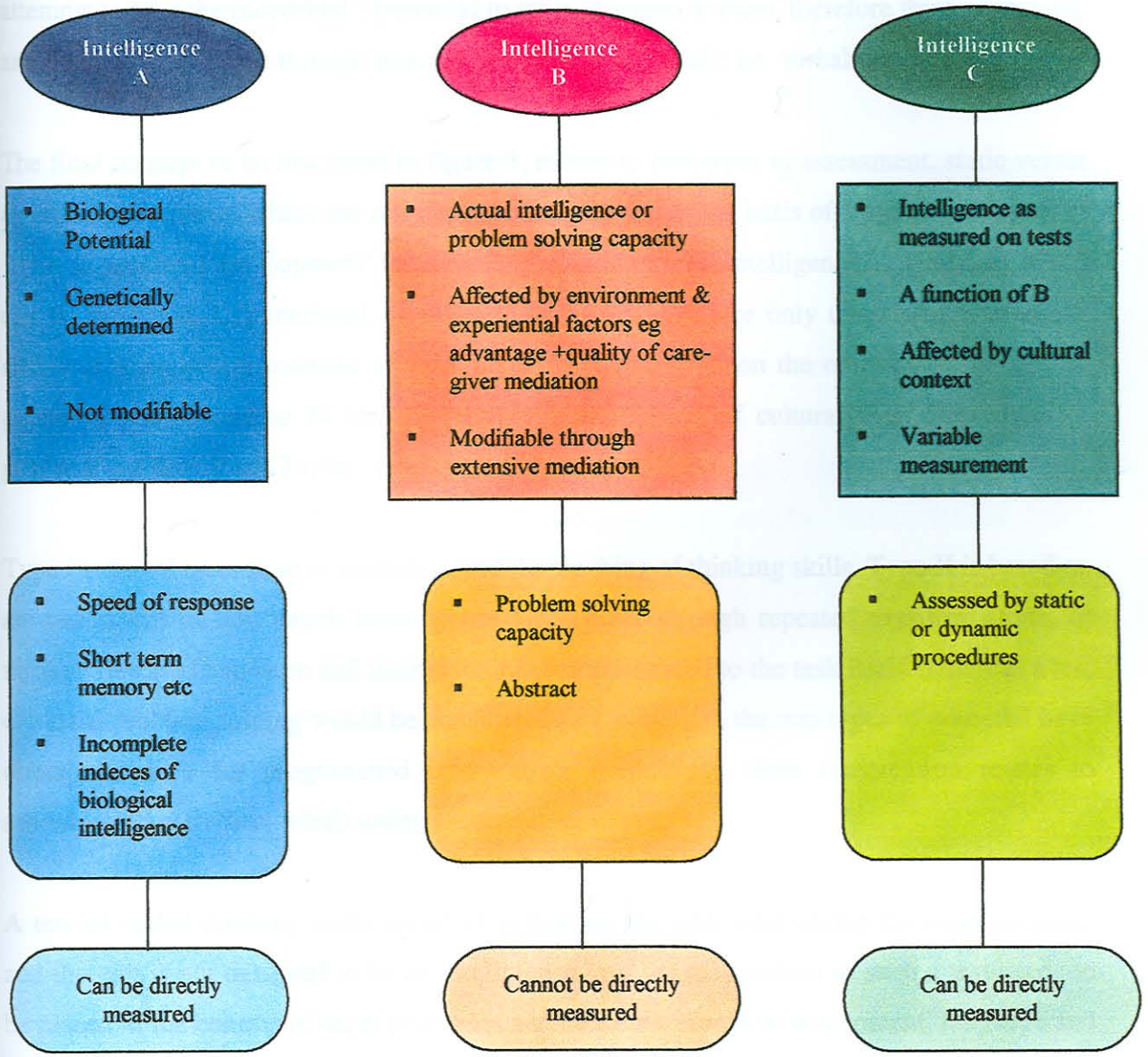

Figure 3.1 Representation of the model of three types of intelligence (Carlson and Wiedl,1992)

Two aspects which may negatively affect test performance, are found in a combination of A and B, and also in C. Thus, factors around the individual (A & B) in terms of (i) genetic potential e.g. speed of response for information processing and memory, and (ii) potential for developing problem solving skills from the environment, form one aspect which may negatively affect test performance. Factors around the test and testing procedure (C) form the other. A test of verbal problem solving, using the context of everyday environmental problems, seeks to use B and effectively test problem solving skills in the context of C. It

attempts to take the individual's potential to infer concepts or rules, therefore think abstractly, and make it measurable through transference to linguistic skills i.e. verbalization.

The final concept to be discussed in figure 3, relates to two types of assessment, static versus dynamic procedures, which are of relevance for culture. On the basis of Vygotsky's theory of 'zone of proximal development' (chapter 2, 2.2), he interprets intelligence as a concept which can be mediated. Conventional, or static tests, tend to measure only unassisted learning of ability at a particular moment in time. Dynamic assessment, on the other hand focuses on potential, and is said to be less susceptible to the effects of cultural bias. Two types of potential are described (Taylor, 1994).

Type I is based on extensive mediation and the teaching of thinking skills. Type II is based on an assessment of how much learning has taken place through repeated exposure alone, or through repeated exposure and instruction or prompts related to the task itself. Whereas a test of verbal problem solving would be considered as a static test, the two types of potential have direct relevance for programmed intervention, particularly when intervention relates to school-based activities, which undergo continuous evaluation.

