

CHAPTER 4

METHODOLOGY

Where there is no discovery, there is no research.

(Leedy, 1989, p.11)

4.1. INTRODUCTION

Research starts with an unresolved problem, and by asking relevant questions and seeking the answers, we bridge the void in our knowledge. It provides the means of arriving at the answer by studying the facts within the parameters of the scientific method. The research methodology is said to be the way in which we proceed to solve the identified problems (Leedy, 1989). It is through being meticulous in this procedure that we may ultimately reap the rewards of research - being a part of the process of discovery. Chapter 4 describes the research methodology used in this research project, which will lead to some discovery of how rural Zulu-speaking children think and solve problems.

This study embodies the fundamental characteristics outlined by Leedy (1989), that are part of the logical sequence of steps that form the methodology. Firstly, the motivation for this research grew out of the inquisitive recognition that a problem existed that required further investigation (chapter 1). It was then articulated as a specific research goal, to be achieved by dividing the main goal into a number of manageable sub-aims (see 4.2.2). A specific plan of procedure was developed giving rise to a logical methodology (chapter 4). Certain critical assumptions were made about the concepts central to the research, and validated through discussion of already discovered truths, (chapters 2 & 3). The principle of using only measurable data in resolving the research problem was adhered to, and finally the helical process of research was recognized, in acknowledging that this research will provide just the beginnings of our understanding of a complex problem.

4.2. AIMS AND SUB-AIMS

4.2.1 Main Aim

The main aim of this study, is to analyse the verbal solutions of rural Zulu-speaking children to everyday problems, elicited through the use of The Test of Ability to Explain for rural Zulu-speaking Children (TATE-ZC), an adapted and translated version of the Test of Problem Solving (TOPS) (Zachman, Jorgenson, Huisingh and Barrett, 1984).

4.2.2 Sub-aims

The main aim will be achieved through the following sub-aims:

4.2.2.1 Sub-aim 1 (Pre-experimental stage)

The adaptation of the Test of Problem Solving (TOPS) (Zachman, Jorgenson, Huisingh and Barrett, 1984), in terms of language, content, stimuli, presentation and administration, and scoring, to make it as 'culture-fair' and 'non-biased' as possible for the target population.

4.2.2.2 Sub-aim 2 (Experimental stage)

The administration of the Test of Ability To Explain - for Zulu-speaking Children (TATE-ZC) (the outcome of 4.2.2.1) to six groups of rural Zulu-speaking children (N= 292), aged 7-12 years, to obtain a comprehensive sample of 'ability to explain' presented by these children in six years of the primary school phase.

4.2.2.3 Sub-aim 3 (Experimental stage)

Analysis of data obtained on the above test forms the basis of the following procedures:

- (i). To perform an item analysis of the 50 questions presented in the TATE-ZC, to evaluate the validity and reliability of each item.
- (ii). To evaluate inter-translator reliability in the translation of test scripts to ensure that the scores obtained for each child were not influenced by inconsistency in translation of material.
- (iii). To evaluate reliability of the scoring criteria, through intra-scorer and inter-scorer reliability measures.

- (iv). To identify age levels at which statistically significant development has taken place, and present a tentative set of ages for criterion-based evaluation for the development of thinking skills in rural primary school children.
- (v). To correlate the total scores obtained with scores for each sub-test, in order to identify which thinking skill/s most accurately reflect overall level of thinking skills and reasoning.
- (vi). To analyse the mean scores of the sub-tests at each age level to attempt to identify a developmental process in the emergence of thinking skills viz. which thinking skill emerges first and which is most challenging.
- (vii). To compare the extent to which school performance correlates with results obtained on the TATE-ZC.
- (viii). To analyse whether gender differences exist in the development of thinking skills, in the sample as a whole and at each age group.

4.3. RESEARCH DESIGN

The research design is the planned strategy used to deal with the central research problem, which indicates an understanding of the data, the process of acquisition of the data and the interpretation these facts. As people were used as the subjects of this research, ethical standards were considered and upheld, and ethical clearance was obtained (Leedy, 1989).

The research design of this project involves two stages (see table 4.1). First, the pre-experimental stage involves the adaptation of the TOPS and two pilot studies of this new test, The Test of Ability to Explain for Zulu-speaking Children (TATE-ZC). The first pilot study focuses on aspects directly related to adaptation of the test, and the second focuses on testing procedure, reliability of the test instrument and the new scoring criteria and guidelines This is followed by an experimental stage in which the TATE-ZC is administered to a sample of 292 children, and statistical analysis is applied.

Table 4.1
Summary of research methodology

Stage	Research Aim	Function	Description
Pre-experimental	Sub-aim 1	Adaptation of the TOPS	Adaptations made to the content, language, picture stimuli and number of test items used.
		Pilot I	Adapted TOPS was administered to 6 children TOPS probing and scoring procedures used Relevant changes were made in terms of 4 aspects above, as well as probing and scoring The TATE-ZC was produced.
		Pilot II	TATE-ZC was administered to 60 children Aspects related to reliability of test procedure, the test instrument and scoring criteria and guidelines were evaluated. Further relevant changes were made
Experimental 1	Sub-aims 2 & 3	Main Study	TATE-ZC was administered to 292 children Reliability of translation of scripts and of scoring for the main study was evaluated Statistical analysis of the scores was implemented

In the initial stages of the pre-experimental stage, an active participation methodology involved the familiarization of the researcher with the social and cultural environment of the children, through working in a school in a rural area of KwaZulu-Natal. Further, consultation with adult members of the target population enabled the researcher to discuss observations made, and to validate the translations of the test, in order to facilitate cultural validity and accuracy of translation.

A descriptive or qualitative design was considered suitable for the pre-experimental stage of test adaptation, particularly pilot study I, as it focuses on process rather than outcome. It is a design in which the experimenter concentrates on validity, so that data represents an accurate picture of what is being investigated. In the process of adaptation of material, review and change are intrinsic to the procedure. A qualitative design is also considered to be an approach concerned primarily with people and is particularly useful in a cross-cultural context (Retief, 1988). This is because it seeks to understand interpersonal relationships, personal values, meanings, beliefs, thoughts and feelings (Leedy, 1989). This was applicable to both the understanding of the responses of the child as well as to an understanding of the socio-cultural context of the child, which further impacted on the content of the test material. Some quantitative analyses were performed in pilot study II.

