CHAPTER 5

RESULTS AND DISCUSSION OF RESULTS

5.1 INTRODUCTION

The information presented in this chapter describes the results obtained from the administration of The Test of Ability to Explain for Zulu-speaking Children (TATE-ZC) to 292 primary school children aged 7-12 years, in the rural area of The Valley of a Thousand Hills, KwaZulu-Natal.

The results will be presented in three sections.

- First, an analysis of the reliability and validity of the test instrument reviewing the item analysis processes, is presented.
- Second the reliability of the test procedure and administration, evaluating reliability of translation procedures and reliability of the scoring criteria is presented.
- Third, an analysis of the data identifying significant differences in the performance of children of different ages in verbal problem solving, correlation of scores for the different thinking skills, an analysis of the relationship between performance on the TATE-ZC and academic performance, and analysis of gender differences is presented.

The results are analysed in terms of the test as a whole, with 50 items, and per five scales or thinking skills of 10 items each. They are analysed for the sample or group as a whole, N=292, and per 6 age groups (7-12 years) of 45-51 children, as well as by gender. Scores of 0-4 points per item or question are referred to giving a total of 200 test points. All points have been converted to percentages for calculations. Results are presented graphically, in terms of statistical procedures and a description and discussion of results where indicated. Figure 5.1 provides an overview of the analysis of results for points (i) – (viii) of sub-aim 3 (experimental stage).