A test of verbal thinking skills aimed at evaluating the inter-relationship between language and thought is not designed to be an intelligence test, yet the creation of such a test needs to be placed in the context of these principles and concepts. However at a content, language and administration level, cultural relevance continues to be of prime importance.

3.4 FACTORS TO CONSIDER IN THE DEVELOPMENT OF A 'NON-BIASED' TEST

There are four main aspects, which have to be considered in the development of a test for children of a different culture and language. They are translation, selection of content that is culturally sensitive, adaptation of stimuli to be representative of that culture and identification of a method of presentation and test administration that would be most compatible with the culture.

3.4.1 Test Translation

Test translation is required when the researcher who designs the test is not a mother-tongue speaker of the target test population, and therefore, does the basic groundwork in his/her own language. It is also required with test adaptation. Effective translation of a test, from the instructions to the test items, is essential to the reliability of the test instrument.

Recognizing that language is not merely a string of words, but a complex product of socialization and culture, translation of test material should encompass two (of the four) types of translation stipulated by Brislin (1980). Pragmatic translation, when the prime interest is in accuracy of information that is conveyed by the two languages, must be combined with ethnographic translation, which places the language within its cultural context (Brislin, 1980).

Guidelines in relation to language usage, have been given when embarking on test translation for psycho-educational tests (Bracken and Barona, 1991). Such tests would include developmental language tests and tests of verbal problem solving.

It is proposed that: test items should consist of simple sentences; that pronouns should be avoided; that test items should not contain metaphors or colloquialisms; that the passive tense in test direction and items should be avoided; and that hypothetical phrasing and the subjunctive mood (would, should, could) in instructions and test items should be avoided.

However, when the issue of language testing is involved, it is often these very forms of sentence complexity that the tester aims to evaluate, thus complying with all these guidelines may not always be appropriate. Some guidelines, e.g. use of pronouns, may be of relevance to a particular language. In a language such as Zulu there is no differentiation between 3rd person singular pronouns he / she/ it, thus avoiding these pronouns in the test stimuli is essential.

The successful cross-cultural use of the Bracken Basic Concept Scale (BBCS, Bracken 1984), a pre-school test for comprehension, reasoning and organizational skills, highlights particular features which the designers of the test believed contributed to its effectiveness.

As stated, positive value is placed on the fact that minimal expression is required from the child and that the focus is on comprehension. However this would not fulfill Taylor's criterion that a test should reflect the latest developments (chapter 3, 3.2.1), as it has been shown that comprehension tests are not as effective as production tests in identifying disparity (Feagens and Short, 1984; Quay, Mathews and Schwarzmeuller, 1977). Due to the fact that the test administrator could more easily influence the child's response in a test of expression as opposed to comprehension, the importance of the skill of the test administrator, not only for the administration of the test stimuli, but also for the accurate translation of the child's response, is stressed.

A positive feature of the BBCS, is that it has one simple instruction which does not change throughout the test. This is positive as it avoids having to give different sets of translated instructions which would have to effectively change the mental set of the child.

The well documented multi-step procedure for test translation (Bracken and Barona, 1991; Brislin, 1980) which follows, provides an effective model for the translation of all aspects of the testing instrument.

- *Source to target translation*, is achieved by complying with the directive that the translator should be truly bilingual and be sufficiently educated to be familiar with underlying concepts and the relatively formal language of test manuals and record forms. The translator must also be made aware of the importance of retaining the level of language and not attempting to simplify it. The need to achieve pragmatic and ethnographic levels of translation is highlighted.

- *Blind back-translation* refers to the production of a translation from the target language back to the original language by an individual with no prior knowledge of the test and its content. Once again certain criteria must be complied with, viz. that the translator should be educated, bilingual and familiar with technical terminology. The back-translation should be compared with the original material to compare grammatical structure, comparability of concepts, word complexity, and overall similarity in wording, meaning and format.

- *Translation- back-translation repetition* should take place till no further errors are noted.

- *Bilingual Review Committee* consisting of individuals from a number of regions where that language is spoken, should then scrutinize the translated text to identify whether region specific words have been included. If no such committee is available individuals from different areas may be asked to review the text, which may result in some corrections being made.

- *Pilot Testing* should then be done to identify any vocabulary words or phrases that systematically failed to elicit appropriate responses. Unusual responses of the subjects to test stimuli should be noted and followed up with members of the community.

- Finally, the *field-testing* may be carried out, yet even at this stage care should be taken to identify any inaccuracies or inappropriate words or phrases.

3.4.2 Adaptation of Content

In developing 'non-biased' test material, the need to be culturally relevant at the level of content has sound theoretical support.

For maximum reliability, it is important to ensure the automatic recognition and basic comprehension of the test stimuli (Bracken and Barona, 1991), thereby freeing the cognitive processes to focus on the specific task at hand. This could only occur if the basic content of the material used, forms part of the child's world knowledge (Blachowicz, 1994). Thus if a test aims to evaluate problem solving in everyday life, the content of the test instrument, social situations, must be derived from the socio-cultural life-style of the culture of the target population. As a test, which aims to test problem solving skills in a western middle-class American society, the TOPS (Zachman, Jorgensen, Huisingh and Barrett 1984) fulfills this criterion.

