

The challenge of cross-cultural assessment—The Test of Ability To Explain for Zulu-speaking Children

Barbara Solarsh and Erna Alant

Abstract

A culturally appropriate test, The Test of Ability To Explain for Zulu-speaking Children (TATE-ZC), was developed to measure verbal problem solving skills of rural, Zulu-speaking, primary school children. Principles of ‘non-biased’ assessment, as well as emic (culture specific) and etic (universal) aspects of intelligence formed the theoretical backdrop. In addition, specific principles relating to test translation; test content; culturally appropriate stimulus material; scoring procedures and test administration were applied. Five categories of abstract thinking skills formed the basis of the TATE-ZC. These were: (a) Explaining Inferences, (b) Determining Cause, (c) Negative Why Questions, (d) Determining Solutions and (e) Avoiding Problem. The process of test development underwent three pilot studies. Results indicate that the TATE-ZC is a reliable and valid test for the target population. A critical analysis of the efficacy of creating a test of verbal reasoning for children from the developing world concludes the article.

Learning outcomes: As a result of this activity (1) the participant will have a clearer understanding of the principles that need to be followed when developing culturally appropriate test material; (2) the participant will understand the process of developing culturally appropriate test material for non-mainstream cultures; (3) the participant will be able to apply the process and principles to other cross-cultural testing situations.

Article Outline

1. Methodology

1.1. Aims

1.1.1. Phase 1: Development of the test—The Test of Ability To Explain for Zulu-speaking Children

1.1.2. Phase 2: Pilot Study I

1.1.2.1. Subjects

1.1.2.2. Procedure

1.1.2.3. Analysis of data

1.1.2.4. Results of Phase 2 (Pilot Study 1)

1.1.3. Phase 3: Pilot Study II—Version 2 of the TATE-ZC

- 1.1.3.1. Subjects
- 1.1.3.2. The test instrument
- 1.1.3.3. Procedure
- 1.1.3.4. Analysis of data
- 1.1.3.5. Results of Phase 3 (Pilot Study II)
 - 1.1.3.5.1. Evaluation of the subject selection procedure
 - 1.1.3.5.2. Evaluation of the orientation procedure
 - 1.1.3.5.3. Evaluation of seating arrangement
 - 1.1.3.5.4. Evaluation of the children's responses to the illustrations
 - 1.1.3.5.5. Reliability of the test instrument in a test–retest trial
 - 1.1.3.5.6. Inter-translator reliability
 - 1.1.3.5.7. Reliability of probing
 - 1.1.3.5.8. Reliability of scoring criteria and inter-scorer reliability
- 1.1.3.6. Summary of the outcomes
- 1.1.4. Phase 4: Pilot Study III—the final version of the TATE-ZC
 - 1.1.4.1. Subjects
 - 1.1.4.2. Test instrument and procedure
 - 1.1.4.3. Analysis of data
 - 1.1.4.4. Results and discussion of the item analysis—Pilot Study III
 - 1.1.4.4.1. Item-test correlation
 - 1.1.4.4.2. Item-scale analysis for five scales (thinking skills)
 - 1.1.4.4.3. Inter-scale correlation
- 2. Discussion and conclusion
- Appendix A. Examples of test stimuli
 - A.1. Example 1:
 - A.2. Example 2:
- References

Cross-cultural and multi-lingual assessment in Black African populations is a daily challenge faced by speech-language pathologists throughout South Africa. When assessing areas such as language, cognition and problem solving, the challenge is to ensure that the true ability of the individual, irrespective of culture or language, is reliably measured. Much research has been produced demonstrating inferior performance of Black South African children on various psychological tests designed for children from western cultures (Bentley, Kvalsvig, & Miller, 1990; Mwamwenda & Mwamwenda, 1991; Viljoen, Levitt, & Tredoux, 1994). There are, to date, no tests to evaluate verbal problem solving, which have been developed and standardized specifically for rural African children. The Test of Ability To Explain for Zulu-speaking Children (TATE-ZC) is a test of verbal reasoning and abstract thinking skills that attempts to account for the many variables requiring consideration, in order to make a test as culture-fair as possible. However, testing is in itself a particularly western middle class phenomenon, as the manner, content and criteria for evaluation are firmly embedded within western middle

class culture and standards (Baddeley, Gardener, & Grantham-McGregor, 1995; Retief, 1988).

Verbal reasoning, which embodies abstract thinking skills, has been targeted in assessments as it represents the interface between language and thought, and is critical to academic progress of all children (Cummins, 1991). However, an explanation given by a child demonstrating reasoning is a complex phenomenon representing more than the integration of language and thought. It is the culmination of: the child's individual ability, nutritional status and socio-economic realities (Bentley et al., 1990), social structures (Ong, 1982), cultural norms (Wood, 1992) and cognitive style (Witkin & Goodenough, 1981), the outcome of the learning opportunities provided by that child's environment (Blank, 1970) and the impact of any formal education the child has received up to that point in his/her life (Fahrmeier, 1975 and Macdonald, 1990; Mwamwenda & Mwamwenda, 1991). Despite this, ability to integrate language and thought, or to encode communication, has been shown to be more significantly related to general cognitive ability than to any of these other factors (Quay, Hough, Mathews, & Jarret, 1980).

An attempt to address and deal with the difficulties faced when evaluating language and thinking in a cross-cultural, multi-lingual context will serve to highlight the range of issues that need to be accounted for in developing test materials in this context. It will also 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 within this dichotomy, so that true potential of rural African children may be identified, and issues of 'difference' and 'deviance' clearly differentiated (Bernstein, 1970 and Labov, 1970). In evaluating cognitive skills such as verbal reasoning and explanations in rural Africa, the context of how culture and social circumstances influence linguistic and cognitive processes, as well as problem solving skills, must be considered.

Culture and disadvantage (Wood, 1992) are two of the significant factors that affect the development of language and cognition. Culture is a fundamental factor affecting primary language development, early cognitive development and communication competence, with a particular emphasis being placed on mother-child communication style (Taylor, 1986). Disadvantage has been shown to impact on secondary language development especially on the development of abstract thinking skills and the acquisition of literacy skills, which contribute to academic achievement (Blank, 1970, Heath, 1992, Macdonald, 1990 and Wood, 1992). Further, cognitive style (Tinajero & Páramo, 1997; Witkin, 1967), and the tradition, oral or literate, to which a culture belongs (Ong, 1982 and Tannen, 1982), also play a significant role. Integration of these aspects forms the basis for the development of 'non-biased assessment'.

Carlson and Weidl (1992) proposed a theoretical model, which represents the inter-relationship among ability, learning and performance. This provides a construct for understanding verbal problem solving in a cultural/emic context and a universal/etic context, and distinguishes three types of intelligence to this end. Intelligence A is genetic potential, Intelligence B is actual intelligence derived from interaction with the environment and Intelligence C is what is measured in intelligence tests or in tests of

verbal thinking skills (see Fig. 1, a diagrammatic representation of the theory). It is Intelligence C that is affected by cultural content of the test, familiarity with test material and testing conditions.

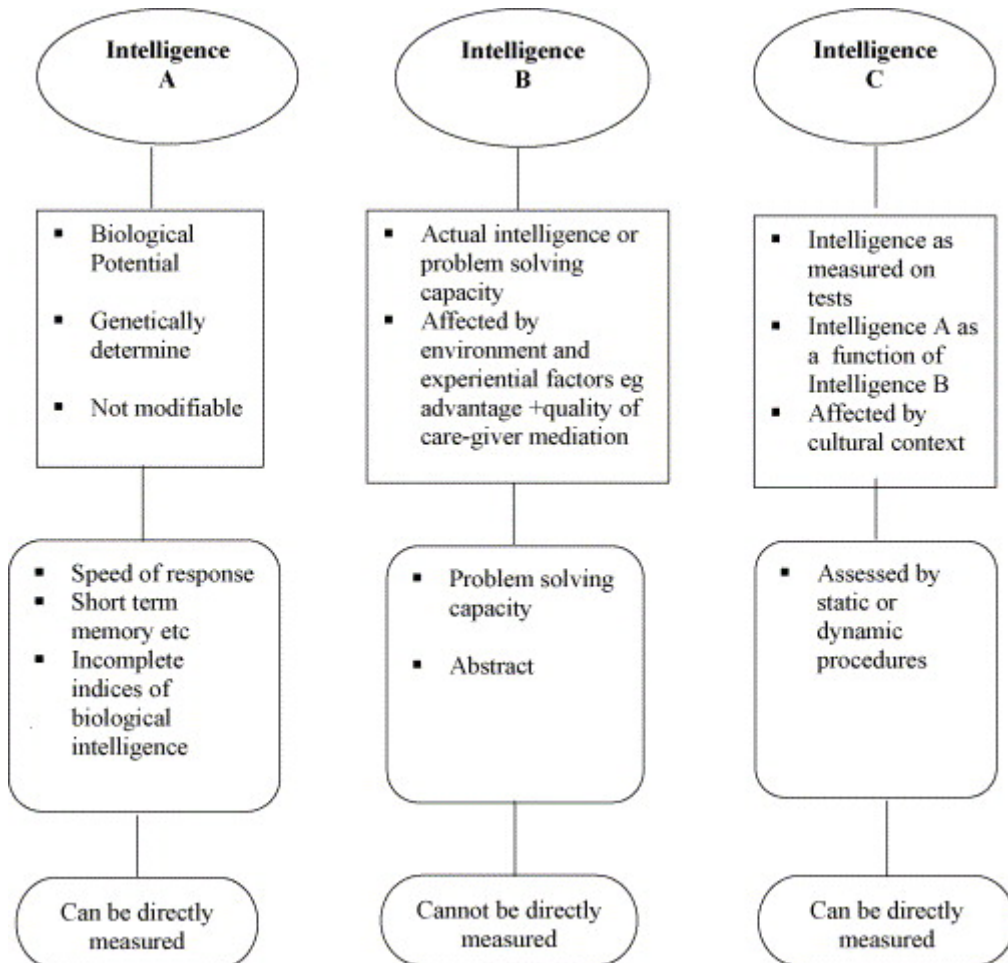


Fig. 1. Representation of the model of three types of intelligence (Carlson & Weidl, 1992). *Note:* This figure was created by the author from the text.