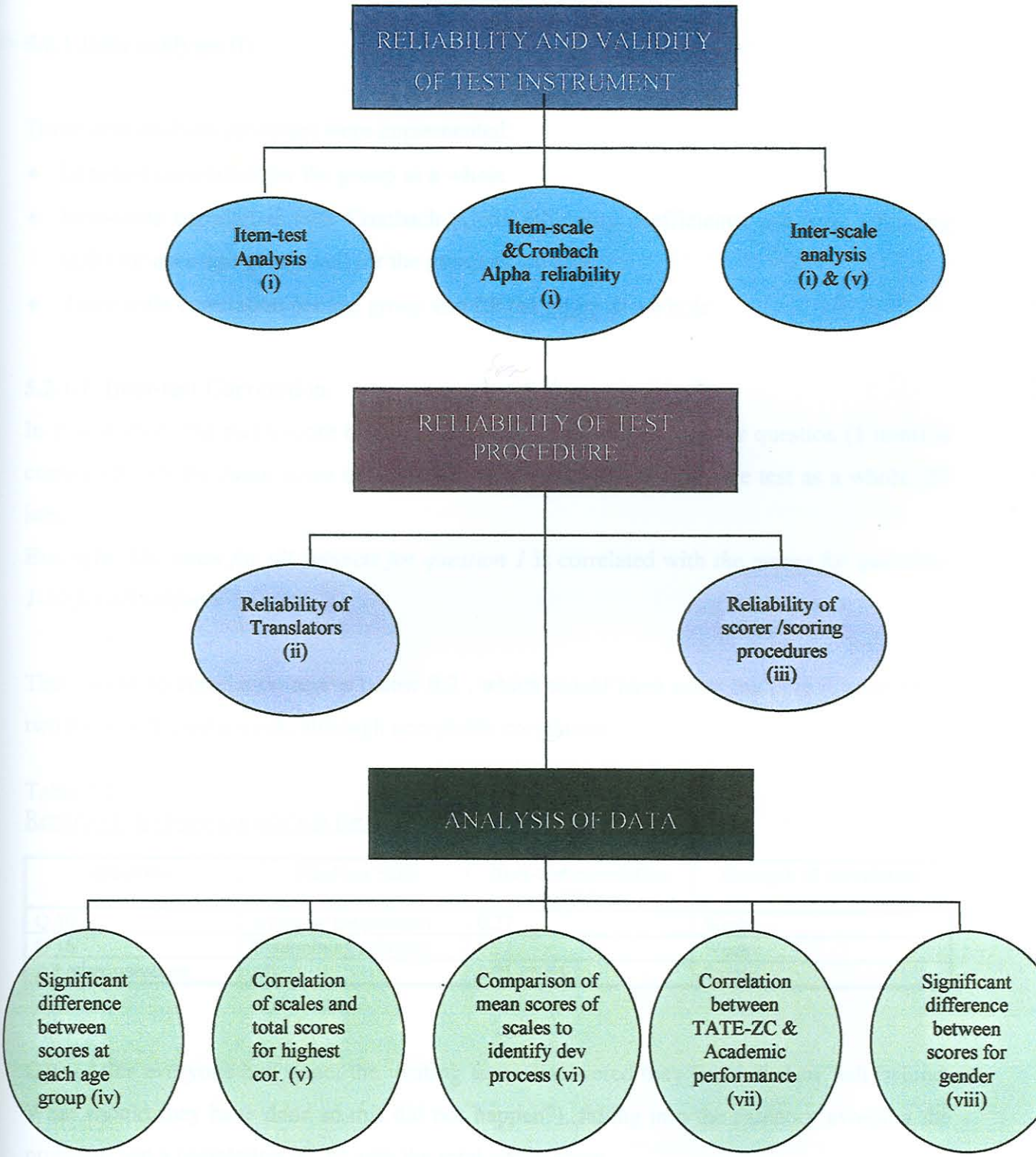

Figure 5.1 Eight areas of analysis of data described in sub-aim 3 (experimental stage)

5.2 RELIABILITY AND VALIDITY OF THE TEST INSTRUMENT

5.2.1 Item analyses (i)

Three item analysis processes were implemented:

- Item-test correlation for the group as a whole
- Item-scale correlation with Cronbach Alpha reliability coefficients, per scale (thinking skill) for each age group, and for the group as a whole
- Inter-scale correlation per age group and for the group as a whole

5.2.1.1 Item-test Correlation

In this analysis the mean score of all subjects (N=292) for each item or question (1 item) is 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* is correlated with *the means for questions 1-50 for all subjects*

There were no correlation scores below 0.2 , which would have made the item invalid. Only two items reflected a weak, although acceptable correlation

Table 5.1
Results of item-test correlation for the group as a whole

Question	Thinking Skill	Item-test correlation coefficient	Strength of correlation
Q 10	Avoiding the problem	0.27	Weak
Q 16	Determining solutions	0.22	Weak
All other questions	All	0.31 - 0.64	Strong

Q10 (After everyone had gone, the visiting team discovered they had left their ball behind. What should they have done so this did not happen?), falling into the category avoiding the problem, had a correlation of .27 with the total group score

Q 16 (The painter does not like the colour of the paint. What should he do about it?), falling into the category determining solutions, had a correlation of .22.

This indicates that of all the 50 items, the subjects as a group, achieved scores that were more different from the total mean score for these 2 questions than for all the other questions. For

the 48 other questions, the means for each question correlated well with the mean for the test as a whole.

When the scores for subjects as one group (N=294) were correlated within the five scales, these two items once again showed a weak but acceptable correlation. This clearly indicates that although these two items are not invalid in themselves on either of the measures, and do not render the test as a whole invalid in any way, they do need to be scrutinized at a qualitative level to ascertain why relatively low correlation were obtained. Discussion of these two questions is included in the item-scale analysis (5.2.1.2). These results indicate reliability and validity of each item in the test instrument.

5.2.1.2. Item-scale Analysis, for 5 scales (thinking skills) with Cronbach Alpha coefficients

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 indicates whether children of a particular age respond consistently to the items in that scale

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

An overview of the results indicates that with the exception of the scale Explaining Inferences, the 12year group did not show the best item-scale correlation. This indicates that although we would expect that by 12years, the children would perform consistently well over all the items in a scale, there was more inconsistent performance for the 12 year group across 10 items in a scale, than for the younger age groups.

This may be due to the fact that, whereas the scores for the 12year group were generally higher due to maturation, their inconsistency of performance over all the items indicated either, great variability of performance amongst the 12 year group or problems with the items. As other age groups showed good item-scale correlation we may conclude that the problem lies within the 12year group. Children who have reached grade 7 without failing have a very wide ability level. More consistent performance of the younger children may be an indication that improvements in education, e.g. outcomes-based approaches and the introduction of grade R (readiness class), are beginning to show results. 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. This 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. This may be supplying them with an increased number of schemata on which they may draw in answering questions.

The high incidence of item-scale correlation across almost all scales for all ages, indicated that all the test items are valid for each scale, with a good distribution of scores.

Results of the item-scale correlation per scale follows, with a discussion of those items within each scale where correlation was weak for a particular age group, i.e. falling between 0.2- 0. 3, or low, i.e. being less than 0.2. Each table also reflects the Cronbach Alpha reliability co-efficient per age and scale.