Culturally relevant content also has implications for effective test translation. It has been noted that test materials and topics that are simple and familiar to the subjects and the translators, are more easily translated than less familiar materials and topics (Nida 1964).

3.4.3 Adaptation of stimuli

The automatic recognition and comprehension of test stimuli can be further increased by presenting a child with test stimuli that do not only reflect 'own culture' in terms of content, but also in terms of representation. This is achieved by redrawing all the picture stimuli with the culturally appropriate characters and settings, so that there may be a total sense of cultural familiarity, comfort and personal identification with the material. This would help to satisfy calls for cultural sensitivity in tests by various authors (Bracken and Barona 1991, Vaughn-Cooke, 1986).

The importance of the quality of the picture stimulus in terms of whether it is a simple line drawing, a detailed drawing, or a totally realistic photograph has general as well as cultural implications. This has been reflected in the TOPS (Zachman, Jorgensen, Huisingh and Barrett 1984) which has been recently revised (TOPS 1997) so that the simple line drawings used in the 1984 edition have been replaced with photographs in the 1997 edition even though the target population has remained the same. If photographs are not possible then detailed drawings would increase children's automatic comprehension of the content. Essentially, the stimuli must have a positive and motivating impact on the child.

3.4.4 Adaptation of presentation methods

In cross-cultural evaluation, examiners may be faced with children from a predominantly oral tradition and field-dependent cognitive style, presenting challenges beyond the test itself. Method of presentation and administration of test material is another significant factor in the development of 'non-biased' test material.

For children from such cultures, there has usually been minimal exposure to more formal language associated with books and literacy or to the type of discrete point questioning which may spontaneously emerge from this experience (Alant, Tesner, and Taljaardt, 1992). There has also, therefore, been minimal exposure to the social environment in which a child and adult spend time together engaged in continuous dialogue about a subject unrelated to immediate reality (Liddel et al, 1994). The test-taking situation is an artificial one which has, therefore, to be learned (Retief, 1988). Thus, we cannot assume that non-mainsteam children will have the sense of familiarity, challenge or desire for achievement and recognition that is

fostered in a child growing up within a field-independent cognitive style, when faced with the test situation.

Principles related to procedure and administration include those related to the individual administering the test as well as to the test itself.

3.4.4.1 Attitude of the tester

The tester plays a pivotal role in eliciting the best possible response from the child (Bracken and Barona, 1991). The attitude of the tester in terms of integrity of intent, knowledge of all relevant issues and sensitivity to cultural norms have been stressed over and above an adequately translated test (Bracken and Barona, 1991). One particular aspect of cultural norm, which is considered, is the seating position of the child in relation to the tester. This influences the cultural norm of children avoiding direct eye-contact with adults as a sign of respect (Badderly, 1995). If the child is seated next to, and not opposite the tester it may avoid contravening a cultural norm which is prevalent in a traditional culture such as that of the Zulu culture.

3.4.4.2 Race of the tester

The race of the tester was found to be a significant factor (Labov, 1977). It has been demonstrated that the limited output of non-mainstream children, which had given support to the early deviance theory of Bernstein (1970) was partly as a result of the factor of the race of the tester. Having a tester of the same race was found to be more significant for achieving better results, than the methodology of administration in terms of individual or group testing (Wober, 1969).

3.4.4.3 Experience of the tester in administering psycho-educational tests

In accepting that the tester should be of the same culture and language as the individuals tested, it is important to then focus on skills required for reliable testing, if the tester is not a trained professional. Although effective training sessions alleviate this problem, previous experience in scientific test administration would help to ensure an integration of testing protocol and more reliable test administration through maintaining a consistent mode of presentation. This is preferable if possible.

3.4.4.4 Effective use of probes in testing

Vaughn-Cooke (1986) discusses the use of probes in terms of the elicitation of a spontaneous language sample in young children. The purpose of such probing is to ensure that despite the spontaneous manner of elicitation, certain structures and concepts could be elicited. The control for what is elicited still lies predominantly with the child.

In a test that does have a specific focus, yet aims to leave as much control in the hands of the child as possible, probes have a different purpose. Here the aim of the probe is to ensure that, although a basic question serves to elicit the response, the child is given an opportunity to give the 'best possible response' referred to in the introduction to this chapter. This probing must however, be neutral and systematic, to ensure that, if the child does elaborate the answer, it is in response to the prompt conveying a message of 'is there anything more you have to say?', rather than in response to prompt conveying an underlying message that the answer is incomplete or not good enough.

In certain cultures, prompts may actually have the opposite effect to the one intended. African children have been shown to become tense, change their answers and talk less when prompted to explain their answers (Fahrmeier, 1975). In addition they have been shown to perform worse in tests where they are asked to make a judgment and then explain the judgment, than in tests where they are only asked to make a judgement (Mwamwenda and Mwamwenda, 1989). If the aim of the assessment is to evaluate ability to explain, then explanation should be elicited as the initial answer. This would avoid a false negative in evaluating explanation skill.

Neutral probes such as 'umhm, mmmm' are intended to give the child the opportunity to add to the answer without feeling that the previous response was wrong or inadequate. These probes do, however, form a natural part of listener response when a speaker gives an extended verbal delivery. It may be considered a probe if given when it appears to the tester that the child has come to the end of the initial response to the question. 'Is that all?' or 'Anymore' may be used as a communication between tester and child that no more time is required, and that in fact the child has no more to say about that question, and the tester can move on.