There are two sets of factors that may negatively affect test performance. The first set is reflected in a combination of genetic intelligence and actual intelligence, A and B, and relates to the individual in terms of genetic potential, e.g. speed of response for information processing and memory, and potential derived from interacting with the environment. The second set of factors relates to the test itself and testing procedure, C. 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 abstract thinking, and make it measurable, through transference to linguistic skills, i.e. verbalization.

The final concept to emerge from Fig. 1 relates to two types of assessment, static versus dynamic procedures, which are of relevance for culture. On the basis of his theory of 'zone of proximal development' Vygotsky (1962) interprets intelligence as a concept, which can be mediated. Conventional, or static tests, tend to measure only unassisted and unmediated learning or ability at a particular moment in time. Dynamic assessment on the other hand, focuses on potential to acquire knowledge or skills, i.e. mediated learning (Vygotsky, 1962) and is said to be less susceptible to the effects of cultural bias. 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, and other tests of language ability, needs to be placed in the context of these principles and concepts. However, cultural relevance continues to be of prime importance, at a content, language and administration level.

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, identification of an appropriate method of presentation and test administration/scoring that would be most appropriate to breach the cultural divide (Bracken & Barona, 1991; Mwamwenda & Mwamwemda, 1989). A discussion on each of these aspects follows.

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. Test translation is also required with test adaptation. Effective translation of a test, from the instructions to the test items and responses, is essential to the reliability and validity of the test instrument.

There are guidelines for language usage when embarking on test translation for psycho-educational tests (Bracken & Barona, 1991), which include developmental language tests and tests of verbal problem solving. Bracken and Barona (1991) propose that test items should consist of simple sentences. Pronouns, passive tense, hypothetical phrasing and subjunctive mood should be avoided in test instructions and items, and test items should not contain metaphors or colloquialisms. However, when the issue is language testing, 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 of the suggestions, e.g. use of pronouns, may have relevance to particular languages. In a language such as Zulu there is no differentiation between third person singular pronouns: he/she/it, and therefore avoiding these pronouns in the test stimuli would be essential.

The well documented multi-step procedure for test translation (Bracken & Barona, 1991; Brislin, 1980), which follows, provides an effective model for the translation of all aspects of the testing instrument. The six steps described below would ensure valid and reliable translation.

- *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.
- *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.
- *Translation-back-translation* repetition should take place till no further errors are noted.
- A *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.
- *Pilot Testing* should then be done to identify any vocabulary words or phrases that systematically failed to elicit appropriate responses.
- Finally, *field-testing* may be carried out, yet even at this stage care should be taken to identify any inaccuracies or inappropriate words or phrases.

Adaptation of test content refers to the selection of meaningful and familiar contexts to use as test items, which must be present in the child's world knowledge (Blachowicz, 1994 and Milosky, 1990). Adaptation of test stimuli refers to how this context is physically represented in the test, i.e. the culture of the characters and settings pictorially represented. For maximum reliability, it is important to ensure the automatic recognition and basic comprehension of the test stimuli (Bracken & Barona, 1991), thereby freeing the cognitive processes to focus on the specific task at hand. Creating a sense of cultural and social familiarity as well as comfort and personal identification with the material, would help to satisfy calls for cultural sensitivity in tests (Bracken & Barona, 1991; Vaughn-Cooke, 1986).

The test-taking situation itself may be seen to be an artificial one which has, therefore, to be learned (Retief, 1988). Not only do children growing up in socially and educationally disadvantaged environments (such as those found in rural South Africa), have minimal exposure to the more formal language of tests associated with books and literacy, they also have had little experience of the type of discrete point questioning which may spontaneously emerge from such exposure (Alant, Tesner, & Taljaardt, 1992). There could also have been minimal exposure to a social environment in which a child and adult spend time together engaged in continuous dialogue about a subject unrelated to immediate reality (Liddel, Kvalsvig, Shabalala, & Qotyana, 1994). Hence, the assumption cannot be made that children will universally experience a sense of familiarity, challenge or desire for achievement in a test situation, that may be fostered in children growing up within a field-independent or western cognitive style.

Factors related to 'non-biased' testing procedure and administration include those related to the individual administering the test as well as to the test itself. Factors related to the

tester include attitude (Bracken & Barona, 1991), race (Labov, 1977) and training (Wober, 1969).

The tester plays a pivotal role in eliciting the best possible response from the child (Bracken & 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 noted as more important than an adequately translated test (Bracken & Barona, 1991). The race of the tester was also found to be a significant factor in terms of the length of verbal responses (Labov, 1977). It has been demonstrated that the limited output of culturally different or 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 has been found to be more significant for achieving better results, than the method used to administer an individual or group test (Wober, 1969).

Accepting that the tester should share the same culture and language as the individuals being tested (Labov, 1977), it may not always be possible to find a tester of a particular culture who is a trained professional. It therefore becomes important to focus on developing the skills required for reliable testing in non-professional individuals. Whereas it would be preferable for all testers to be trained in scientific test administration, effective training sessions of the available tester, stressing consistency in mode of presentation would assist in ensuring the most reliable test administration procedure as possible.

Having a tester of the same race, language and culture partially addresses the unfamiliarity of the test situation. However, if further steps were to be taken to create a test-friendly environment, it would increase validity of results. Some of these factors relate to the testing process itself, and include use of probes and the relationship between method of elicitation and response required (Vaughn-Cooke, 1986).

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, target structures and concepts could be elicited. However, in a test that does have a specific focus, the aim of the prompt is to encourage the child to give the best possible answer. This probing must be systematic and neutral and invite the child to elaborate on an answer, rather than judge the answer as incomplete or inadequate.

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

answer. Targeting the explanation immediately would avoid a false negative in evaluating explanation skill (Mwamwenda & Mwamwemda, 1989).

Neutral probes such as ‘umhm and 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 are also a natural part of listener response to a speaker's extended verbal delivery. Thus, they may be considered probes if given when the tester thinks that the child has come to the end of the initial response to the question. ‘Is that all?’ or ‘Anymore?’ may be used to signal that no additional time is required, and that the child has no more to say in response to the question, and the tester can move on.

Asking questions as an elicitation technique for culturally different or non-mainstream children has been validated in the following recent research. Non-mainstream children have been found to be more verbal when presented with a question as opposed to when they are given a statement to validate or refute (Baddeley et al., 1995). They have been shown to perform better when giving descriptive continuous language responses rather than on discrete point naming tasks, as description and discourse form a naturalistic part of cultural communication (Pena & Quinn, 1997). These findings based on studies with Puerto Rican and African–American children may also apply to a rural Zulu culture, which could also be classified as a field dependent culture (Tinajero & Páramo, 1997; Witkin & Goodenough, 1981). Communication in these cultures is typified by social discourse dealing with the ‘here and now’, thus asking a question which elicits such a response in a test situation, may be seen to be a task with which children from field dependent cultures are more familiar (Witkin, 1967).

As has been previously stated the testing situation is not natural, but rather one which has to be learnt, particularly for non-mainstream cultures (Retief, 1988 and Baddeley et al., 1995). Tests such as the Test Of Problem Solving (TOPS) (Zachman, Jorgenson, Huisingh, & Barret, 1984), assume that the test context and process are so familiar to the child being tested that no training item is necessary. When one cannot make such an assumption about the test population, a training item would be essential to give the child an example of what constitutes a good answer in this context (Pena & Quinn, 1997). The training item could be seen as the learning opportunity for the non-mainstream child having to face the unfamiliarity of the test situation.

Finally, one of the important criteria in the development of cross-cultural tests is that the test itself should provide the basis for guidelines for intervention (Vaughn-Cooke, 1986). A criterion-based, rather than norm based reference and scoring system was therefore considered a more valid and useful method for describing results. Research has shown that at an oral and text based level, reasoning and inferential skills have been developed through intervention programmes (Hansen & Pearson, 1983; Winne, Graham, & Prock, 1993). Thus, by identifying cognitive skills required for each level of schooling, criteria in terms of levels of abstract thinking skills (see Table 4) could be matched, and form the basis of programmatic intervention.

It can thus be seen that a commitment to reliably assessing children of a culture and language that is different to a middle class western society, requires recognition of three areas of influence. First, there are factors within the child and the intellectual, social and cultural environment in which the child develops. Second, there are factors in testing and in the test setting, in which the tester has an influence. Finally, there is the test itself, in which the language of the test, the context of the stimuli, the representation of the stimuli, and the methodology, administration and scoring of the test all play a role in the creation of reliable and valid evaluation procedures. This research project aims to account for factors in the second and third areas of influence in the development of a test of verbal reasoning, so that the ability of rural Zulu-speaking children may be reliably tested.

1. Methodology

Ethical clearance for the study was obtained from the Ethics Committee at the University of Pretoria. Permission to use schools in The Valley of a Thousand Hills was obtained from the Circuit Inspector for that area. Permission to use specific schools was obtained from principals of the schools.

1.1. Aims

The aim of this study was to develop a test for the evaluation of verbal pragmatic reasoning, in rural Zulu-speaking primary school children.

This study was conducted in four phases:

- Phase 1: The development of the test in the pre-pilot phase;
- Phase 2: Pilot Study I;
- Phase 3: Pilot Study II;
- Phase 4: Pilot Study III.