A quantitative analytical survey design, which would give sound scientific credibility to the work of the qualitatively produced test instrument was considered suitable for the experimental stage. This methodology tends to focus on reliability and seeks to control natural phenomenon, and test them against the hard facts of reality (Leedy, 1989). Statistical procedures applied at this stage, enable the experimenter to extract the information about the data in which stratified sampling of six different age groups of children are compared in their performance on the test procedures administered.

4.3.1 THE PRE-EXPERIMENTAL STAGE (Sub-aim 1):

There are three phases to the pre-experimental stage (Table 4.1):

- Adaptation of the TOPS (1984)
- Pilot study I
- Pilot study II

4.3.1.1 Adaptation of the Test of Problem Solving (TOPS)(1984)

4.3.1.1.1 The TOPS as a model:

The TOPS, as the model of problem solving, reasoning and explaining used in this research consists of 15 simple line drawings of social situations and 50 questions. There are 3-5 questions for each picture. The questions are based on five identified thinking skills associated with the ability to reason and solve problems. 10 questions randomly presented throughout the test are related to each thinking skill. Each question is scored for linguistic and cognitive accuracy on a three point scale of 0-2 with a maximum possible score of 100. Scores are computed for each thinking skill and as a total score. The scores may be related to an age-equivalent score with a ceiling age of 15.9 years for a score of 100, and a minimum age of 3.5 years for a score of 23. Standard scores and percentiles may also be computed. There is no demonstration item and test time is stated as 20-30 minutes.

A number of aspects of the TOPS were retained in the development of the TATE-ZC.

- The procedure of presenting a picture, asking questions and evaluating the verbal response was felt to be appropriate (chapter 2, 2.7.5).
- The structure of 15 test pictures and 50 test questions was used.
- The analysis of the thinking skills in terms of the five identified skills was felt to be valid. The thinking skills identified, Explaining Inferences (EI), Determining Causes (DC), Negative Why Questions (NW), Determining Solutions (DS) and Avoiding problems (AP), represent different processes of the problems solving procedure and are therefore felt to be a valid reflection of verbal ability to solve problems presented (chapter 2, 2.7).

4.3.1.1.2 Aspects of the TOPS that were adapted

The following aspects of the TOPS were adapted at the pre-experimental stage, in the attempt to make it 'non-biased' and culturally appropriate:

- (i) The content of the test material
- (ii) The language of the test
- (iii) The picture stimuli
- (iv) The number of test items used.

(i) Adaptations made to the content

Adaptation of content is essential to content validity of the test instrument being constructed. It helps to facilitate the accuracy with which the test instrument measures the factors under study (Leedy, 1989). The content, which was reviewed, was the particular social settings of each picture stimulus upon which the questions are based. It is these social settings which had to be adapted to the socio-cultural life experience of rural Zulu-speaking children.

Table 4.2
Procedure for the adaptation of content

Criterion	Procedure	Results	Recommendations
Adaptation of situational Context	<p>This was achieved by consulting an Educational Officer with more than ten years of experience, living and working in Kwadedagendlale or The Valley of A Thousand Hills. Each test item was reviewed in relation to the context of the experiences and realities of rural children in the area. 15 social contexts were reviewed.</p> <p>See Appendix A for details of situations reviewed</p>	<p>Situations retained 66% (10)</p> <p>Situations to be altered: 33% (5)</p> <p>7 Appropriate situations were identified</p> <p>See Appendix B for reasons for changes and new situations identified</p>	<p>The 33% or 5 situations considered to be inappropriate should be replaced with the situations identified as more suitable.</p> <p>Two of the additional situations identified needed to be piloted as potential trial items</p>

(ii) Adaptations made to the language of the test.

This process aimed to produce 57 questions in the Zulu commonly used in the Kwadedangendlale district of KwaZulu-Natal.

As a result of the adaptations to the situational contexts above, a number of new questions had to be devised which were suitable to the new situations. 34 of the original questions were retained, 23 questions were reformulated by the researcher, to fulfill the criteria of the test. These 57 questions were then translated into Zulu.

In order to ensure that the Zulu translators were accurate in their interpretation of the English, the American English (AE) structures and vocabulary were modified to that of South African English (SAE). Adaptations were made to 3 words and 3 sentences (see Appendix C for details).

In order to ensure a systematic and reliable approach to the translation of the test questions, a multi-step approach to translation (Brislin, 1980) was taken (chapter 3, 3.4.1).

Table 4.3

A representation of the 4 stages followed in the translation procedure:

Stage 1: English to Zulu	Stage 2: Zulu to English	Stage 3: Review and Correction	Stage 4: Confirmation
<p>A professor of Zulu at the University of Durban-Westville was consulted, and she supervised a lecturer (a mother-tongue Zulu speaker) in the translation of the questions into Zulu. It was emphasized that colloquial Zulu was required. The purpose of the test and its basic concepts were explained, and the translator was also provided with the questions and the picture stimuli to facilitate the best possible understanding of the contexts.</p>	<p>A senior lecturer in the Zulu Dept at Natal Univ did a back translation of the Zulu questions. This translation was done by an English mother-tongue speaker. This was done because of the complexity of the phraseology when a literal translation from Zulu to English is produced, and it required a translator who would understand the nuance of the Zulu and effectively translate this into comparable English. The translator here was provided with the Zulu questions and the picture stimuli. See Appendix D for examples of this procedure.</p>	<p>The researcher reviewed the English back translation and 12 questions were noted to require clarification or revision. The researcher discussed and clarified the accurate intent of each problematic question with the translator and adjustments were made to the original Zulu translation.</p>	<p>A Zulu-speaking adult living in the Valley of a Thousand Hills reviewed the translation and confirmed its accuracy. These 57 questions were used in the pilot project.</p>

(iii) Adaptation of Illustrations:

In order to facilitate the automatic recognition and comprehension of stimuli, as well as creating a sense of cultural familiarity, comfort and personal identification, (see chap 3, 3.4.2 and 3.4.3), the illustrations were all redrawn with African characters and settings.

An artist experienced in the illustration of primary health care manuals was consulted. She was presented with a description of each illustration and the original test stimuli were 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. 17 pictures were drawn and used in the pilot study.

(iv) Adaptations to number of items presented:

At this stage the only changes made in this area were the presentation of 17 and not 15 pictures, and the posing of 57 and not 50 questions. This was done as it was felt that two

social contexts and their accompanying questions should be piloted for use as training items as lack of such items was recognized as a significant limitation of the TOPS (chapter 3, 3.4.4.6).

This pre-experimental stage to this point, produced a test instrument consisting of 17 culturally appropriate stimulus pictures, and 57 accurately translated questions to be presented in pilot study I.