3.4.4.5 Method of elicitation and type of response required

The most important aspect about method of elicitation is that is must match exactly the target skill to be tested, i.e. the type of response required. In problem solving, the target may be judgement of conservation in Piagetian tests, sentence completion, mathematical problems or ability to give verbal explanations.

It may not always be possible, to make the way in which a response is elicited, socio-culturally familiar or 'non-biased'. An example of such a situation is where a low socio-economic, field -dependent culture may be evaluated for skills related to education, literacy and verbal reasoning which are not an integral part of that social class or culture. This stems from the fact that for lower socio-economic families, the interactions between parent and child have been shown to be predominantly information or instruction giving, rather than of explanation and reasoning, more common to middle class families (Heath, 1992). Household size within certain cultures also contributes to this problem. Within a rural African context, for example, where the average household size is 7, it has been shown that interaction between parent and child decreases as a product of household size, increasing the interaction between children (Liddel, Kvalsvig, Shabalala and Quotyana, 1994). This decreases the potential for learning to take place within Vygotsky's concept of 'zone of proximal development' (chapter 2, 2.2).

The urgent need for the development of test material in the areas such as language and thinking for non-mainstream children is well recognized internationally, and must be based on valid information about how children perform in these areas. The only way available to a tester to obtain this information about children's thinking, may not be able to meet all the criteria of being fully culturally sensitive.

Thus, a strategy such as posing a verbal question to a child in response to a picture stimulus may be culturally unfamiliar. Non-mainstream children may be unfamiliar with question-answer interaction as a natural part of socialization, as well as having had limited exposure to books and illustrations. However, Ong (1982) has pointed out that there are very few cultures today that have had no exposure to literacy. In addition children who would be tested for language and thinking skills would be attending schools, and although the quality of the

education may be questionable, they will have had some exposure to questions, literacy and books.

Asking questions as an elicitation technique for non-mainstream children has been validated by recent research. Non-mainstream children have been found to be more verbal when presented with a question than when given a statement to validate of refute (Badderly, 1991). They have been shown to perform better on descriptive continuous language responses rather than discrete point naming tasks, as description and discourse form a naturalistic part of cultural communication (Pena and Quinn, 1997). These findings based on studies with Puerto Rican and African-American children, would also apply to a rural Zulu culture, being similarly field dependent. With the emphasis on the social here and now, explanation of social situations may be seen to be a task with which these children are more familiar.

3.4.4.6 Use of training items

As has been previously stated the testing situation is not natural, but one which has to be learnt (Retief, 1988). This is particularly relevant for non-mainstream cultures. There are tests, e.g. the TOPS (1984), which assume that the test context is so familiar to the children being tested that no training item is necessary. When one cannot make such an assumption about the test population a training item is essential to give the child an example of what constitutes a good answer in this context. The training item can be seen as the learning opportunity for the non-mainstream child having to face the unfamiliarity of the test situation.

3.4.4.7 The social environment of testing

The socio-emotional state of the child during testing has been shown to have a significant impact on test performance (Brinton and Fujiki, 1993). This is especially significant for non-mainstream children who have been shown to produce minimal utterances within the unfamiliar and threatening context of evaluation (Labov, 1977). Having a tester of the same race, language and culture, partially addresses this issue, however if further steps were to be taken to create a test-friendly environment, it would increase validity of results.

Rapport may be established with the tester in two ways. Within a dyad at the time of testing, or if a number of children are to be tested, within a group before the children are individually assessed. The group experience offers immediate support for the children who will then not

be immediately faced with the strange adult. The general purpose of testing, what will be expected of the children, the length of the test and an emphasis on it being non-threatening could be conveyed. Games could be played to reduce tension, and tester and children could each tell about themselves to set the context for e.g., verbal expression.

Within the dyad, it would not be possible to create a social atmosphere, but it would offer more opportunity for personal discourse. Both contexts, would offer the opportunity to decrease the negative effects the socio-emotional state of the non-mainstream child on test performance.

In conclusion, if all the above factors were to be implemented in the development of tests for non-mainstream children, it would go a long way to increasing the probability of creating 'non-biased' test material. The creation of the non-biased test in relation to the above must also be able to answer the question: Does the basic method of testing of, for example, problem solving and ability to explain, reflect a considered decision based on all other possible ways of evaluating the particular skill.

3.5 DIFFERENT WAYS OF EVALUATING VERBAL PROBLEM SOLVING SKILLS

In the process of solving a problem, a child must have the ability to understand what exists, to experiment with a range of possible solutions, to select the appropriate solution, and to demonstrate either through action, words or numerals, the outcome of the problem solving process that has just taken place (Blachowicz, 1994). Giving a verbal explanation is one outcome of the problem solving process, which may be analysed in order to understand how children think.

In the TOPS (Zachman, Jorgensen, Huisingh and Barrett, 1984), the problem-solving task is put into place by posing a number of questions about a picture stimulus of a social situation familiar to the child. The complexity of the task is manipulated by the linguistic structures of the questions that have been formulated. Each type of question calls upon different levels of thinking skill in order to give the correct answer. Thus the type and quality of explanation given is a direct reflection of the child's ability to understand the question, to search for the

correct answer, to select the appropriate lexicon and syntax relevant to the question, and present the solution.

The processes of problem solving and the understanding of causality have been evaluated in a number of different ways since the early work done by Piaget in 1956, was started.