The descriptive or qualitative design used in Phases 1–3 included an active participation methodology. An active participation methodology involved the familiarization of the researcher with the social and cultural environment of the children through having worked in a rural school in KwaZulu-Natal for 2 years. Further, consultation with adult members of the target population enabled the researcher to validate observations made, informing each phase of the research. A descriptive or qualitative design was used, as it focuses on process rather than outcome. A qualitative design is also considered to be useful in a cross-cultural context (Retief, 1988). A quantitative analytical survey design was used in Phase 4 for reliability and validity testing, to give credibility to the qualitatively produced test instrument (Leedy, 1989). Statistical procedures were applied to the data in which results of a stratified sample of children in six age groups were evaluated.

1.1.1. Phase 1: Development of the test—The Test of Ability To Explain for Zulu-speaking Children

In developing the TATE-ZC, the Test Of Problem Solving (TOPS) (Zachman et al., 1984) was selected as a model. The following aspects of the TOPS were retained:

- The presentation mode of presenting a picture, asking questions and evaluating the verbal response was felt to be appropriate for use with children from a predominantly field dependent culture.
- The test structure of 15 test pictures and 50 test questions was used as this required the concentration and attention of the child for no longer than 25 min, equivalent to a lesson in school.
- The evaluation of thinking skills in terms of the five categories identified in the TOPS which had been thoroughly researched for that test (Zachman et al., 1984), was considered to be valid. The thinking skills analyzed were Explaining Inferences (EI), Determining Causes (DC), Negative Why Questions (NW), Determining Solutions (DS) and Avoiding Problems (AP). These skills represent different verbal processes of the problem solving procedure (Zachman et al., 1984), and are therefore regarded as a valid reflection of verbal ability to solve problems presented.

While retaining aspects of the TOPS noted above, other aspects were identified as requiring revision in this initial phase. The context from which the test material was drawn, the language of the test, the manner in which stimuli were pictorially represented and the total number of test items required for the inclusion of a training item, were considered necessary to adapt to make the test ‘non-biased’ and culturally appropriate.

The context was the particular social setting of each picture stimulus upon which the questions were based. These social settings had to be formulated in terms of the socio-cultural life experience of rural Zulu-speaking children. An educational officer (originally trained as a teacher) responsible for the management of school-based projects at The Valley Trust (a Primary Health Care Non-Government Organization situated in The Valley of a Thousand Hills) was consulted. Each stimulus from the original TOPS was shown to her and a similar context within the context of rural KwaZulu-Natal was identified. Seventeen contexts were identified, and questions devised so that there was a minimum of 10 different questions in each of the five thinking skill categories.

An artist experienced in the illustration of primary health care manuals was consulted. She was presented with a description of what each illustration should represent as well as the original test stimuli being used as a guideline. She was also informed of the purpose and context for which the illustrations were required, to give her a holistic overview of the task. The test stimuli were all drawn with African characters and settings.

Finally, the questions were translated into Zulu using the multi-step approach to translation as described by Brislin (1980). Table 1 provides a description of how the translation process was carried out.

Table 1.

A Representation of the four stages followed in the translation procedure

	Translation English to Zulu	Back- Translation Zulu to English	Review of Back- transla- tion	Correction	Confir- mation
Translator/ reviewer	A lecturer in the Dept of Zulu, University of Durban-Westville	A senior lecturer in the Zulu Dept at Natal University	The re- searcher	The lecturer in the Dept of Zulu, University of Durban-Westville	An adult living in the Valley of a Thousand Hills
Native language	Zulu	English	English	Zulu	Zulu
Rationale	Colloquial Zulu was required	A translator was required who understood the nuance of the Zulu and could translate this into comparable English	For accurate review of the English back translation	For correction of inaccuracies noted	For accurate review of the final translation

	Translation English to Zulu	Back- Translation Zulu to English	Review of Back- transla- tion	Correction	Confir- mation
Process	The purpose of the test and its basic concepts were explained. The translator was provided with the questions and the picture stimuli to facilitate the understanding of the contexts	The translator here was provided with the Zulu questions and the picture stimuli	The back-translated questions were compared with the original English questions	Twelve questions required revision. The researcher clarified the accurate intent of each problematic question with the translator. Corrections were made to the Zulu translation	The corrected translation was compared with the English question. These 57 questions were used in the pilot project

Phase 1 resulted in the production of 17 pictures, with 57 questions. The two additional social contexts and their accompanying questions were included to pilot as training items in addition to 15 test items. This represented Version 1 of The Test of Ability To Explain for Zulu-speaking Children—the TATE-ZC.

1.1.2. Phase 2: Pilot Study I

The main aim of Pilot Study I was to evaluate whether the adaptations made to the TOPS resulted in a test instrument that was culturally and linguistically appropriate for use with rural Zulu-speaking children, and whether there were additional aspects to the TOPS that needed revision. Both test content and manner of test administration required review in relation to culture-fair testing (Baddeley et al., 1995).

The above aim was realized in terms of the following sub-aims:

- to test the appropriateness of the social contexts used in the test;
- to evaluate the clarity of the illustrations;
- to test the accuracy of the translation of questions used in the test.

Additional sub-aims in terms of additional aspects of the TOPS to be considered for adaptation:

- to review the effectiveness and consistency of the probing described in the TOPS manual;
- to apply and review the scoring procedure used in the TOPS, which was based on a three-point scale, used to evaluate both linguistic accuracy and accuracy of content of the answers to the questions.

1.1.2.1. Subjects

Permission to carry out Pilot Study I was obtained from the headmaster at a primary school in The Valley of a Thousand Hills. The teachers identified six children, one boy and one girl per age category. Three age categories were targeted, 7.6–7.11 years, 9.6–9.11 years and 12.6–12.11 years, covering the age range of the primary school child. Subjects in all the pilot studies had to comply with the following criteria:

- not have failed a class;
- no known sensory impairment;
- no known medical or emotional disorder.

1.1.2.2. Procedure

Permission to work in the identified school was obtained from the headmaster, and relevant teachers were asked to identify the subjects described above. The researcher met with all the children 1 week before testing, and explained the test process and what was expected of them. The researcher met with the Zulu mother-tongue Research Assistant (RA) who was trained and employed by the Human Science Research Council (Durban), and who had extensive experience in psychometric testing of rural children and translation of test material. A one hour training session took place.

Before each child was individually tested s/he was made to feel at ease, the lapel microphone was attached and then testing began. The researcher was present (in the background) throughout testing to observe consistency of presentation by the RA. Testing took place in a small adequately lit staff room. At the end of testing, each recording was translated and transcribed by the RA and all probes were noted. The answers were scored and analyzed according the identified aims.

1.1.2.3. Analysis of data

The results of this early phase were qualitatively analyzed. The appropriateness of the social contexts used, the clarity of the illustrations and the accuracy of the translation of questions were evaluated in terms of whether the answer given by each of the children in the pilot study accurately related to the context, illustrations and language presented. This evaluation was made by the RA and researcher.

The review of the effectiveness and consistency of the probing described in the TOPS manual was evaluated by analysing the pattern of probes identified in the translation and transcription process. Consistency and number of probes presented per question were noted.

The review of the scoring procedure used in the TOPS, which was based on a three-point scale (2, 1, 0), was used to evaluate both linguistic accuracy and accuracy of content of the translated answers.

One credit point was allocated for linguistic accuracy and one credit point for semantic accuracy.

1.1.2.4. Results of Phase 2 (Pilot Study 1)

The results from the pilot study and actions generated are presented in Table 2.

Table 2.

A Representation of the results of Pilot Study I in terms of the five sub-aims outlined

Sub-aim	Result	Problems	Outcome
(i) Evaluate appropriateness of situational contexts	Sixteen out of 17 social contexts were found to be appropriate. The children scored 1 for semantic accuracy in 16 pictures	One social context was found to be inappropriate. The context of the pregnant lady needing to get to hospital was omitted	Fifteen pictures were retained as test items as in the TOPS. One picture was randomly selected as a training item
(ii) Evaluate clarity of illustrations	Sixteen pictures were accurately identified. The children scored 1 for semantic accuracy in 16 pictures	One picture was misinterpreted. The children said the pregnant woman holding her stomach was holding a heavy object	This item was excluded
(iii) Evaluate accuracy of translation	Sixteen questions were accurately translated and understood. The children scored 1 for semantic accuracy in 16 pictures	None	No alteration in terms of translation was required

Sub-aim	Result	Problems	Outcome
(iv) Review probing	Poor consistency and number of probes was noted	Different amounts and types of probes given to different children	A structured system of probing was formulated in terms of number and type
(v) Apply and review the scoring system	Scores were calculated for linguistic and semantic accuracy. The three-point scale was limited. It did not give a graded score for semantics, and Zulu linguistics could not be scored on an English syntactic structure	Scoring criteria for linguistic accuracy were not considered to be valid. The three-point scale provided a limited range of scores for such a complex problem as verbal reasoning	Scoring criteria were designed for cognitive content only. A five-point scoring system was developed based on theoretically sound scoring criteria (see Table 4)

A number of recommendations that emerged from Pilot Study I were implemented. The illustration of one item presented (the pregnant lady) was misinterpreted by all the children. Further discussion about this item with The Educational Officer, who lived in The Valley of a Thousand Hills, revealed that children of 7–9 years were sent away when mothers went into labour, thus the social context was not actually suitable across the age range. This item was excluded. All the children's responses to the other items indicated an understanding of the illustrations and social contexts. Of the 16 remaining items, one was randomly selected as a training item. From the answers presented by each of the children, the RA affirmed that the Zulu questions were accurately understood.

Observation by the researcher and evaluation by the RA concurred that the lack of specificity in the instructions for probing in the TOPS, resulted in different amounts and types of probes given in terms of the child and the response given. A structured system of probing was therefore recommended. An evaluation of the three-point scoring system used in the TOPS applied with the Zulu-speaking children, indicated that linguistic accuracy could not be scored when answers were translated due to the variable of the translator and vastly different syntactic structures of the two languages. As the content of the answer would be the only factor contributing to the score, it was felt that the level of skill should be identified with a greater degree of accuracy—hence the introduction of the five-point scale.