4.3.1.2 Pilot study I: The adapted version of the TOPS

4.3.1.2.1 Aims of pilot study I

4.3.1.2.1.1 Main Aim of 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 ZS children, and whether there were additional aspects of the TOPS that needed adaptation.

4.3.1.2.1.2 Sub-aims of pilot study I

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

- (i) To test the accuracy of the translation of questions used in the test
- (ii) To test the appropriateness of the social contexts used in the test
- (iii) To evaluate the clarity of the illustrations
- (iv) To review the effectiveness and consistency of the probes given
- (v) To apply and review the scoring procedure of the TOPS.

4.3.1.2.2 Subjects of pilot study I

The 6 children involved in this part of pilot study had to comply with the same criteria as the children in the main study (see 4.3.2.2). This sample, using three of the six age groups, covers the age and gender distribution for the main study. The following table identifies subjects selected.

Table 4.4
Subjects in pilot study I

Subjects	Age	Grade
2 - (1 male & 1 female)	7.6 - 7.11 years	Grade 2
2 - (1 male & 1 female)	9.6 - 9.11 years	Grade 4
2 - (1 male & 1 female)	12.6 - 12.11 years	Grade 7

4.3.1.2.3 Setting of pilot study I

The pilot study took place at a primary school in Kwadedagendiale. Testing was done in a small staff room where electricity was available for use of the tape recorder.

4.3.1.2.4 Procedure for pilot study I

Preparation: One week before testing was scheduled, the researcher visited the school where the headmaster helped in identifying the subjects. The researcher then met with the children, and requested their co-operation. She explained what would be required of them, and emphasized that the testing was confidential and unrelated to school evaluation.

The research assistant: The research assistant (RA), a mother-tongue Zulu-speaker, had participated in many research studies through the Human Science Research Council, and was thus familiar with test protocol. The researcher met with her for one hour, a week before the pilot study, and the project and test were discussed. She was given a test manual in order to familiarize herself with the test.

Administration: Each child was tested individually. The child was made to feel at ease, the RA presented the test instructions, the researcher attached the lapel microphone, then testing began. All 57 questions were administered consecutively. Probing was carried out according to the instructions given in the TOPS manual.

During testing of the first child, the researcher, who sat quietly in the background, left the room to see if the child's responses would alter in her absence. No significant change was noted, thus it was felt that this was not a significant factor.

Testing time varied from 30 minutes for the grade 2 children to 15 minutes for the grade 7 children. At the end of the testing each child was given a chocolate as a reward. The pilot study took place over 2 days (2 hours per session). The adapted version of the TOPS was administered to 3 subjects, one from each age group, per session. Immediately after the session, the research assistant translated the responses into English. The scripts were reviewed by the researcher.

4.3.1.2.5 Results of pilot Study I

The results of, and action generated from pilot study I are presented in Table 4.5.

Table 4.5

A representation of the results of pilot study I in terms of the 5 sub-aims outlined.

Sub-aim	Result	Problems	Outcome
(i) Evaluate accuracy of translation.	All questions were accurately translated and understood.	None	No alteration in terms of translation were required
(ii) Evaluate appropriateness of situational contexts	16 social contexts were found to be appropriate.	1 social context was found to be inappropriate. The context of the pregnant lady needing to get to hospital was omitted. It became evident that children were sent away when mothers went into labour and it was felt this context was culturally and socially inappropriate	It was therefore excluded. 15 pictures were retained as test items. 1 picture was used as a training item.
(iii) Evaluate clarity of illustrations	16 pictures were accurately identified	1 picture was misinterpreted. The children said the pregnant woman holding her stomach was holding a heavy object.	This item was excluded.
(iv) Review probing	Poor consistency of probing	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 TOPS scoring system.	Examples for scoring were inappropriate. Scores were calculated for linguistic and cognitive accuracy. 3 point scale was limited	Due to the change in some contexts, some scoring examples were inappropriate. Scoring criteria for linguistic accuracy were not considered to be valid (see chap 3.6.2.2). The 3 point scale provided a limited range of scores for such a complex problem (see chap 3, 3.6.2.3)	Appropriate examples were provided. Scoring criteria were designed for cognitive content. A 5 point scoring system was developed based on theoretically sound scoring criteria (see chap 3, 3.7, table 3.1)

The recommendations of pilot study I were implemented, and pilot study II followed. This extensively revised and adapted test instrument is now referred to as the Test Of Ability to Explain for Zulu-speaking Children (TATE-ZC).

4.3.1.3. Pilot Study II

4.3.1.3.1 Aims of pilot study II

4.3.1.3.1.1 Main aim of pilot study II:

To administer the Test of Ability To Explain for ZS Children (TATE-ZC) to three groups (20 per group) of rural Zulu-speaking (ZS) children (N=60) of different ages (7, 9, 12years) spanning the primary school phase, to evaluate the effectiveness and consistency of the administration and scoring procedures.

4.3.1.3.1.2 Sub-aims of pilot study II

- (i) To evaluate the subject selection procedure
- (ii) To evaluate the test orientation procedure
- (iii) To evaluate the seating for testing
- (iv) To evaluate the children's responses to the illustrations
- (v) To evaluate reliability of the test instrument through the use of a test-retest trial
- (vi) To evaluate the reliability of the translation of the scripts
- (vii) To evaluate the reliability of probing in terms of consistency, and consistency over time slots
- (viii) To evaluate the scoring criteria and guidelines using the 5 point scale, and the reliability of the new 5 point scale on an inter-scorer reliability trial.

4.3.1.3.2. Subject selection for pilot study II

The subjects were selected from a junior primary school and an adjacent senior primary school in the Kwadedagendlale district of Kwazulu-Natal. These schools were identified as suitable by the Educational Officer (Table 4.2) at The Valley Trust (a community development organization in the area). The schools were judged by her to be representative of rural schools in the area, and were close to the main road.

All subjects for pilot study II met the criteria for subjects to be selected in the main study (see 4.3.2.2).