The first guidelines for evaluating explanation and causality resulted from a study in which a *written sentence completion task* was administered to 7.0, 8.0 and 9.0 year olds (Piaget 1956). The task required children to complete the sentence following the word 'because' . Piaget proved that children of 7-8 years were more successful answering empirical (physical) questions, e.g. the man fell off his bicycle because....., than logical questions e.g., Paul says the little cat cannot swallow the big dog because..... , which was achieved by 9 years. It was only after 10 years that deductive questions such as, half of nine is not four because..... could be answered. Thus developmental norms for empirical and logical explanations began to emerge. Piaget also proposed that children would understand psychological relations first, physical relations next, and finally logical relations.

A *metalinguistic approach* to the study of causality in terms of judging acceptability of silly sentences revealed that, in support of Piaget, young children of 4.0 performed better on affective (psychological) items rather than on logical items. Children were also shown to perform better on acceptable, rather than unacceptable sentences. This method of evaluating reasoning was criticized for a number of reasons. Donaldson (1986) points out that the instinct of children to acquiesce may result in false positives and false negatives. Further, the assumption that answers are either acceptable or unacceptable is false. As the sentences are all out of context, children may create an inner context in which to validate some of the sentences that are strange to them. Thirdly, children may not use the causal connective to judge acceptability, but rather the acceptability of the two clauses in terms of their world view.

Getting children to *judge synonymety* of sentences, and use of memory recall in which the child has to match a target causal sentence, with one of five other possible sentences with the same meaning were also found to be problematic for the evaluation of causality and reasoning. This is due to the significant reliance on memory in these tasks (Donaldson, 1986).

Using a *comprehension task* to evaluate reasoning, in which picture alternatives of the options are presented, is also not considered an accurate indication of reasoning. This is due to the fact that the children are provided with cues in selecting the correct response, and they are therefore not providing a product of their own thinking processes (Donaldson, 1986), which would be culturally unique.

Production tasks have been used to evaluate causality, but these tend to have been tightly structured around the production of 'because' sentences, and have not been very discriminating in identifying stages of development (Donaldson, 1986). Attempts to chart the development of causal connectives have also been attempted, by studying spontaneous speech samples. It is interesting to note that even with older children of 9-10, these studies revealed the use of predominantly affective statements, expressing simplistic psychological causality.

In the *psycholinguistic studies* of children's explanations carried out by Donaldson (1986), she did aim to relate linguistic and cognitive processes, by using verbalization around a toy for which an action resulted in a particular effect. However, her studies focused in particular on the development of reasoning in the younger child.

Finally, while a test such as the Test of Auditory Reasoning and Processing Skills (TARPS) (Gardner, 1992) provides a formal means of evaluating reasoning, the focus on auditory processing requires the child to produce one-word answers only. Although language may have been manipulated at a competence level, the process is not reflected in the type of answer presented.

Reasoning skills and the development of the ability of the child to put thinking processes into words, is related specifically to the primary school child. This is particularly relevant to the use of decontextualized language. Through understanding whether the inability of the child to express reasoning is contributing to school failure, effective intervention strategies may be identified. As literacy is about language and the manipulation of meaning, a test instrument for the above purpose must be able to access the way in which language manipulates meaning and how the child deals with this. The test instrument must provide the child with the opportunity to formulate and encode an explanation, which may then be evaluated.

3.6 TOWARDS A PROTOCOL FOR THE INVESTIGATION OF CHILDREN'S VERBAL SOLUTIONS TO PROBLEMS OF EVERYDAY LIFE.

As a test of problem solving, reasoning, understanding of causality and of explaining, the Test Of Problem Solving (Zachman, Jorgensen, Huisingh and Barrett, 1984) embodies the features which make it an appropriate model for evaluating thinking skills, which may also be applied within a cross-cultural context.

3.6.1 Reasons for the selection of the Test Of Problem Solving (Zachman, Jorgensen, Huisingh and Barrett, 1984) as a model for evaluating verbal problem solving.

3.6.1.1 Aim of the Test

The stipulated aim of the TOPS is to evaluate a child's ability to integrate semantic and linguistic knowledge with reasoning ability, which is the essence of verbal problem solving. It is also a test, which grew out of a trend in educational philosophy and trends in the field of speech pathology, to evaluate the functional, as well as formal aspects of language.

3.6.1.2 Content of the Test

This test uses social or pragmatic reasoning as the unit of evaluation of problem solving abilities, rather than problem solving related to physical reality or pure logic. This is of particular relevance in cross-cultural testing, as much research has already been done to indicate that culturally different children are slower to acquire reasoning about physical reality when performing Piaget's tests (Dasen, Ngini and Lavallee, 1979; Mwamwenda and Mwamwenda 1986). There is minimal research in the area of verbal reasoning, in a pragmatic context.

3.6.1.3 Mode of Testing

The mode of testing used in the TOPS is that of production. This would permit the researcher to impose minimal constraints on the response given by the child, and permit the test to reveal the reasoning intrinsic to that particular culture. A production study also forces application of

language to thought processes. Verbal reasoning is critical in education, therefore evaluating it provides hard evidence for the lobby, to improve education for non-mainstream children.

3.6.1.4 Cultural Adaptability

The TOPS is a test, which lends itself to cultural adaptation in terms of the test stimuli which could be translated, redrawn, and the social content could be revised.

3.6.1.5 Translation

This could be easily achieved as the socio-cultural context could ensure that there were no culturally inappropriate words or concepts.