The recommendations of Pilot Study I were implemented, and Pilot Study II followed using Version 2 of the TATE-ZC.

1.1.3. Phase 3: Pilot Study II—Version 2 of the TATE-ZC

The main aim of Pilot Study II was to administer Version 2 the Test of Ability To Explain for Zulu Speaking Children, to evaluate the effectiveness and consistency of the selection, administration and scoring procedures.

The main aim was implemented through the following sub-aims:

- Review of aspects of the subject selection procedure to ensure children fell into accurate age categories and the effectiveness of gaining parental permission through letters.
- Review of the test orientation procedure, as the socio-emotional state of the child has been shown to have a significant impact on testing (Brinton & Fujiki, 1993), particularly in a cross-cultural setting (Labov, 1977).
- Review of the seating for testing, as there is a cultural norm in the Zulu culture, which prevents children having direct eye-contact with an adult. Baddeley et al. (1995) have proposed that the child sits next to, rather than opposite the tester.
- Analysis of the children's responses to the illustrations. A positive response was considered an indication that they enjoyed the testing, and that the stimuli had achieved the goal of being culturally and socially appropriate (Bracken & Barona, 1991, Vaughn-Cooke, 1986).
- Assessment of reliability of the test instrument through the use of a test–retest trial.
- Assessment of reliability of the translation of the audio-taped responses through comparison with a second translator, in order to exclude inaccurate translation as a variable influencing the child's test score.
- Assessment of the reliability of probing in terms of consistency of presentation of probes by the RA throughout the morning.
- Assessment of the reliability of the new five-point scale, the scoring criteria and guidelines for using the five-point scale, using a blind inter-scorer reliability trial.

1.1.3.1. Subjects

Three groups (20 per group) of rural children ($N = 60$) of three different ages (7.6–7.11, 9.6–9.11, 12.6–12.11 years) spanning the primary school phase were included. A sample of 20 per group was chosen so that descriptive statistical analysis could be applied.

The subjects selected were from a different junior primary school and adjacent senior primary school to that used in Pilot Study I. The schools were situated in the Kwadedagendlale district of Kwazulu-Natal. These schools were identified as suitable by the Educational Officer at The Valley Trust as they were judged by her to be

representative of rural schools in the area, and were close to the main road for easy access to the schools in all weather.

Subjects were identified by the teachers through a structured process (see Table 3) in which the exact same criteria used in Pilot Study I were applied. An equal distribution by gender was required. Twenty children in each age group who fulfilled the criteria were then randomly selected. Although the aim states that the sample should span the range of the primary school phase, it was felt that Grade 2 children should form the lower range. The researcher's clinical experience (during a school-readiness testing project), indicated that Grade 1 children provided minimal verbal output in a one-to-one testing situation. Information obtained from the school-readiness project revealed that few Grade 1 children had attended a pre-school, thus exposure to Grade 1 was their first experience of formal learning. In light of problems encountered in cross-cultural testing, it was felt that Grade 2 children, who had already had one year to adapt to a formal learning context, would perform more reliably in a test situation. Thus, Grade 2, Grade 4 and Grade 7 pupils provided the range for the primary school phase.

Table 3.

Description of steps taken in the selection of subjects

Steps	Procedures
Step 1	The researcher selected all possible children falling into the specified age groups, and drew up separate screening sheets for boys and girls. Children in each of the groups were identified in this way
Step 2	The researcher met with the 12 relevant class teachers, gave a brief explanation of the purpose of the research, and enlisted their help with the completion of the selection sheets, which were to be collected the following day
Step 3	Each teacher reviewed the identified children in his/her class, and had to answer the following questions
	Has the child ever failed?
	Does the child appear to hear well?
	Does the child appear to see well?
	Does the child appear to have any emotional problems?
	Does the child have any known medical problem, e.g. epilepsy?
	Is the child a high/average /low achiever?

Steps	Procedures
Step 4	The researcher then selected children fulfilling the specified criteria
Step 5	These children were each given a letter (Zulu) in which the researcher briefly explained the procedures of the research and requested the parent's permission for their children's participation

1.1.3.2. The test instrument

The test instrument used in Pilot Study II was Version 2 of the TATE-ZC, which included the revised scoring and test administration procedures and included one training item. Scoring criteria were revised to create a five-point scale, as opposed to a three-point scale. Specific criteria for allocating each score were identified to differentiate the features required for scores 0–4 (see Table 4). A new set of examples of answers and related scores was devised in terms of the new situational contexts.

Table 4.

A summary of scoring criteria, theoretical frameworks and allocated score

Scoring criteria	Theoretical framework	Score
No response	TOPS (1984)	0
The response is inappropriate or irrelevant	TOPS (1984)	
Simple affect is used as the explanation	Donaldson (1986)	
There is indication that the question has been understood, although the most relevant information is not presented	TOPS (1984)	1
Precise vocabulary is not used	TOPS (1984)	
The answer is vague and imprecise in relation to the question	TOPS (1984)	
The answer may be correct but does not relate to the context of the stimulus	Van den Broek (1997)	
There is indication that the question has been	TOPS (1984)	2

Scoring criteria	Theoretical framework	Score
understood in that the response relates accurately to the question		
Use is made of precise vocabulary	TOPS (1984)	
Physical causality is expressed	Piaget (1956), Donaldson (1986)	
One concrete/observable factor is presented in the answer	Kemper and Edwards (1986)	
The response reflects clear understanding of the context	TOPS (1984), Blachowicz (1994)	3
Use is made of precise vocabulary	TOPS (1984)	
Two or more concrete factors are presented in relation to physical causality between two events	Piaget (1956), Donaldson (1986), Kemper and Edwards (1986)	
One abstract factor in relation to psychological or logical causality may be expressed between: two actions OR an event/agent and an action	Piaget (1956), Donaldson (1986), Kemper and Edwards (1986)	
The answer relates directly to the context presented	Van den Broek (1997)	
The response reflects a clear understanding of the context	TOPS (1984), Blachowicz (1994)	4
Use is made of precise vocabulary	TOPS (1984)	
Physical (concrete) or psychological (abstract) causality is presented in addition to abstract logical causality. The relationship between the rule and criterion is clearly verbalized	Piaget (1956), Donaldson (1986), Kemper and Edwards (1986), Cummins (1985), Blachowicz (1994), Van den Broek (1997)	
A creative well formulated answer is presented	Van den Broek (1997)	

1.1.3.3. Procedure

Permission to use the school and test the children was obtained and the subject selection process by the teachers (see Table 3) completed. The research project met ethical standards for educational research by obtaining permission from the circuit inspector and principals as community representatives to test the children (Code of Federal Regulations, 2001; The Guidelines on Ethics for Medical Research of the Medical Research Council of South Africa, 1993). Each child who was selected for the research project received a letter in Zulu to obtain parental permission for their participation. As many of the participant's parents were illiterate this proved a difficult exercise. Parents were invited to come and discuss the research with the researcher at school if they were uncertain.

The orientations and testing were conducted in a comfortable testing environment and occurred each school morning for 3 weeks. Children participated in the test orientation procedure in groups of 10, as it was felt that a group of this size offered sufficient opportunity for both individual interaction with each child as well as group interaction to ease the children into the testing environment. It took approximately 30 min to orientate each group and the R and RA were present throughout.

Each group of 10 children was called to the library and they were seated in a circle with the R and RA. Through a translator (the RA) the R welcomed and thanked the children, and gave a brief description of what was expected. Emphasis was placed on the fact that they did not have to be afraid and that their performance results would be confidential. An 'ice-breaker' game followed in which each individual, including the R and RA had a chance to say his/her name and simultaneously do an action of their choice to informally introduce themselves and encourage individual and spontaneous response.

It was explained to the subjects that they would be required to talk into a tape recorder, and in order to obtain intelligible recordings to facilitate accuracy of data analysis, each child would be given the opportunity to practice by talking about themselves, their families and interests. The R and RA also participated and used the opportunity to encourage the children to speak loudly and engage in a conversation with the RA. Before the children were sent back to their classrooms to be called individually for testing at a later stage, they were asked if they were willing to participate and then instructed not to talk about the test to each other. Each child was then called individually and tested in the following manner:

- Each child was seated comfortably next to the RA and the child was put at ease through informal conversation.
- The RA presented the test instructions to the child and the training item was presented according to instructions devised by the researcher. The training item gave the child practice with verbal expression and provided an example of a 'good' answer as one, which may contain more than one reason (see Table 4). The lapel microphone was then

attached to the child's collar, the child was asked to say his/her name clearly, and testing began.

- The child was given 10–15 seconds to look at each picture, before the questions were initiated.
- In order to achieve consistency with regard to the elicitation of solicited responses, the presentation of probes was controlled and systematic. In addition to naturalistic responses of ‘mmm’ during an extended monologue, two different control probes were used. When the child appeared to have come to the end of the initial spontaneous response, the examiner used the probe ‘mmm’. When it appeared that the child had nothing further to say, another probe, ‘is that all?’ was used, to which the child could respond in the affirmative or by saying ‘no’ and elaborating on the answer. Following either response, the RA would then proceed with the next picture/question.
- On completion of the last test question, an indication of the child's response to the picture stimuli was obtained by asking the child the two questions. The first was, “Did you like the pictures?” An attempt was made to counter bias in the responses through the child wanting to please, by asking a more qualitative question, “Did you like them a lot, just a little or a medium amount?”
- The child was then thanked, reminded not to talk about the test and sent back to class.
- The tape recordings were then translated and transcribed by the RA including all interjections and probes used by the RA. Audio-taped responses of 20 children were translated and transcribed blind by a second translator for reliability testing and the evaluation of consistency of probing.
- Each script was scored by the researcher according to the newly devised criteria (see Table 4). Twenty randomly selected scripts were then scored by the second scorer.
- A test–retest trial was carried out after a waiting period of 2 weeks in order to limit the drawbacks of this method due to the practice effect (Leedy, 1989). Nine children (15% of the original sample) were randomly selected, three from each age group, and were re-tested using the TATE-ZC.