Table 4.6
The subjects in pilot study II

Gender	Number	Grade	Age Range
Male	10	Grade 2	7.6 - 7.11 years
Female	10	Grade 2	7.6 - 7.11 years
Male	10	Grade 4	9.6 - 9.11 years
Female	10	Grade 4	9.6 - 9.11 years
Male	10	Grade 7	12.6 - 12.11 years
Female	10	Grade 7	12.6 - 12.11 years

As shown in Table 4.6, there was an even distribution in terms of gender and number per grade. 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. This was done as previous clinical experience of the researcher (during a school-readiness testing project), had demonstrated to the that Grade 1 children provided minimal verbal output in a 1:1 testing situation. 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 previously discussed, with cross-cultural testing (chapter 3, 3.3.4), it was felt that Grade 2 children, who had already had one year to adapt to a formal learning context, would perform better in a test situation. Thus Grade 2, Grade 4 and Grade 7 pupils provided the range for the primary school phase.

Table 4.7

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?
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 parents permission for their children's participation.

The table above describes the steps taken in the selection procedure for pilot study II.

4.3.1.3.3 The Test instrument

The test instrument used in pilot study II, is the TATE-ZC, implementing the revised scoring and test administration procedures (see main study 4.3.2.7.4).

4.3.1.3.3.1 The revised scoring criteria for the TATE-ZC

These criteria were revised in terms of the theoretical background presented (chapter 3, 3.7) to create a 5 point scale evaluating accuracy of the cognitive processes involved in the problem solving activities (see Appendix G). A new set of scoring examples was devised in terms of the new situational contexts (see Appendix H).

Scoring of test results includes the following steps:

- Using the 5 point scale (see Appendix G), each answer is scored on a range of 0 - 4 on the score sheet designed for this purpose (see Appendix F).
- Examples for the scoring of each answer on the 5 point scale are used as a guide.
- The maximum score obtainable per category of thinking skill is 40.
- The maximum score for the test as a whole, is 200 which is then calculated as a percentage for convenience.

4.3.1.3.4 Procedure for pilot study II

(i) The Researcher (R) and the Research Assistant (RA)

- The R and RA were present throughout the testing procedure. Both participated in the orientation of each group of children.
- The R carried out the administrative functions of calling the children, recording their names, timing of each child and ensuring that no technical problems occurred.
- The RA was now familiar with the test and test protocol, and administered the test. Specific training at this point occurred in terms of number and manner of prompts permitted (see 4.3.2.4.5).

(ii) The setting

The orientations and testing took place in the library of the Resource Centre adjacent to the junior primary school, where electricity and a comfortable testing environment were available.

(iii) Duration of the fieldwork

Testing took place each school morning for 3 weeks. The R and RA returned two weeks later to complete testing for the reliability study. It took approximately 30 minutes to orientate each group of 10 children, and the test took 15-20 minutes per child.

(iv) Group orientations

Each group of 10 children was called to the library and they were seated in a circle with the R and RA. The R welcomed and thanked the children, and gave a brief description of what they would each have to do. The RA acted as translator. Emphasis was placed on the fact that they

did not have to be frightened and that their performance results would be confidential. This was followed by an 'ice-breaker' game in which each individual, including the R and RA had a chance to say his/her name and simultaneously do an action. This action was to be imitated by the whole group and thereafter an attempt would be made by the group to remember each child's name and action. After this icebreaker the children all sat down together in the circle once again.

The R then explained that in order to further get to know one another and because the children would be asked to speak into the tape recorder, they should practice by each one individually telling a little 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, they were instructed not to talk about the test to each other.

(v) Individual testing

- Each child was seated comfortably next to the RA and the child was put at ease.
- The RA then gave the test instructions to the child and the training item was presented. This served to address issues related to effective 'non-biased' testing to give the child practice with verbal expression and to give an example of a 'good' answer which may contain more than one reason. 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 elicitation of responses a controlled and systematic presentation of probes was given. An approach was adopted whereby, in addition to naturalistic responses of 'mmmm' during an extended monologue, the controlled probes would be used.

Two different probes were used. When the child appeared to have come to the end of the initial spontaneous response, the examiner would give one probe 'mmm'. When it appeared that the child had nothing further to say another probe, 'is that all?' was given, to which the child could respond in the affirmative or by saying 'no' and elaborating on the answer. 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 following two questions:
 - Did you like the pictures?
 - Did you like them a lot, just a little or a medium amount?
- The child was then thanked, given a sweet, reminded not to talk about the test and sent back to class.

(vi) Translation and scoring of the scripts

The tape recordings were then translated and transcribed by the RA (4.3.2.7.3).

Each script was scored by the researcher.

4.3.1.3.5. Reliability Control Measures

Reliability control measures were used to ensure a high degree of accuracy of the test instrument and administration used (Leedy, 1989).

Four types of reliability control measures are included in pilot study II:

- (i) Reliability of the test instrument
- (ii) Reliability of translation of responses
- (iii) Reliability of probing
- (iv) Inter-scorer reliability.

(i) The reliability of the test instrument.

This was evaluated using the Test Re-test Method. A coefficient of stability is obtained by applying the same instrument to the same individuals on two consecutive occasions. In order to limit the drawbacks of this method due to the practice effect and fast changes that occur with children over time, a two week time period was used between the two testing sessions

(Leedy, 1989). 9 randomly selected children, 3 from each age group, were re-tested using the TATE-ZC.

A t test for dependent measures was used to see if the test means obtained by a child was significantly different for the two test performances. This comparison was performed for total scores and for each sub-test.

(ii) Reliability of translation of responses.

A random 20 (33%) of scripts (at least 6 from each age group) were translated from the original audio-tape, by a second translator. Each pair of sentences was then compared and the number of agreements and disagreements per pair of sentences were calculated as a percentage. A pair of sentences was considered to agree if the vocabulary and meaning were the similar and if the sentences were allocated the same score (for examples of agreement and disagreement see Appendix I).

The score was calculated by the formula
$$\frac{\text{Total no of agreements}}{\text{Sum of agreements+disagreements}} \times 100$$

A 95% level of agreement was required.

(iii) Reliability of probing

The two identified probes “mmmm”, and “is that all?” and any other interjections, were inserted into the 20 translated scripts. The scripts with identified probing, were evaluated by the researcher by comparing probing at different time slots during the day to check consistency of administration throughout the day. There was a maximum of three time slots during the morning. The responses of 20 randomly selected children, (at least 6 per time slot) were evaluated. The 2 translators were required to identify all probes present in each answer inserting a (1) for “mmm” and a (2) for “is that all”?

Each of the 50 pairs of sentences could score 2 points to give a total of 100.

- 1 point was deducted if the sentences differed in the probes identified.
- 1 point was deducted if there was inconsistency between the two translators.

Thus if the probes were not present in both sentences, that sentence would score 0 for presence of probes, but 1 for consistency between translators. A total score for probes was calculated and a mean score for each time slot calculated. A minimum of 95% in each time slot was required for probing to be considered reliable and consistent over the different time slots.