3.6.1.6 Ease of Administration

As this test requires only the booklet of picture stimuli, and a battery operated tape recorder, administration is relatively simple, which is appropriate if testing is done in a rural context.

3.6.1.7 Elicitation of targets

The TOPS targets five specific types of thinking skills in which explanation of causality or reasoning is required and which would facilitate the development of cognitive and academic language proficiency. The content or type of explanation to be presented is directly influenced by the linguistic structure of the question, which may increase the level of skill needed to answer the question effectively. Thus the questions themselves serve to identify these thinking skills. Questions have been validated as a method of eliciting responses cross-culturally (chapter 3, 3.4.4.5) Identification of these thinking skills provides a structure for the evaluation of the broad category of problem solving and explaining.

3.6.1.8 Non-timed test

It has been shown that subjecting 'non-mainstream' children to timed tests, subjects them to stress reducing the quality of output (Vaughn-Cooke, 1986). The TOPS is not a timed test, and takes from 15-25 minutes to administer, depending on the age of the child.

In light of the above discussion, creating a 'non-biased' test instrument, would require significant changes that would have to be made to the TOPS as a model to achieve what is required.

3.6.2 Limitations of the TOPS as a model for a cross-cultural test of verbal reasoning

3.6.2.1 Problem solving versus Explanation

The TOPS claims to be a test of problem solving, but in actual fact tests only one aspect of the larger process. Whereas problem solving has been described as the outcome of a process in which a child applies thinking skills to a problem (chapter 2, 2.7.1), explanation refers to the verbal process which is the overt manifestation of underlying thinking skills. Problem solving may occur in relation to a problem in the physical world, a mathematical problem, a literary problem or a social problem and thinking skills are applied in all cases. Fundamental to all problem solving is the understanding which grows through verbalization of the process. This may occur overtly, covertly or it may have become so automatic with learning and development that previous steps taken can be eliminated in reaching the resolution of the problem. Thus in the process of achieving Piaget's stage of formal operations (Ginsberg and Opper, 1969) ability to give verbal explanations must be achieved. Analysing children's explanations is one way of evaluating cognitive and academic language proficiency, which Cummins (1985) believes, is essential for successful education. If the format of the test is that of questions and answers about social situations, then the child is demonstrating the ability to explain. Testing children's ability to explain immediately clarifies the aspect of problem solving being evaluated and is thus felt to better reflect what the test is evaluating.

If such a test is to be used to test thinking skills cross-culturally, the name of the test should reflect this aspect.

3.6.2.2 Scoring of linguistic, semantic and concept criteria

There is an emphasis in the TOPS on the quantitative scoring of errors of syntax (grammar) and semantics (vocabulary) (Zachman, Jorgenson, Huisingh and Barret, 1984). Thus a score of 2, 1, or 0 is allocated, based on the relevance of the answer to the problem (concept) and on the quality of the answer linguistically and semantically. It is further stated that even if an

answer appears to be conceptually correct, 1 point is deducted for a syntactic or semantic error. This penalty appears to be inappropriate in a test of problem solving, which implies that it is the solving of the problem that is the issue and not the child's ability to use an irregular past tense verb correctly. The deduction of 50% of the possible maximum score also appears to be inappropriate.

Quantitative scoring for linguistics would also be problematic in a cross-cultural and cross-linguistic context. Dialectal and cultural factors may influence the type of words and grammar used resulting in a false negative, as the explanation given may be accurate. Further, if the child's answers have to be translated to be scored by a therapist, then it would not be ethically acceptable to score for syntax as the language skill of the translator would be an interfering variable.

Thus emphasis should be placed on the quality of the explanation as a reflection of the child's thinking skill, with some consideration being given to accurate use of vocabulary which is significant for learning.

3.6.2.3 The limitations of a three point scoring scale

The three-point scale above is very restricted in dealing with the gradation of responses that may be obtained in relation to the complexity of the concepts it aims to evaluate. It was also found that with such a limited range of possible scores, errors could be made in the allocation of scores (Bernhardt, 1990) Within the point-scale of only three points, 0 1 2, the greatest degree of error was found in the allocation 0, i.e. the answer that was considered incorrect (Bernhardt, 1990). It would appear that in order to earn 2 points, the child's response had to be correct within the thinking skill targeted, and the linguistic and semantic structure. To earn 1 point, there had to be evidence of the thinking skill, though this may be expressed 'imprecisely' in terms of the linguistic and semantic structures. To earn 0 points, two possibilities existed. First, the child's response could be irrelevant in terms of the thinking skill targeted. Second, the thinking skill reflected in the response was acceptable though not the most 'appropriate or concise', but it was also imprecise in terms of linguistic and semantic structure. This led to the documentation of inconsistency in scoring (Bernhardt, 1990).

Thus a scoring system with a wider range of possible scores would result in the creation of a greater range of ability identified, increasing reliability of the test for the skills being evaluated.

3.6.2.4 Use of Probes

The authors of the TOPS state that probing is to be encouraged to clarify responses or obtain more information, and that it is left to the discretion of the examiner to quickly determine whether probing is necessary. They stipulate that there are three allowable probes: "Tell me more" "Explain it to me better" and "What do you mean?" In terms of the previous discussion on the effective use of probes (chapter 3, 3.4.4.4), the above instructions may create a situation in which process of administration is inconsistent resulting in inaccurate scores being allocated.