1.1.3.4. Analysis of data

Each sub-aim was individually analyzed, with both informal observation and statistical methods being used.

- *The subject selection procedure* was analyzed by a total count of number of children correctly selected within the given age range when birthdates were checked. Total count of number of letters returned to the school was recorded.

- *Review of the test orientation procedure.* Each child was observed during testing for signs of distress or discomfort by the researcher and RA.
- *Review of the seating for testing.* The extent to which the children made eye-contact with the RA during testing was informally observed.
- *Analysis of the children's responses to the illustrations.* Total number of positive or very positive responses was calculated as a percentage.
- *Assessment of reliability of the test instrument* was evaluated by using the Test Re-test Method, described in detail by Leedy (1989). A coefficient of stability was obtained by applying the same instrument to the same individuals on two consecutive occasions (Leedy, 1989). A *t*-test for dependent measures was used to determine whether the test means obtained differed significantly. This comparison was performed for total test scores and for each sub-test for each of the children (Steyn, Smit, du Toit, & Strasheim, 1994).
- *Assessment of reliability of the translation.* Twenty subjects' test responses (at least six per age group) were selected at random for blind translation and transcription by a second translator. The R carried out a blind scoring of the second translation to see if the child's core would differ in relation to a different translation. Comparisons of the two translations and scores of each Zulu response were made by the R, and the number of agreements per pair of sentences was calculated. A pair of sentences was allocated the full score if the vocabulary and meaning were similar and if the sentences were allocated the same score.

The formula: total no. of agreements divided by the sum of agreements + disagreements $\times 100$, yielded a percentage of sentences considered to be accurately translated. This excluded inaccurate translation as a variable influencing test scores. A 95% level of agreement was required (Leedy, 1989).

- *Assessment of the reliability of probing* in terms of consistency of presentation of probes by the RA throughout the morning. The 20 responses evaluated for reliability of translation were used to evaluate the reliability of the use of probes, and thereby exclude inconsistent probing as an uncontrolled variable influencing a child's test score. All probes and interjections used by the RA were noted in each of the scripts and compared over three 90 min time slots during the morning, viz. 9:00–10:30 h, 11:00–12:30 h and 12:30–14:00 h. A minimum of six children had to have been tested in each time slot. The two translators were required to identify all probes present in each answer by inserting a (1) for “mmm” and a (2) for “is that all”? Each of the 50 pairs of sentences could score two points to give a total of 100.
- one point was deducted if the sentences differed in the probes identified by the translators, thereby influencing the quality of the response;

- one point was deducted if the probes were not consistently present in each presentation as judged by the two translators.

A total score for the accurate use of probes was calculated with a mean score for each time slot. A 95% consistency rate (Leedy, 1989) in each time slot was required for probing to be considered reliable and consistent over the different time slots.

- *Assessment of the reliability of the new five-point scale, the scoring criteria and guidelines for using the five-point scale*, using a blind inter-scorer reliability trial. Two scorers evaluated the translated responses of 20 different randomly selected scripts (33% of scripts), with at least six children per age group. Results of this process would ensure the reliability of the scoring criteria and scoring guidelines (see Table 4) devised by the researcher. A Pearson Correlation Coefficient (r) was calculated, in which inter-scorer reliability for Scorer 1 (S1) and Scorer 2 (S2) was measured for the total scores and each of the sub-test scores. Both scorers were lecturers in speech pathology, one of them being the researcher. The scripts were scored independently by each of the scorers, using the scoring criteria and scoring guidelines devised for the TATE-ZC.

1.1.3.5. Results of Phase 3 (Pilot Study II)

1.1.3.5.1. Evaluation of the subject selection procedure

The subject selection procedure using the teachers identified 60 children who fitted the age categories accurately in addition to fulfilling all the other selection criteria. Steps 1–4 (Table 3) were therefore retained for Pilot Study III. Step 5 revealed that only a very small percentage (<20%) of letters sent to parents were returned—all of them giving permission for their children to participate. The poor response from parents was interpreted by the headmaster of the school as being due to low levels of literacy amongst the parents, and poor understanding of the need to actively respond, as parents never received communications from the school in this way. He reported that parental participation in their children's education was generally limited and felt it would be impossible to track all the parents at home or work to gain written or even verbal consent. In consultation with the headmaster, it was felt that as permission for the research had been obtained from the Education Circuit Inspector, endorsed by the headmaster, and was not chemically invasive, this was sufficient for the testing to be implemented. In Pilot Study III no letters were sent to parents, but it was agreed that the various principals would answer any queries from the parents, should they arise. The authors acknowledge, however, that all efforts should be made to secure parental permission and propose that as participation in research studies becomes more routine in these contexts, parents will become more willing to provide parental consent for participation in such studies.

1.1.3.5.2. Evaluation of the orientation procedure

Each child was observed by the researcher throughout the testing procedure. Qualitative evaluation revealed that none of the children displayed apprehension, rather eagerness for

their chance and an ease of communication with the RA. The children spoke clearly, attended to the stimuli and responded to all the test questions.

1.1.3.5.3. Evaluation of seating arrangement

From observation, seating the children adjacent to the tester worked well. It was observed that even the younger children did not hesitate to engage in eye contact with the tester intermittently. The researcher thus decided to retain this seating arrangement.

1.1.3.5.4. Evaluation of the children's responses to the illustrations

A Likert scale was used with graded responses of mildly positive, positive or very positive. Forty two percent of children indicated a very positive response (liked the pictures very much), whereas 58% indicated a positive response (they liked them a medium amount). There were no mildly positive or negative responses. The researcher considered the possibility that the caricatured nature of some of the illustrations may have affected the children's responses, thus all 16 illustrations were re-drawn, ensuring the greatest natural likeness to real life pictures as possible, as this was the style used in the TOPS (see Appendix A).

1.1.3.5.5. Reliability of the test instrument in a test–retest trial

Reliability of the test was determined using the *t*-test for paired samples (Leedy, 1989) to measure significance of the difference between mean scores of Test 1 and Test 2.

The correlation between Test 1 and Test 2 was significant for the total score ($t = 0.934$ with $p < 0.001$), and the sub-tests Explaining Inferences ($t = 0.818$ with $p < 0.05$), Negative Why ($t = 0.867$ with $p < 0.05$), Determining Solutions ($t = 0.940$ with $p < 0.001$) and Avoiding Problem ($t = 0.914$ with $p < 0.05$). The correlation for the sub-test Determining Cause was not significant ($t = 0.631$), but as this did not influence the significance of the total score, it was not considered to be significant to the overall result.

These results indicated that the TATE-ZC is a reliable tool for the measurement of thinking skills, and that performance on the test will be consistent if it is re-administered to students, i.e. the test is consistent in measuring level of skill.

1.1.3.5.6. Inter-translator reliability

Percentage agreement for the total sample was 96.13%, with percentages for each of the three age groups falling above the 95% level of agreement (see Table 5).

Table 5.

Inter-translator reliability calculated as a percentage

Age group	Sample size (<i>n</i>)	% Agreement between the two translators
7	7	96.29
9	7	95.43
12	12	96.67
Total group	20	95.13

This result indicates a sufficiently high level of reliability between the two translators, and that the translated responses were therefore accurate. However, an inter-translator reliability measure would be necessary whenever new untried translators are used.

1.1.3.5.7. Reliability of probing

Agreement between the translators as to the consistency in the use of probes throughout the three testing periods was between 97 and 98% indicating a high level of reliability in this aspect of the test administration (see Table 6). These results indicated that the instructions for the use of probes were clear and implementation was reliable. In addition the use of probes in the TATE-ZC did not influence the quality of the responses, an important concern in cross-cultural testing.

Table 6.

Reliability of testers' use of probes, calculated as a percentage, across three time slots

Age group	Sample size (<i>n</i>)	First 90 min	Second 90 min	Third 90 min
7	7	97.67%, <i>n</i> = 3	98%, <i>n</i> = 2	99%, <i>n</i> = 2
9	7	99.5%, <i>n</i> = 2	95.33%, <i>n</i> = 3	96.5%, <i>n</i> = 2
12	12	95.5%, <i>n</i> = 2	97.5%, <i>n</i> = 2	96%, <i>n</i> = 2
Total group	20	97.55%, <i>n</i> = 7	97%, <i>n</i> = 7	97.16%, <i>n</i> = 6

Key: first 90 min: 9:00–10:30 h; second 90 min: 11:00–12:30 h; third 90 min: 12:30–14:00 h.

1.1.3.5.8. Reliability of scoring criteria and inter-scorer reliability

A Pearson Correlation Coefficient (r) was calculated, for inter-scorer reliability for the total test score ($r = 0.9662$) and for the sub-tests Explaining Inferences ($r = 0.9185$), Determining Cause ($r = 0.8785$), Negative Why ($r = 0.9318$), Determining Solution ($r = 0.8823$) and Avoiding Problem ($r = 0.9031$) with all correlations being significant at $p < 0.001$, indicating reliability of scoring criteria and scoring guidelines.

1.1.3.6. Summary of the outcomes

The Test of Ability To Explain for Zulu-speaking Children, the test instrument developed to this point, consisted of a test booklet in which 16 pictures of different contexts relevant to the life of a rural Zulu-speaking child, and 53 Zulu questions with the English translation below to evaluate ability to explain, were adjacently arranged. One picture was a demonstration item, and 15 pictures the test stimuli. Three of the questions related to the demonstration item, and 50 questions were test items. There were two to four questions per picture (see Appendix A, for examples).