(iv) Inter-scorer reliability

2 different scorers scored the translated responses of 20 scripts (33%), with at least 6 per age group, according to the Scoring Criteria (see Appendix G) and Scoring Guidelines (see Appendix H).

4.3.1.3.6 The results of pilot study II

(i). Evaluation of the subject selection procedure

The subject selection procedure was found to be efficient and reliable in identifying appropriate subjects. Steps 1-4 (Table 4.7) are therefore retained for the main study. Step 5, however was problematic. Only a very small percentage 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. In consultation with the headmaster, it was felt that as permission for the research had been obtained from the Education Circuit Inspector, and endorsed by the headmaster, this was sufficient for the testing to be implemented. In the main study no letters were sent to parents, but it was agreed that the various principals would answer any queries from the parents, should they arise.

(ii). Evaluation of the orientation procedure

It was felt that this orientation was most important in reducing apprehension about the test process, familiarising the children with the research assistant, and emphasizing the need for individual responses that could be elaborated with prompting. It also offered a vital opportunity to train the children in responding to testing by letting the children practice verbal expression when telling about themselves, and encouraging them to speak clearly so tape recordings would be audible. It was evident in children of all ages that generating individual

responses was not familiar to them, in that they tended to copy each other's actions and all told about themselves following the exact format of the child before.

It was felt that this orientation should continue to be part of the test process for the main study despite the fact that it was time consuming.

(iii). 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 at certain times, thus this factor may not be impacting on child-adult relations to such an extent in a 1:1 test situation. This seating arrangement is retained for the main study.

(iv). Evaluation of the children's responses to the illustrations

It was felt that it was important to get some feedback as to how the children responded to the illustrations used as test stimuli. A Likert scale was used with choices of neutral, negative, positive or very positive responses 42% of children indicated they like the pictures very much, whereas 58% indicated they like them. The researcher considered the possibility that the caricatured nature of some of the illustrations may have affected the children's responses, and may even evoke a more negative response from adults.

All 16 illustrations were re-drawn for the main study, ensuring the greatest natural likeness to real life pictures as possible (see Appendix J).

(v). Evaluation of reliability of the test instrument in a test-retest trial

The significance of correlation in a test-retest reliability trial was calculated. A t-test for paired samples (n) was calculated to measure significance of correlation between scores of Test 1 and Test 2 (see Table 4.8). A period of 2-3 weeks lapsed between the two tests. A total of 9 children were re-tested, with 3 from each age group.

There was a high and significant correlation at $p < .001$ level of significance, between the 2 tests for total score and for the sub-test Determining Solutions (DS). There was a significant correlation at $p < .05$ level of significance for the sub-tests Explaining Inferences (EI), Negative Why (NW) and Avoiding the Problem (AP). The correlation between Test 1

and Test 2 was not significant for the sub-test Determining Cause (DC), however as this did not influence the significance of the total score, it is not considered to be of critical importance.

Table 4.8

Results of the t-test for paired samples, calculated to measure significance of correlation in a test-retest reliability trial (N=9)

Correlation of Total and sub-test scores on test (Test 1) and retest (Test 2) trials	Correlation for 2 tail significance
Total T1/T2	n =.934 *
EI T1/T2	n =.818 **
DC T1/T2	n =.631
NW T1/T2	n =.867 **
DS T1/T2	n =.940 *
AP T1/T2	n =.914 **

* p<.001 ** p<.05

These results indicate 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 repeated by students, i.e. the test is consistent in measuring level of skill.

(iv) Calculation of inter-translator reliability

The inter-translator reliability was calculated by scoring for agreement and disagreement between translators, and a percentage of agreement was then calculated. Percentage agreement for the total sample at each age level was > 95% (Table 4.9).

Table 4.9

Inter-translator reliability calculated as a %

Age Group	% Agreement between the two translators
7 N= 7	96.29 %
9 N=7	95.43 %
12 N=12	96.67 %
TOTAL N=20	95.13%

This indicates a sufficiently high level of reliability between the two translators, and that the translated responses were therefore accurate.

In spite of the above result, an inter-translator reliability measure is considered necessary for the main study as there are three research assistants, and it is important to ensure that scores obtained by students are not affected by inconsistency in translation among them.

(vi) Evaluation of reliability of probing

Reliability of probing was calculated to evaluate whether there had been consistency of probing over the 5 hour testing period each day, and over the test period as a whole. The daily testing time was divided into 1 ½ hour slots with a school break. Probing was calculated for at least 2 children from each age group, over each of the time slots, over the full test period (Table 4.10).

Two probes were counted. One “mmm” (coded as 1), when child appeared to come to the end of her/his verbal output, and “is that all?” (coded as 2), at the end of each response.

The probes were marked onto the transcription of verbal response from the audio-tapes. Each of the two probes had to be present in the responses to score a point, but slight variation in position of the probes in the sentence was permitted in relation to slight differences in the translations. A percentage of consistency was calculated between the two raters.

Consistency of probes was scored to be > 95 %, indicating a high level of consistency in the test administration procedure. This result indicates a high level of reliability of the test administration process measured.

These results indicate that the instructions for probes and implementation of consistent probing, was easily achieved, and it is therefore unnecessary to repeat this process in the main study.

Table 4.10

Reliability for the consistency of probing, calculated as a %, across 3 X 2hour time slots, for each age group and for the total sample

Age Group and N	Time slot 1	Time slot 2	Time slot 3
7 N=7	97.67% N=3	98 % N=2	99 % N=2
9 N=7	99.5 % N=2	95.33 % N=3	96.5 % N=2
12 N=6	95.5 % N=2	97.5 % N=2	96 % N=2
TOTAL N=20	97.55 % N= 7	97 % N=7	97.16 % N=6

Key:Time slot 1 -9:00- 10:30

Time slot 2 -11:00- 12:30

Time slot 3 -12:30- 14:00

(vii) Evaluation of scoring criteria- Inter-scorer reliability

A Pearson's Correlation Coefficient (r) was calculated, in which inter-scorer reliability for scorer1 (S1) and scorer 2 (S2) was measured for the total scores and each of the sub-test scores (Table 4.11). Both scorers were lecturers in speech pathology, one of them being the researcher. The scripts of 20 subjects, at least 6 per age group, were scored by each of the scorers independently using Scoring Criteria (Appendix G) and Scoring Guidelines (Appendix H) devised by the researcher.