In a cross-cultural context, it may not always be a trained clinician administering the test, so determining the 'need' for probing is open to abuse. Further it has been shown that non-mainstream children may interpret probing to have negative implications (Badderly, 1991; Fahrmeier, 1975). The above examples of probing may all be interpreted as ' what you have given is not adequate' which may have a negative effect on the child's attitude to the test.

Probing should therefore be neutral with some communication between child and tester to indicate the child has nothing further to contribute.

3.6.2.5 Culture-bound content of the TOPS

It has been noted that early socialization, cognitive styles, socio-cultural realities all impact on the development of thinking skills and the ability to problem solve (chapter 2, 2.2, 2.3, 2.4, 2.5). Thus in terms of a middle class American child the TOPS is culturally relevant. However in the context of the non-mainstream child, whether in America or in the developing world, the TOPS would require extensive revision. This is most evident in the social contexts selected for use in this test.

3.6.2.6 Lack of a training item

A training item may have been felt to be superfluous in the TOPS due to assumptions made about the target population in terms of familiarity with the test situation. In addition, the

minimally structured approach to probing may be seen as a means of facilitating the best possible answer.

However, in a cross-cultural context, the assumptions about the target population cannot be made, nor is it valid to have such a loosely structured approach to probing. Having at least one training item would give the child an understanding of the task, as well as an example of what constitutes a good answer. In addition the child would understand the type of cues that would be given to help achieve such an answer.

3.6.2.7 Poor test validity in scoring criteria

The TOPS has been challenged in relation to the content validity of the scoring criteria for evaluating responses (Bernhardt, 1990). It has been demonstrated that when speech pathology clinicians and students evaluated responses by referring only to the scoring criteria outlined in the manual there was less than 50% agreement in the scores obtained. This led to the conclusion that the scoring criteria were therefore, not conceptually meaningful. The TOPS does provide a number of examples to clarify the scoring criteria, but it was felt by Bernhardt (1990) that unless the scoring criteria alone could provide a valid reflection of the child's ability, it was not possible to use this test with any degree of confidence.

The behavioural descriptions used in the scoring criteria, e.g. "a response that is acceptable, but does not contain the most concise or appropriate information, but is linguistically and semantically imprecise reflecting vagueness, ambiguity, confusion, incompleteness or immaturity" (TOPS, 1984,manual p 14,), are the basis on which norm referenced evaluations are made. Thus if there is little confidence in the criteria, a norm referenced conclusion lacks validity.

Research in a number of different areas of cognition (discussion to follow) has shown that norm-based referencing is not appropriate for non-mainstream children, and that criterion-based or correlational evaluations are preferable (Badderly, 1995; Viljoen et al ,1994).

3.7 TOWARDS A CONCEPTUAL FRAMEWORK FOR THE SCORING OF A 'NON-BIASED' TEST OF ABILITY TO REASON AND EXPLAIN

A number of tests have been devised to describe and evaluate cognitive skills of children in developing countries which demonstrate that criterion-based or correlational evaluation procedures are preferable to norm referencing. In addition the inappropriateness of using foreign norms on a particular culture has been repeatedly demonstrated.

The administration of the Bender Gestalt Test of visual motor perception to primary and high school Zulu-speaking children school confirmed the inappropriateness of the foreign norms to Zulu-speaking children (Viljoen, Levett and Tredoux, 1994). In a series of tests administered to children in Jamaica, which included an expressive test of semantic categories, a test for comprehension based on the ability to repeat silly sentences, and a test of ability to learn new vocabulary in a foreign language, mean scores were used and compared with scholastic achievement scores (Badderly, 1995). This was done to evaluate the reliability of the cognitive tests in correlating with scholastic achievement. Here, the test was not norm referenced to existing standardized norms, but used as a correlation score, resulting in a high level of validity.

These two studies, in addition to the previous discussion on the difficulties of norm referencing in a cross-cultural context (chapter 3, 3.2) support the notion that, in the initial stages at least, criterion-based referencing (and correlational comparisons), and not norm-based referencing, would be more suitable as a scoring system for a 'non-biased' test . If these directives were to be applied to a test of verbal reasoning, the content validity of the criterion-based system would have to be ensured by being based on a firm theoretical foundation of particular criteria and the role they play in indicating verbal reasoning ability.

In taking cognizance of the documented weakness of the content validity of the scoring criteria in the TOPS (Bernhardt, 1990), it is important to ensure such content validity if a similar test is to be designed.

The theoretical basis on which the scoring criteria for a test of ability to reason and explain are formulated, may be drawn from: the work of Piaget (1956 ; 1968); studies of children's

expression of causality in narratives (Kemper and Edwards, 1986); psycholinguistic studies of children's' explanations (Donaldson, 1986); the concept of cognitive/academic language proficiency and the need for the development of decontextualized language within an educational context; the role played by world knowledge in solving problems and presenting explanations (Blachowicz, 1994); and indicators from developmental process in event comprehension (van den Broek, 1997)

Considering the above theoretical framework as well as the complexity of verbal reasoning, a five point scale, **0 1 2 3 4**, was felt to be preferable.