Each question related to one of the five thinking skills identified in the TOPS (see Table 7). The thinking skill linked to each picture was dependent on the context of the picture. There was no specific pattern in which thinking skills were targeted. Thus, the total 50 questions were made up of 10 questions per five thinking skills, which were randomly presented according to the possibilities offered by the different contexts. Table 7 presents a brief description of each thinking skill and the type of question asked.

Table 7.

Description of the five thinking skills identified in the TOPS

Thinking skill	Description	Question
Explaining Inferences	The child is asked to explain how he knows that something he sees is true	How do we know that ...?
Determining Causes	The child is asked to state a logical cause for a situation he sees in the picture	Why did ...?
		How did ...?

Thinking skill	Description	Question
The Negative Why	The child is asked to give a reason why one would not behave in a certain way	Why would'nt ...?
		Why are'nt ...?
Determining Solutions	The child is asked to solve various situational problems that are illustrated and presented verbally	A statement is presented followed by a question:
		What should ...?
		What could ...?
Avoiding Problem	The child is asked to find ways the problem presented could have been avoided	What could they have done so that ... did not happen?

The final test package therefore, consisted of the test booklet of pictures and questions, a booklet of Scoring Criteria with model answers and allocated scores, a form for the recording of relevant details for each subject, and a scoring form.

1.1.4. Phase 4: Pilot Study III—the final version of the TATE-ZC

The aim of Pilot Study III was to evaluate the reliability and validity of the test instrument, the TATE-ZC, through the implementation of two item analysis procedures and one inter-scale correlation.

The main aim was realized through the implementation of the following three item analysis processes (Steyn et al., 1994) to ensure that the items and scales of the test were all measuring the same construct of verbal thinking skills:

- Item-test correlation in which the mean of each item was correlated with the mean for the test as a whole.
- Item-scale correlation in which the mean of each item within each scale (thinking skill) was correlated with the mean for that scale. This correlation was calculated for

(i) the mean of an item and the mean obtained for all items in a scale in the same age category;

(ii) the mean of the item and the mean obtained for all ages for that particular scale.

• Inter-scale correlation in which validity was measured by correlating the means for the scales

(i) within each age group;

(ii) for the test as a whole.

1.1.4.1. Subjects

A stratified sample of 292 children from six primary schools in The Valley of a Thousand Hills participated in Pilot Study III. Fifty children in six age categories from Grade 2 to Grade 7 were selected in accordance with the selection criteria and procedure described in Table 3. Fifty children per age category was considered a valid number for the required statistical procedures (Steyn et al., 1994). Six-month age categories were used to limit the age range and to create a buffer between age groups, as a child of 7.11 years may be closer in ability level to a child of 8.1 years than to a child of 7.1 years.

The final sample consisted of six groups of children with 45–51 per group with absenteeism on the day of testing resulting in fewer than 50 in some groups (see Table 8).

Table 8.

Subjects in the item analysis

Age category	Grade	Female	Male	Total no.
7.6–7.11	2	27	24	51
8.6–8.11	3	27	24	51
9.6–9.11	4	26	24	50
10.6–10.11	5	23	25	48
11.6–11.11	6	23	22	45
12.6–12.11	7	26	21	47
Total		152	140	292

1.1.4.2. Test instrument and procedure

The test instrument was the final version of the TATE-ZC produced from Pilot Study II. The procedure described for administration in terms of the orientation followed by individual testing was complied with. All recommendations from Pilot Study II were implemented.

1.1.4.3. Analysis of data

The reliability and validity of the test instrument was calculated using the ITEMAN Conventional ITEM Analysis Programme (as cited in Steyn et al., 1994). The following statistical procedures were applied:

1. Item-test analysis—The Pearson Product-Moment Correlation Method was used to correlate the mean item scores and total mean score for the test as a whole. In this analysis, the mean score of all subjects ($N = 292$) for each item or question (one item) was correlated with the mean score obtained for all subjects ($N = 292$) for the test as a whole (50 items).

Example: The mean for all subjects for question 1 was correlated with the means for questions 1–50 for all subjects.

2. Item-scale analysis—The Cronbach Alpha reliability coefficient was calculated for each age and for the total sample. This item-scale analysis correlated the means of items within each scale (thinking skill), for each age group and for the total group. Each item in the scale was correlated with its specific scale total. This indicated whether children of a particular age responded consistently to the items in that scale.

Example: The mean for question 1 falling into the scale Explaining Inferences was correlated with the mean for the 10 questions falling into the scale Explaining Inferences.

3. Inter-scale correlation—The Pearson's Correlation Coefficient was calculated for each age group, for the test as a whole and for the sample as a whole. These calculations evaluated the extent to which all scales consistently measured the construct of thinking skills.

1.1.4.4. Results and discussion of the item analysis—Pilot Study III

1.1.4.4.1. Item-test correlation

A correlation between an item and the test as a whole that was less than 0.2 was considered invalid and would have to be excluded (Steyn et al., 1994). There were no correlation scores below 0.2, with only two items out of 50 reflecting a weak (0.22 and 0.27) although acceptable correlation. Forty-eight items showed a strong correlation

(0.31–0.64) indicating a high level of validity in these test items. All 50 test items were therefore valid (Steyn et al., 1994).

1.1.4.4.2. Item-scale analysis for five scales (thinking skills)

Correlation scores that were close to 1.0 indicated a strong correlation between the item and the scale for children of a particular age and were represented as an alpha score (Steyn et al., 1994). This would indicate that children of a particular age performed at a similar level for a particular thinking skill. The results are summarised in Table 9.

Table 9.

Item-scale correlation using Cronbach Alpha coefficients correlation across all ages

	7 year category, n = 51	8 year category, n = 51	9 year category, n = 50	10 year category, n = 48	11 year category, n = 45	12 year category, n = 47	Total group, n = 292
Explaining Inferences	0.722	0.648	0.697	0.647	0.561	0.786	0.814
Determining Cause	0.581	0.632	0.582	0.557	0.659	0.632	0.726
Negative Why	0.518	0.582	0.545	0.669	0.603	0.471	0.717
Determining Solution	0.562	0.487	0.368	0.515	0.568	0.388	0.661
Avoiding Problem	0.555	0.399	0.505	0.624	0.530	0.519	0.678

Note: Alpha scores presented.

All correlations were positive and above 0.3 across all age groups. The correlations that were the lowest were for the 9-year-old group ($\alpha = 0.368$) and the 12 year group ($\alpha = 0.388$) in the scale Determining Solutions (see Table 9). The correlations that were the highest were for the 7 year group ($\alpha = 0.722$) and the 12 year group ($\alpha = 0.786$) in the scale Explaining Inferences. Correlations for the group as a whole (all ages) were all above 0.661, with Explaining Inferences having the highest correlation ($\alpha = 0.814$) and Determining Solutions the lowest ($\alpha = 0.661$) (see Table 9).

The alpha scores from the results of the item-scale analysis for each thinking skill indicated that with the exception of the scale Explaining Inferences, the 12 year group did not show the best item-scale correlation (see Table 9). It is expected that the effects of maturation and exposure to schooling would result in the 12-year-old children performing consistently well over all the items in a scale for all scales. However, there was more

inconsistent performance for the 12 year group across 10 items in the scales Determining Cause, Negative Why, Determining Solution and Avoiding Problem, than for the children in the younger age groups. This was reflected in a wide range of scores within the scales, as opposed to scores that were all close to the mean.

Whereas the scores for the 12 year group were generally higher due to maturation, their inconsistency of performance over all the items in four out of five scales could have been due to the fact that children who have reached Grade 7 without failing have a very wide ability level. Their exposure to schooling has not shown uniform benefits. More consistent performance of the younger children may be an indication that improvements in education, e.g. outcomes-based approaches and that the introduction of grade R (readiness class), are beginning to show results (White Paper on Education and Training, 1995). In addition, in the last 4–5 years many more children are being exposed to television in the rural areas with a significantly large number of homes in The Valley of a Thousand Hills having a television set (personal observation from 1997 onwards). Stimulation provided by television may be showing a cumulative effect for the younger children in terms of world knowledge. Children with a greater level of potential, which was not previously tapped when stimulation was minimal, may be showing an increased response to the stimulation provided by television, in terms of a greater awareness of the world and benefits of educational programmes. An increased number of schemata on which they may draw in answering questions may have been stored in this way.

The high incidence of item-scale correlation across scales for all ages, indicated that all the test items are valid for each scale, with a good distribution of scores (Steyn et al., 1994). The range of alpha scores indicating higher correlations for Explaining Inferences, Determining Cause and Negative Why scales could indicate that these thinking skills develop earlier than Determining Solutions and Avoiding Problem which achieved lower alpha scores. Further research is required to confirm this hypothesis.

1.1.4.4.3. Inter-scale correlation

Inter-scale correlation using Pearson's Correlation Coefficient was calculated for each age group between each scale and for the test as a whole, as well as the sample as a whole (Table 10).

Table 10.

Pearson inter-scale correlation ($n = 292$)

Thinking skill	Explaining Inference	Determining Cause	Negative Why Question	Determining Solution	Avoiding Problem
Explaining Inference	1.000				
Determining Cause	0.68421*	1.000			
Negative Why Question	0.70588*	0.70082*	1.000		
Determining Solution	0.65885*	0.69594*	0.66578*	1.000	
Avoiding Problem	0.64630*	0.70248*	0.66447*	0.65873*	1.000

* $p < 0.0001$.

In Table 10, the Pearson Correlation Coefficient indicated that when the group as a whole was considered, the high correlation between each pair of thinking skills (above 0.64) confirmed that all these thinking skills were valid and contributed to the overall construct of ability to explain.