Table 4.11

Pearson's Correlation Coefficient in which inter-scorer reliability for scorer 1 (S1) and scorer 2 (S2) was measured for the total scores and for each of the sub-test scores

	Total S2	EI S2	DC S2	NW S2	DS S2	AP S2
Total S1	$r = .9662 *$					
EI S1		$r = .9185 *$				
DC S1			$r = .8785 *$			
NW S1				$r = .9318 *$		
DS S1					$r = .8823 *$	
AP S1						$r = .9031 *$

* $p < .001$

Inter-scorer reliability is high at the $p < .001$ level of significance indicating reliability and accuracy of Scoring Criteria and Scoring Guidelines.

Due to the critical importance of accurate scoring, it was felt that inter-scorer reliability measures should form part of the main study to further validate the scoring criteria for the TATE-ZC.

4.3.1.4 Summary of the outcomes for the pre-experimental phase (sub-aim 1)

A summary of the outcomes of the pre-experimental stage is presented in Table 4.12. The outcomes are described in terms of:

- Aspects of the TOPS that were retained
- Adaptations to the TOPS that were made
- Testing procedure and administration
- Reliability measures.

Table 4.12
Summary of Outcomes for the pre-experimental stage

Area of Outcome	Aspect	Description
1. Aspects of the TOPS that were retained	1.1 Basic test procedure	<ul style="list-style-type: none"> • Procedure of presenting a picture and asking related questions
	1.2 Thinking skills targeted	<ul style="list-style-type: none"> • The 5 thinking skills identified in the TOPS as a comprehensive measure of pragmatic verbal problem solving, viz Explaining Inferences, Determining Cause, Negative why , Determining solutions, Avoiding the Problem
	1.3 Number of pictures and questions in the main test	<ul style="list-style-type: none"> • 15 pictures and 50 questions, with 10 questions per thinking skill
2. Adaptations made to the TOPS	2.1 Content of pictures used as stimuli	<ul style="list-style-type: none"> • 10 of the original social contexts were retained • 5 new social contexts were devised for the test • 1 new social context was used as a training item
	2.2 Translation	<ul style="list-style-type: none"> • Translation of the test into Zulu according to the procedure outlined by Brislin (1980) proved to be reliable.
	2.3 Illustrations	<ul style="list-style-type: none"> • All illustrations used as stimuli were drawn using African characters and settings • The first set of illustrations tended to have a caricatured quality and needed revision • A set of life like pictures was then produced for the final test
	2.4 Scoring	<ul style="list-style-type: none"> • Scoring excluded linguistic performance, and focused on pragmatic reasoning only • A 5 point scoring scale replaced the 3 point scoring scale • A set of scoring guidelines was devised for each of the 5 scores • A new set of culturally appropriate examples for allocating a score to an answer was drawn up for each question

3. Test procedure and administration	3.1 Subject selection	<ul style="list-style-type: none"> Using teachers to identify subjects was found to be a reliable way for selecting subjects Letters of permission to parents was replaced by permission from headmasters and their preparedness to field queries from parents
	3.2 Orientation procedure	<ul style="list-style-type: none"> This was considered to be an important part of training children on how to perform in a formal test situation
	3.3 Seating	<ul style="list-style-type: none"> Seating the children adjacent to the RA was fine, although the children did not appear to avoid eye contact with the adult
	3.4 Probing	A structure probing procedure replaced the more informal approach to probing found in the TOPS
4. Reliability measures	4.1 Test-retest measure	<ul style="list-style-type: none"> Evidence for the test instrument as reliable and valid was shown
	4.2 Reliability of translation of test scripts	<ul style="list-style-type: none"> Reliability was shown to be >95% Inter-translator reliability is considered necessary for the main study as 3 different translators are used.
	4.3 Reliability of probes	<ul style="list-style-type: none"> A high level of reliability of probes was noted This was also shown to be consistent throughout the testing period for each day Instructions for probing were, therefore, easy to comply with
	4.4 Inter-scorer reliability	<ul style="list-style-type: none"> A high level of agreement was shown to exist Inter-scorer reliability will be carried out for the main study, to confirm reliability for the main study

4.3.2 THE EXPERIMENTAL STAGE (sub-aim 2 & 3): THE MAIN STUDY

4.3.2.1. Aims of the main study

Sub-aim 2 (see 4.2.2.2) & sub-aim 3 (see 4.2.2.3) state the aims for the experimental stage, viz, to administer the TATE-ZC to a sample of rural Zulu-speaking primary school children and analyse the performance of the children on the TATE-ZC.

4.3.2.2. Selection of subjects for the main study

Subjects were selected on the basis of a stratified purposive sample. Stratification occurred at six age groups and in six different grades. The purposive sample attempted to account for uncontrolled variables through stated criteria for selection.

The children who were selected for the main study had to meet the following criteria:

- **Age:** The chronological age of each of the children had to fall within the specified 6 month age range for the grade.
- **Gender:** An equal distribution of male and female was needed to control for social and developmental factors that may influence performance (Babbie, 1992).

- **Academic Record:** The children had to have a record of no failure or repetition of a school year, to exclude low cognition as a factor.
- **Sensory Impairment:** The children had to be identified by teachers as appearing to have no visual or auditory impairment which may impact on development and test performance.
- **Social and Medical History:** Children had to be identified by teachers as having no social or medical problem which may impact on development and test performance.

4.3.2.3 Selection of schools for the main study

Six schools in the Valley of a Thousand Hills participated in this study - 4 primary schools (Grade 0-7), 1 junior primary schools (Grade 0-4) and 1 senior primary school (Grade5-7). The schools were selected on the basis of the principals' willingness to participate, and their access to roads. All the schools were within a 10 km radius, thus drawing on children in similar contexts, giving recognition to the fact that even within a disadvantaged community, there is a range of 'poverty' and 'relative wealth'. The subject selection was based primarily on identifying children who complied with the criteria for selection, and who fell within the 6 month age range identified per grade. There was, therefore, no attempt to identify a critical number of children per school. Pupil:teacher ratio's varied from approximately 30-40:1 (Table 4.13).