Specific scores would be allocated on the following theoretical bases:

Firstly, the scale is based on three types of explanations, which were first identified by Piaget (1926) and developed by Donaldson (1986). Thus physical, psychological and logical relations would be expressed as empirical, intentional and deductive reasoning, with empirical reasoning receiving the lowest score of the three and deductive reasoning receiving the highest. Affective reasoning, assigning simplistic emotions as the explanation, present in children from 2.6 years (Donaldson, 1986) would receive a score of 0 in this context

Secondly, the "event chain" approach to the analysis of causal structure in narratives (Kemper and Edwards, 1986), provides the bias for differentiating between observable and non-observable causal relations. This concept presents structures for evaluation, in which the above modes of reasoning become inter-linked as causally cohesive events in the narrative, which proceed in a logical sequence. The "event chain" consists of three types of events: physical states and actions, which are considered observable thus empirical, and mental states (intentions and cognitions) which are considered unobservable and thus are intentional and deductive. Actions have been shown to appear first in children's narratives, followed by physical states, and by 6 years, mental states are included. Score allocation would consider this developmental process, as well as the concepts of observable and non-observable causes, which may be described as concrete or abstract responses.

Thirdly, Cummins' (1985) concept of decontextualized language and Blachowicz (1994) emphasis on the role of world knowledge, has influenced the criteria in the evaluation of the

thinking skills, particularly for 'determining solutions' and 'avoiding the problem'. This may also be expressed as concepts of concrete and abstract answers, in which the concrete answer relates to an explanation for which there is evidence in the stimulus picture therefore there is limited reliance on abstracted world knowledge and spontaneous reasoning. The abstract answer refers to an explanation drawn from decontextualized world knowledge and spontaneous logical reasoning, which is causally associated by the child, with the question being asked.

Criteria derived from event comprehension (van den Broek, 1997) would take into consideration the strength of causal connections particularly as a child moves from single causes to a recognition of multi-causality. Within and between episode reasoning may be seen if the answer presented, relates to the information directly presented in the picture, or whether the answer extends beyond that particular 'episode' creating new associations. Shift from concrete observable actions as cause, to the internal states such as goal and motivations also represents a shift from concrete to abstract thinking.

Finally, making the appropriate inference about the type of reasoning demanded by the linguistic structure of the question would be scored. If the question is formulated to determine cause, answering the question by presenting a solution that may be appropriate to the context, indicates that the child has failed to make the correct inference.S/he has not been able to match the rule or context to the criterion demanded in the question and apply a cognitive process to formulate the correct relationship between these two parts. Being able to verbalize this relationship would credit the child with the highest possible score.

Devising a scoring structure based on the above theoretical foundations may therefore increase the content validity of the TOPS, and make use of its basic scoring structure reliable.

The following table provides an integrated summary of scoring criteria identified, the theoretical frameworks from which it was derived, and the score which would be allocated.

Table 3.1

A summary of scoring criteria, theoretical frameworks and allocated score

Scoring Criteria	Theoretical Framework	Score
* No response. * The response is inappropriate or irrelevant. * Simple affect is used as the explanation.	TOPS (1984) TOPS (1984) Donaldson (1986)	**0**
*There is indication that the question has been understood, although the most relevant information is not presented. * Precise vocabulary is not used. *The answer is vague and imprecise in relation to the question. *The answer may be correct but does not relate to the context of the stimulus	TOPS (1984) TOPS (1984) TOPS (1984) Van den Broek (1997)	**1**
* There is indication that the question has been understood in that the response relates accurately to the question. * Use is made of precise vocabulary. * Physical causality is expressed. * One concrete / observable factor is presented in the answer.	TOPS (1984) TOPS (1984) Piaget (1968); Donaldson (1986) Kemper and Edwards (!986)	**2**
* The response reflects clear understanding of the context. * Use is made of precise vocabulary. * Two or more concrete factors are presented in relation to physical causality between two events * One abstract factor in relation to psychological or logical causality may be expressed between: two actions OR an event/agent and an action * The answer relates directly to the context presented	TOPS (1984); Blachowicz (1994) TOPS(1984) Piaget (1968);Donaldson (1986); Kemper and Edwards (1986). Piaget (1968); Donaldson (1986); Kemper and Edwards (1986) Van den Broek (1997)	**3**
* The response reflects a clear understanding of the context. * Use is made of precise vocabulary. * Physical (concrete) or psychological (abstract) causality is presented in addition to abstract logical causality. The relationship between the rule and criterion is clearly verbalized A creative well formulated answer is presented	TOPS (1984); Blachowicz (1994) TOPS (1984) Piaget (1968); Donaldson (1986); Kemper and Edwards (1986); Cummins (1985); Blachowicz (1994); Van den Broek (1997) Van den Broek (1997)	**4**

3.8 SUMMARY

In this chapter, the theoretical issues related to the development of 'non-biased' tests were discussed, and the practical ways in which these principles could be implemented in the development of a 'non-biased' test for verbal reasoning and explanation, was described.

This discussion was presented from a perspective of different attempts made in the past to devise 'non-biased' test instruments. Criteria for such a test were described, based on directives given by Vaughn-Cooke (1986) for the development of 'non-biased' developmental language tests. The important factors of test translation, adaptation of content and stimuli, as well as presentation and administration of 'non-biased' tests was discussed. Evaluation of verbal problem solving and reasoning from a number of different perspectives was discussed.

In the development of a protocol for the investigation of verbal problem solving and explaining, reasons for the selection of the TOPS (1984) as a model for the development for such a test, as well as its limitations were presented. Finally a conceptual model for increasing content validity of the scoring criteria for a test of verbal reasoning and explaining was presented.