2. Discussion and conclusion

At the outset of this research, an assumption was made, that if the content of a test related to etic/universal intelligence, i.e. knowledge that a rural the child could derive from his/her natural environment, then the child would be given the opportunity to demonstrate best potential in verbal problem solving, and should disprove previous research indicating inferior psychometric test outcomes in rural children (Bentley et al., 1990; Mwamwenda & Mwamwemda, 1991; Viljoen et al., 1994). In addition, a criterion referenced evaluation system based on cognitive skills required for each educational level, as opposed to a norm referenced system, was devised based on the belief that a criterion based reference system was more appropriate and useful in a cross-cultural context (Vaughn-Cooke, 1986). The first step in testing the outcomes of using a culture-fair test would be to develop the test, which is the outcome of the research project described. The second step would be to implement the test.

Many factors contributing to ‘culture fairness’ were considered, yet the one feature that stands out as possibly not being culture-fair, is the very heart of what was being investigated—the child's ability to demonstrate abstract verbal reasoning within a western testing paradigm. Abstract verbalization requires the child to verbalize, and deal with

language in an analytic decontextualized manner (Cummins, 1991), drawing inference and making predictions (Verzoni & Swan, 1995). It is the aspect that would shift the child from one cultural and cognitive thinking style to another (Ong, 1982), or from field dependence to field independence (Tinajero & Páramo, 1997; Witkin & Goodenough, 1981), and may be the key to enabling rural children to achieve a better outcome in a western academic educational system (Alant et al., 1992, Cummins, 1985 and Macdonald, 1990; Mwamwenda & Mwamwenda, 1991). If the TATE-ZC can be used to effectively identify ways in which to achieve more positive academic outcomes by focusing on oral abstract reasoning, as opposed to using the test to highlight comparative limitation in the ability of rural children (Viljoen et al., 1994), tests such as the TATE-ZC could contribute towards better educational achievement for rural children.

The aim of education is to develop literate language skills, which enable an individual to interpret and express complex ideas, thoughts and abstract reasoning. These skills gradually develop from oral language skills (Tannen, 1982), and include thinking and reasoning skills such as those targeted in the TATE-ZC. It has been proposed that there is an urgent need for research into the area of academic language proficiency (Cummins, 1991) in disadvantaged African children in order to address the issue of poor academic progress (Pakendorf, 1996). In administering the TATE-ZC we have the potential to identify with some degree of accuracy the extent to which a rural Zulu-speaking child possesses the thinking skills fundamental to the development of Cognitive Academic Language Proficiency (Cummins, 1985), adequate inferential comprehension of text (Trabasso, 1980 and Van den Broek, 1997) and good potential for academic success. It is thus in the domain of diagnosing a rural Zulu-speaking child as having a genuine language-learning disability, that the TATE-ZC may achieve the full status of 'culture-fair' and 'non-biased'. It is one of a few valid instruments for evaluating rural-Zulu speaking children in relation to standards set by their own community. Thus, the TATE-ZC may contribute to a small set of 'non-biased tests' in the field of speech-language pathology for diagnosing a child as having an authentic language-learning disorder.

While the researcher made every attempt to deal with the many requirements for ensuring that the TATE-ZC was culture fair, which were evaluated in three pilot studies, limitations were noted. The test environment at the different schools varied from a small room to a chair and table at a 'quiet spot' on a veranda, and children who had completed the test may have talked about it to friends. Despite detailed guidelines for scoring, a small number of answers were difficult to categorize. However, the benefit is that the research has been of practical use in producing a tool specifically for rural Zulu-speaking children.

In South Africa today, there is a growing recognition of the need to improve quality of service to Black African first language clients, thus the great need for the development of culturally and linguistically relevant assessment material is evident. Demographics inform us that of the 38 million people living in South Africa today, approximately 33 million speak one of the nine Black African languages, yet, to date there are only a handful of mother-tongue Black African language speech-language therapists. The

critical question is, who will develop these tests, and in terms of current reality, is it acceptable and possible for non-Zulu speaking therapists to fulfil this role?

This research has confirmed that, with appropriate consultation with community members, with close adherence to translation protocol and with an in-depth insight into the critical issues related to multi-cultural testing, cross-cultural test development is possible. The complexity of the process requires commitment to the task and a humble respect for the bilingual Zulu speakers on whom we depend. A statistically valid test such as the TATE-ZC has been produced providing evidence for the fact, that non-Zulu speaking therapists can make a contribution.

In the development of the TATE-ZC many issues of cross-cultural materials development are addressed. South African Speech-language Therapists and Audiologists, as well as therapists throughout the world, are therefore urged to take up the challenge of 'culture fair' and 'non-biased' assessment and treatment, and to develop materials culturally appropriate for the children they test.

References

Alant et al., 1992 E. Alant, H. Tesner and E. Taljaardt, Narrative performance in context: Analysis and implications within a South African context, *Child Language, Teaching and Therapy Monograph* **8** (1992) (2), pp. 188–204.

Baddeley et al., 1995 A. Baddeley, J.M. Gardener and S. Grantham-McGregor, Cross-cultural cognition: Developing tests for developing countries, *Applied Cognitive Psychology* **9** (1995), pp. 173–195.

Bentley et al., 1990 A. Bentley, J. Kvalsvig and R. Miller, The cognitive consequences of poverty: A neo-Piagetian study with Zulu children, *Journal of Applied Cognitive Psychology* **4** (1990) (6), pp. 451–459.

Bernstein, 1970 B. Bernstein, A sociolinguistic approach to socialization: With some reference to educability. In: F. Williams, Editor, *Language and poverty: Perspectives on a theme. Institute for research on poverty monograph series*, Markham Publishing Company, Chicago (1970).

Blachowicz, 1994 C.L.Z. Blachowicz, Problem-solving strategies for academic success. In: G.P. Wallach and K.G. Butler, Editors, *Language learning disabilities in school-age children and adolescents. Some principles and applications*, Macmillan College Publishing Company, New York (1994), pp. 304–322.

Blank, 1970 M. Blank, Some philosophical influences underlying pre-school intervention for disadvantaged children. In: F. Williams, Editor, *Language and poverty: Perspectives*

on a theme. *Institute for research on poverty monograph series*, Markham Publishing Company, Chicago (1970), pp. 62–80.

Bracken and Barona, 1991 B.A. Bracken and A. Barona, State of the art procedures for translating, validating and using psycho-educational tests in cross-cultural assessment, *School Psychology International* **12** (1991), pp. 119–132.

Brinton and Fujiki, 1993 B. Brinton and M. Fujiki, Language skill and the development of social competence in school age children, *Language, Speech and Hearing Services in Schools* **24** (1993) (4), pp. 194–198.

Brislin, 1980 R.W. Brislin, Translation and content analysis of oral and written materials. In: H.C. Triandis and J.W. Berry, Editors, *Handbook of cross-cultural psychology*, Allyn and Bacon, Boston (1980), pp. 389–444.

Carlson and Weidl, 1992 J.S. Carlson and K.H. Weidl, Principles of dynamic intelligence: The application of a specific model, *Learning and Individual Differences* **4** (1992), pp. 153–166.

Code of Federal Regulations, 2001 Code of Federal Regulations (2001). *Title 45, Public welfare of the Department of Health and Human Services, National Institutes of Health, Office for Protection from Research Risks*. Effective as of December 13, 2001. Maryland, USA.

Cummins, 1985 J. Cummins, The construct of language proficiency in bilingual education. In: J.E. Alatis and J.J. Staczek, Editors, *Perspectives on bilingualism and bilingual education*, Georgetown University Press, New York (1985), pp. 209–231.

Cummins, 1991 J. Cummins, Language development and academic learning. In: L.M. Malavé and G. Duquette, Editors, *Language, culture and cognition, multilingual matters series vol. 69*, Longdunn Press, Bristol (1991), pp. 161–175.

Department of Education, 1995 Department of Education, White paper on education and training, Office of the President, Cape Town, South Africa (1995).

Donaldson, 1986 M. Donaldson, *Children's explanations: A psycholinguistic study*, Cambridge University Press, Cambridge (1986).

Fahrmeier, 1975 E.D. Fahrmeier, The effects of school attendance on intellectual development in Northern Nigeria, *Child Development* **46** (1975), pp. 281–285.

Hansen and Pearson, 1983 J. Hansen and P.D. Pearson, An Instructional study: Improving the inferential comprehension of good and poor fourth grade readers, *Journal of Educational Psychology* **75** (1983) (6), pp. 821–829.

Heath, 1992 S.B. Heath, *Ways with words*, Cambridge University Press, Cambridge (1992).

Kemper and Edwards, 1986 S. Kemper and L.L. Edwards, Children's expression of causality and their construction of narratives, *Topics in Language Disorders* **7** (1986) (1), pp. 11–20.

Labov, 1970 W. Labov, The logic of non-standard English. In: F. Williams, Editor, *Language and poverty: Perspectives on a theme. Institute for research on poverty monograph series*, Markham Publishing Company, Chicago (1970).

Labov, 1977 W. Labov, The study of language in its social content. In: W. Labov, Editor, *Sociolinguistic patterns*, University of Pennsylvania Press, Philadelphia (1977), pp. 153–190.

Leedy, 1989 P.D. Leedy, *Practical research* (4th ed.), MacMillan Publishing Company, New York (1989).

Liddel et al., 1994 C. Liddel, J. Kvalsvig, A. Shabalala and P. Qotyana, Defining the cultural context of children's everyday experiences in the year before school.. In: A. Dawes and D. Donald, Editors, *Childhood and adversity. Psychological perspectives from South African research*, David Phillip Publishers, Cape Town, South Africa (1994), pp. 51–65.