Table 4.13
Selected schools and teacher:pupil ratios

School	Total No of Children	No of teachers	Pupil Teacher ratio
1 Senior primary	440	11	40:1
2 Junior primary	501	15	33:1
3 Primary	797	20	40:1
4 Primary	584	14	42:1
5 Primary	564	14	40:1
6 Primary	586	20	29:1
Total	3472	94	37:1

The description of the schools that follows has been included to contextualize the learning environment of rural Zulu-speaking children. All schools were situated on sand roads, with a high pupil:teacher ratio (see Table 4.13). All schools had minimal resources, with most

classrooms having only desks for children and a table and chair for the teacher. There was minimal evidence of teaching materials or wall charts. There was electricity in the classrooms, but this was sparingly used, thus most classrooms were fairly dark and sparse. Children had their own exercise books and some text books were used. All schools were included in the government nutrition programme, thus children were receiving some nutritional supplements. It was noted that the majority of children made a small purchase, e.g. a lollipop, a small packet of chips or even a small packet of biscuit crumbs from the bottom of the biscuit boxes, from the local women who sold food and sweets outside each school at break-times and after school. Attempts had been made by some schools to create gardens around the classrooms to improve the atmosphere of the school, and all had a rough sandy sports field adjacent to the school. All schools were surrounded by expansive and beautiful rolling hills, with clusters of traditional homes as well as simple more western homes dotted on the hills and along the roads, which resulted in children walking long distances to school in many cases.

4.3.2.4 Research Assistants

Three female Zulu-speaking research assistants (RA) were involved in the data collection. All the RA's work in this capacity for the Child Development Unit at The University of Natal, and are therefore experienced in the process of scientific data collection. Two training sessions took place. In the first the researcher gave the RA's some theoretical background to the study, and explained the process up to and including pilot studies I and II. The second session involved training in terms of the subject selection procedure, the orientation session, the test itself, test administration with an emphasis on probing technique, and discussion on how to monitor and control number of subjects per age group.

Each RA was given a testing kit including: a copy of the TATE-ZC, a tape recorder, a lapel microphone, extra batteries, audio cassettes, the front cover page for each test script, to record each child's details, school, date of testing and RA involved, an exam pad, stationery and packets of sweets as rewards for the children.

Using a hired vehicle, the RA's moved independently from school to school. Besides on-going telephonic communication to deal with any arising queries, the researcher monitored

the data collection on two occasions. Once in the early phase of setting the research up and once during the data collection to ensure that data collection was reliable.

4.3.2.5 The test instrument used: The Test of Ability To Explain for Zulu-speaking Children (TATE-ZC)

The TATE-ZC consists 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, are adjacently arranged. 1 picture is a demonstration item, and 15 pictures are the test stimuli. 3 of the questions relate to the demonstration item, and 50 are test items. There are 2 - 4 questions per picture.

Each question relates to one of the five thinking skills identified. The thinking skill linked to each picture is dependent on the context of the picture. There is no specific pattern in which thinking skills are targeted. Thus the total of 50 questions is made up of 10 questions per 5 thinking skills, which are randomly presented according to the possibilities offered by the different contexts. For a detailed breakdown of which pictures and questions are associated with each thinking skill (see Appendix E).

The thinking skills categories tested are those involved in the problem solving process (TOPS, 1984). The following table presents a brief description of each one and the question it poses.

Table 4.14
Description of the five thinking skills (TOPS, 1984)

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.....?
The Negative why	The child is asked to give a reason why one would not behave in a certain way.	Why wouldn't.....? Why aren't.....?
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 the 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?

In addition to the test booklet of pictures and questions, a form for the recording of relevant details for each subject, and a scoring form (see Appendix F) were also prepared.

4.3.2.6 Equipment

The following equipment was used to ensure clear audio-tape recordings of the answers presented by the children.

- 3 x Philips D6280 computer compatible cassette recorders were used..
- 3 x AIWA lapel microphones.

4.3.2.7 Procedure for the main study:

4.3.2.7.1. Subjects in the main study

292 children participated in the main study. The RA's followed steps 1 - 4 of the procedure outlined in pilot study II (see Table 4.7), in the selection of subjects at the different schools. The target was 50 children per age group, with an equal gender distribution. Due to logistical and practical problems experienced, e.g. listing girls in the boys list as it was sometimes difficult to identify from the name, tape recordings being inadequate in a few instances, the following subjects participated as subjects in this study (Table 4.15).

Table 4.15
Subjects in the main study

Age	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

As stated, it was not intended that a critical number of children be selected from each school. Emphasis was placed on age and passing the selection criteria. Table 4.16 describes the distribution of children per school and grade.

Table 4.16:

Distribution of subjects per school

School	Grades	Female	Male	Total
1	Grades 5-7 10-12 years	35	35	70
2	Grades 0-4 5-9 years	24	24	48
3	Grades 0-7 5-12 years	68	44	112
4	Grades 0-7 5-12 years	13	13	26
5	Grades 0-7 5-12 years	9	11	20
6	Grades 0-7 5-12 years	3	13	16
Total		152	140	292

4.3.2.7.2. Data Collection

The data was collected over a period of 6 weeks, with each RA testing approximately 80-100 children. The RA's travelled to the schools on a daily basis and testing took place from 9:00am – 2:00pm each day. Testing was interrupted from time to time by sports events, school functions, choir competitions etc.

In five out of the six schools testing took place in a room where the child and RA were seated at a table. In one school testing took place at a chair and table on the veranda of the school, but the setting was generally quiet.

Group orientations followed the procedure outlined in pilot study II (4.3.1.3.4 , (iv)) and were followed by individual testing, following the procedure outlined in pilot study II (4.3.1.3.4, (v)). Children were not asked to comment on whether they liked the pictures in the main study.

On entry into the testing setting, each child was welcomed by the RA, thanked for participating and informal interaction took place for a few moments to settle the child. Thereafter, the following instructions were given by the RA:

- I am going to show you some pictures and then I will ask you some questions about the pictures.
- Nobody except myself and the R will be hearing your answers.

- This is not like a school test because there is no right and wrong answer. I am just interested to hear what you think and have to say about the pictures.
- You must give the very best answer that you can.
- I am going to tape-record what you say so that I don't have to try and write it down as you speak. Please speak loudly so we can hear you clearly on the tape recorder.
- When you have completed the test and go back to your classroom, I would like to ask you not to talk about the test to the other children.

These instructions were followed by the presentation of the training item.

- Lets look at the first picture together.
- Can you tell me what you see?
- Good. Now you can answer the questions. You can tell me as much as you like.

Each training question was individually presented. If the child gave a good answer, the RA praised the child and stated clearly what was good about the answer, encouraging the child to give more than one answer if desired. If the child gave a poor answer, the RA asked facilitating questions to elicit the answer and/or presented the child with an appropriate answer, so that by the end of the training item the child understood how to answer a question in the best possible way. Three questions were asked, thus the child went through the above procedure three times.

On completion of the training item, the RA proceeded as follows:

- Now let's start the test. You must try to answer the questions as I have explained to you.

The lapel microphone was attached, and the test administered in one sitting, with all 50 questions being administered to each child. Testing took 15-25 minutes per child, depending on the age and competence of the child. Finally the child was again thanked, given a sweet and again reminded not to talk about the test.

During testing, the two identified probes, 'mmmm' to encourage further elaboration, and 'is that all?' when child appeared to have completed the answer, were consistently presented for each question.

4.3.2.7.3 Translation and transcribing of the data

After testing was complete, the RA's went through each cassette, translating and transcribing what the children had said, by listening to each answer then writing it down in English. Each script was concurrently coded for the relevant probes (1) or (2). The cover page for each child was attached to the translated script, and it was then ready to be scored. On completion of all translation and transcription, 11 subjects were randomly selected, and each of the three RA's translated and transcribed those scripts as has been described above, which would form the basis for the inter-translator reliability test, to be carried out at a later stage.

4.3.2.7.4 Scoring of scripts

Using the scoring criteria (Appendix G) and guidelines giving examples for each score for each question, the researcher scored 30 scripts. After a period of 6 weeks, these 30 scripts plus all the remaining scripts were scored by the researcher resulting in an accumulated 125 hours of scoring or 25 minutes per script. The 30 scripts that were re-scored formed the basis of the intra-scorer reliability test.

An additional 29 (10%) randomly selected scripts were photocopied, and with the test booklet, the scoring criteria and guidelines, given to a second scorer. The second scorer is a speech pathology lecturer, with extensive experience in test administration and scoring. The results of her scoring were used as the basis for the inter-scorer reliability test.

In the scoring of each question, the scorer read the answer presented by the child, then reviewed the options provided in the scoring guidelines. An attempt was made to identify the level of complexity of the answer. Was the answer wrong or irrelevant? Did the child present one concrete fact in the answer, identified in the illustration? Did the child present two concrete facts indicating an awareness of multiple causation or did the child present psychological causality or intent showing use of abstract thinking? Did the child give a full answer in which the critical relationship or factor was clearly identified and clearly expressed. Scores from 0-4 were accordingly allocated. Some answers may be obscure and require repetitive review until the appropriate score can be identified. In such cases references is made to the examples presented in Appendix G and the child's answer compared with the range of answers in the attempt to identify the correct level.

When each question had been analysed and scored, scores of each subject were calculated out of 200, which was converted to a percentage for total score (50 questions) , as well as for each of the five sub-tests or scales (5 x 10 questions).

4.3.2.7.5 Collection of academic results

When all the tests had been scored, the RA's returned to the schools and requested permission to record academic results for each child. Permission was granted and the results of the July exams or tests were used. The subjects targeted were Zulu or literacy, as the children's first language and language used for semantic contextual reasoning, and mathematics or numeracy as the subject in which logico-deductive reasoning would take place. Exact marks are presented where available. Where symbols (A-E) or ratings (1-5) were used in the lower grades, consistent percentage values were allocated for statistical procedures.

These scores form the basis of the correlational analysis, to evaluate the relationship between pragmatic reasoning skill and academic performance.

4.3.2.8 Data Analysis and Statistical Procedures for the main study:

All statistical procedure for the main study were done using SAS.

Statistical procedures presented in Table 4. 16 were implemented to evaluate:

1. Validity and reliability of the test instrument
2. Reliability of the testing procedure
3. Analysis of the data

Table 4.17

Table of statistical procedures used in the experimental stage

Process	Procedure	Application	Aim
1. Validity and reliability of the Test Instrument	ITEMAN Conventional Item Analysis Program (1989)	1.1 <u>Item- Test Analysis</u> - The mean of all subjects on an item is correlated with the mean for all subjects for the total score The Pearson product moment correlation between the item scores and total score for that test, is calculated.	Sub-aim 3, No (i)
		1.2 <u>Item-scale Analysis &The Cronbach Alpha reliability coefficient</u> - This is calculated for each age and the total population	Sub-aim 3, No (i)
		1.3 <u>Inter-scale correlation</u> : Correlation of scales at different age groups	Sub-aim 3, No (i) No (v)
2. Reliability of Test procedure	Friedman Procedure using the BDMP programme. This is a non- parametric test for paired samples* Pearson's Correlation Coefficient*	2.1 <u>Reliability of Translation</u> : A comparison of % error between 3 Translators on the same 5 scripts.	Sub-aim 3, No (ii)
		2.2.1 <u>Reliability of scoring procedure</u> : Inter-scorer correlation in which scorer 1 and scorer 2 each scored 26 scripts	Sub-aim 3, No (iii)
		2.2.2 <u>Reliability of scoring procedure</u> Intra-scorer correlation in which scorer1 scored the same 30 scripts twice, with 6 week break between score 1 and score 2)	Sub-aim 3 No (iii)
3. Analysis of the data	ANOVA with Scheffe's Procedure, for definite difference in means* Pearson's Correlation Coefficient* Descriptive statistics- Comparison of mean scores*	3.1 <u>Significant difference for age</u> : Significant difference in means between the different age groups for the total score	Sub-aim 3 , No (iv)
		3.2 <u>Significant difference for age</u> : Significant difference in means between the different age groups for each scale or thinking skill	Sub-aim 3 , No (iv)
		3.3 <u>Significant difference for gender</u> : Significant difference in mean scores between the two genders for the test as a whole and for each scale	Sub-aim 3, No (viii)
		3.4 <u>Correlation between TATE-ZC and academic performance</u> . Correlation of 3 sets of scores, TATE-ZC, Zulu/literacy and mathematics/numeracy	Sub-aim 3 , No (vii)
		3.5 <u>Comparison of mean scores for each age group and thinking skill</u> Comparison of mean scores to detect developmental process	Sub-aim 3, No (vi)

* = (Steyn, Smit, du Toit and Straheim, 1994)

4.4 SUMMARY

Chapter 4 presents a discussion on the methodology used in this study. The aims and the sub-aims are stated, followed by a description of the experimental design. Procedures at the pre-experimental stage included the process of adaptation of the TOPS and two pilot studies of this adapted version. The results of this stage produced the TATE-ZC, which was the test instrument used in the experimental stage. The description of the experimental stage, included the aims, the procedures, method of analysis and statistical procedures. Statistical procedures were described in terms of parametric and non-parametric statistical procedures, reliability measures and descriptive statistics.