Macdonald, 1990 C.A. Macdonald, *Crossing the threshold to standard three in black education*, Humans Science Research Council, Pretoria, South Africa (1990).

Medical Research Council of South Africa, 1993 Medical Research Council of South Africa (1993). *Guidelines on ethics for medical research* (revised ed.). Pretoria, South Africa.

Milosky, 1990 L.M. Milosky, The role of world knowledge in language comprehension and language intervention, *Topics in language disorders* **103** (1990), pp. 1–13.

Mwamwenda and Mwamwenda, 1989 T.S. Mwamwenda and B.B. Mwamwenda, Assessing Africans' cognitive development: Judgment or judgment and explanation, *Journal of Genetic Psychology* **151** (1989) (2), pp. 245–254.

Mwamwenda and Mwamwenda, 1991 T.S. Mwamwenda and B.B. Mwamwenda, African's cognitive development and schooling, *International Journal of Educational Development* **11** (1991) (2), pp. 129–134.

Ong, 1982 W.J. Ong, *Orality and literacy*, Routledge, London (1982).

Pakendorf, 1996 Pakendorf, C. (1996). *The translation of the Peabody Picture Vocabulary Test-Revised into Northern Sotho: A preliminary investigation*. Unpublished

Masters Thesis in Communication Pathology. Pretoria, South Africa: University of Pretoria.

Pena and Quinn, 1997 E.D. Pena and R. Quinn, Task familiarity: Effects on the test performance of Puerto Rican and African children, *Language, Speech and Hearing Services in the Schools* **28** (1997), pp. 323–332.

Piaget, 1956 J. Piaget, The child's conception of physical causality, Routledge and Kegan Paul Ltd., London (1956).

Quay et al., 1980 L.C. Quay, R.A. Hough, M. Mathews and O.S. Jarret, Predictors of communication encoding: Age, socio-economic status and cognitive ability, *Developmental Psychology* **15** (1980), pp. 221–223.

Retief, 1988 Retief, A. (1988). *Method and theory in cross-cultural psychological assessment. Human Science Research Council Investigation into research methodology.* Research report: 6. Pretoria, South Africa: Human Science Research Council.

Steyn et al., 1994 A.G.W. Steyn, C.F. Smit, S.H.C. du Toit and C. Strasheim, Modern statistics in practice, JL van Schaik Publishers, Pretoria, South Africa (1994).

Tannen, 1982 D. Tannen, Spoken and written language: Exploring orality and literacy, Ablex, Norwood, NJ (1982).

Taylor, 1986 O.L. Taylor, The nature of communication disorders in culturally and linguistically diverse populations, College-Hill Press, California (1986).

Tinajero and Páramo, 1997 C. Tinajero and M.F. Páramo, Field dependence-independence and academic achievement: A re-examination of their relationship, *British Journal of Educational Psychology* **67** (1997), pp. 199–212.

Trabasso, 1980 Trabasso, T. (1980). *On the making of inferences during reading and their assessment.* Technical Report No. 157. Centre for the Study of Reading. Urbana, IL: University of Illinois.

Van den Broek, 1997 P.W. Van den Broek, Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In: P.W. Van den Broek, P.J. Bauer and T. Bourg, Editors, *Developmental spans in event comprehension and representation*, Lawrence Erlbaum Associates, Mahwah, NJ (1997), pp. 321–342.

Vaughn-Cooke, 1986 F.B. Vaughn-Cooke, The challenge of assessing the language of non-mainstream speakers. In: O.L. Taylor, Editor, *Treatment of communication disorders in culturally and linguistically diverse populations*, College Hill Press, San Diego, CA (1986), pp. 23–48.

Verzoni and Swan, 1995 K. Verzoni and K. Swan, On the nature and development of conditional reasoning in early adolescence, *Applied Cognitive Psychology* **9** (1995), pp. 213–234.

Viljoen et al., 1994 G. Viljoen, A. Levitt and C. Tredoux, Using the bender gestalt in South Africa: Some normative data for Zulu-speaking children, *The South African Journal of Psychology* **24** (1994) (3), pp. 145–151.

Vygotsky, 1962 L.S. Vygotsky, Thought and language, Massachusetts Institute of Technology Press, Cambridge (1962).

Winne et al., 1993 P.H. Winne, L. Graham and L. Prock, A model of poor readers' text-based inferencing: Effects of explanatory feedback, *Reading Research Quarterly* **28** (1993) (1), pp. 53–65.

Witkin, 1967 H.A. Witkin, A cognitive styles approach to cross cultural research, *International Journal of Psychology* **2** (1967) (4), pp. 233–250.

Witkin and Goodenough, 1981 H.A. Witkin and D.R. Goodenough, Cognitive styles: Essence and origin, University Press, New York (1981).

Wober, 1969 M. Wober, The meaning and stability of the ravens matrices test among Africans, *International Journal of Psychology* **4** (1969), pp. 229–235.

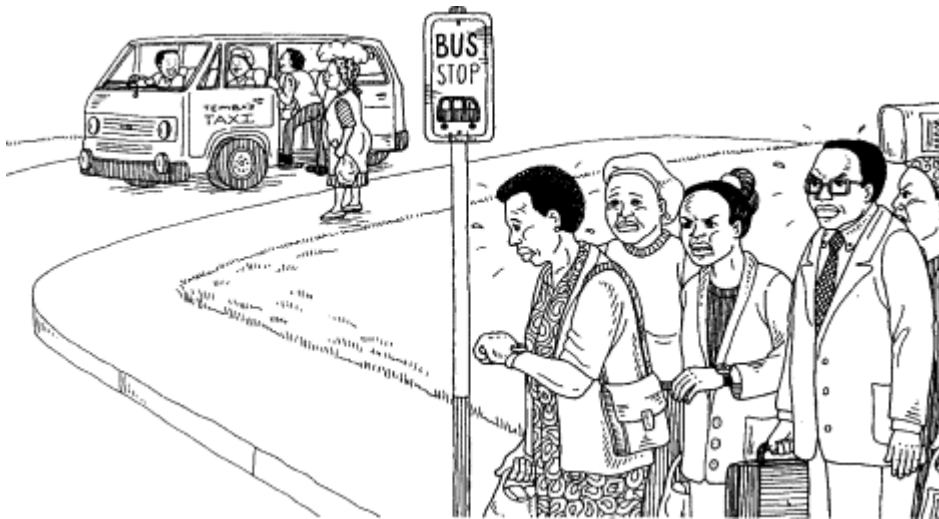
Wood, 1992 D. Wood, Culture, language and child development, *Language and Education* **6** (1992) (2), pp. 123–140.

Zachman et al., 1984 L. Zachman, C. Jorgenson, R. Huisingsh and M. Barrett, Test of problem solving (TOPS), LinguiSystems, Illinois, East Moline City (1984).

Appendix A. Examples of test stimuli

Adapted pictures and questions from the TATE-ZC:

A.1. Example 1:



Question 1:	Bonke labantu bahlele ukuya esitheshini saseThekwini ngebhasi. Yini eyenza bagibele itekisi esikhundleni sebhasi?
	All of these people arranged to go to Durban Station by bus. Why are they riding in the taxi instead of the bus? (Determining Cause)
Question 2:	Lo mshayeli wetekisi akawazi umgwaqo oya esitheshini saseThekwini. Kufanele enze njani?
	The taxi driver does not know the way to the Durban Station. What should he do about it? (Determining Solution)
Question 3:	Umshayeli wetekisi ulahlekile ohambeni lwakhe oluya esitheshini ngabagibeli bakhe bashiywa yisitimela. Bekumele enzeni ukuze angaduki.
	The taxi driver lost his way to the station and his passengers missed the train. What should he have done so he did not get lost? (Avoiding Problem)

A.2. Example 2:

Question 1:	Sazi ngani ukuhti sekuyisikhathi sekhefu esikoleni?
	How do we know it is now time for break (recess) at school? (Explaining Inference)
Question 2:	Omunye wabafana ubephele u-R5.00 esikoleni, kodwa manje eseno- R2.00. Angenzani njengoba imali ilahlekile?
	One of the boys brought R5.00 to school, but now he only has R2.00. What should he do about the money that is lost? (Determining Solution)
Question 3:	Ikuphi okumele lomfana akwenze ukuze ingmlahlekeli imali kusasa?
	What should the boy do so his money does not get lost in the future? (Avoiding Problem)

Self study questions:

1. Which of the following intelligences as distinguished by Carlson and Weidl (1992) is measured in tests of verbal thinking skills?
 - a. Genetic potential intelligence (Intelligence A).
 - b. Actual intelligence derived from interaction with the environment (Intelligence B).
 - c. Intelligence affected by cultural content of the test, familiarity with test material and testing conditions (Intelligence C).
 - d. All of the above.
 - e. None of the above.
2. Which of the following aspects should be considered in the development of testing material for children from different cultural contexts?

- a. Test translation.
- b. Selection of culturally sensitive content.
- c. Adaptation of stimuli to be representative of that culture.
- d. Identification of a suitable method of presentation and administration.
- e. All of the above.

3. Indicate which options does not apply to the process of test translations?

- a. Source to target translation.
- b. Changing the presentation method.
- c. Translation-back-translation.
- d. Blind-back-translation.
- e. A bilingual review committee.

4. It is important to test verbal reasoning abilities of children as:

- a. It is critical to the academic progress of all children.
- b. It deals with language in an analytic decontextualized manner.
- c. It focusses on oral abstract reasoning.
- d. a and b.
- e. All of the above.

5. Results from this study indicate that:

- a. The TATE-ZC is a reliable and valid test for use in South Africa, with children of varying cultural backgrounds.
- b. The test is one valid instrument for evaluating rural-Zulu speaking children in relation to standards set by their own community.
- c. Contributes to a small set of South African based “non-biased tests” in the field of speech-language pathology.
- d. b and c.
- e. All of the above.