Scale 1: EXPLAINING INFERENCES

Correlation per question was high (>0.3) across the age groups indicating consistency of test items in evaluating inferencing skills for all ages. There was a very high correlation for the total group (N=292) on all items in the scale, with a high Cronbach Alpha reliability co-efficient of 0.814 for the total group. On reviewing the correlation per age group, there is not an absolutely consistent pattern of increasing consistency with age, indicating that younger children may in fact perform more consistently over all the items than the older ones. However, when the scores for all the children (N=292) were grouped, a better distribution of scores and hence higher correlation resulted.

Table 5.2
Item-scale correlation per age group and for the total group for the scale explaining inferences

	Age 7 N=51	Age8 N=51	Age9 N=50	Age10 N=48	Age11 N=45	Age12 N=47	Total Group N=292
Q1	.63	.44	.60	.11	.41	.51	.61
Q2	.19	.36	.38	.61	.63	.67	.63
Q3	.72	.61	.55	.56	.40	.79	.69
Q4	.52	.62	.49	.56	.42	.69	.67
Q5	.46	.24	.50	.51	.31	.69	.60
Q6	.65	.45	.52	.47	.67	.60	.59
Q7	.54	.63	.59	.41	.42	.41	.59
Q8	.61	.53	.51	.48	.38	.61	.55
Q9	.46	.59	.52	.55	.41	.53	.60
Q10	.50	.46	.57	.55	.49	.32	.64
Alpha	.722	.648	.697	.647	.561	.786	.814

There was more consistent significant correlation for the 7 and the 12year groups indicating that the items for the scale of explaining inferences, are particularly accurate for these two age groups. In most items the 7year children scored 2 and 3, whereas the 12year olds consistently scored higher, with 8-10 items showing a mean score of 3.5 - 4. This may be due to the fact that explaining inferences, in which the information is directly requested (How do we know that....?) is a thinking skill that emerges early and is the easiest. When children are asked to give an explanation in response to a linguistically more complex question (e.g. negative why), level of performance is reduced.

A low correlation of .11 was noted for Question 1(How do we know these people are at a wedding?), for the 10year group. Review indicated that this item was not discriminating for this age group, and in fact 69% of the children scored 3 points, and no children scoring 0. The 10year old children were able to pick up the concrete items in the picture and thus, by presenting 2 of them using accurate vocabulary, they scored 3 points.

For the 7 year group, Question 2 (How do we know this man has a problem talking on the phone?) showed a low correlation score of .19 when compared with the scale total. With 94% of children scoring 0 on this item, it was evident that the item was too difficult for children of this age. The correlation improved over the ages for this question, indicating improved ability to answer this question with maturity. It therefore has merit as a developmental item. This question demonstrated a recurrent problem the children revealed with answering questions.

There was a tendency to focus in on the first part of the question, the main clause, which could stand alone. As soon as the child deciphered something meaningful, (How do we know the man has a problem?), there was a tendency to stop processing the rest of the sentence or the complementary clause (..talking on the telephone). Thus the problem of an incorrect answer arose from poor processing of the question.

The Cronbach Alpha reliability coefficient was high across all ages.

Scale 2: DETERMINING CAUSE

In this scale, the correlation for the total group were again higher than the correlation within the different ages, with the 11 year group showing the best distribution for the scale Determining Cause.

In the 9 year group, Question 5 (Why is the painter painting the school?) had a particularly low correlation, which showed little improvement for the 10year group. For the 9year group, 86% of children scored 2, with a small distribution of scores around it. In the 10 year group 75% scored 2 with a slightly better distribution of scores around it. Thus there was not much variability in the quality of the answer given showing the limitations of this question for these age groups. The 7, 8 and 12year groups showed greater variation in the answer presented, with the 11 year group showing greatest variability with an even better distribution than for the group as a whole.

Question 7 (What made the ice-cream drip?), showed a low correlation of .23 for the 7year group. Lack of ability to rely on scientific schemata may have contributed to this. It has been noted that ice-cream is often sold outside the schools by the local women, so most children have some exposure to it.

Table 5.3
Item-scale correlations per age group and for the total group for the scale determining cause

	Age 7 N=51	Age8 N=51	Age9 N=50	Age10 N=48	Age11 N=45	Age12 N=47	Total Group N=292
Q1	.34	.36	.59	.53	.51	.43	.54
Q2	.56	.44	.52	.47	.59	.45	.53
Q3	.51	.42	.42	.37	.46	.38	.49
Q4	.36	.60	.37	.69	.42	.61	.59
Q5	.53	.57	.15	.20	.66	.50	.53
Q6	.39	.40	.30	.45	.46	.47	.46
Q7	.23	.57	.40	.41	.47	.57	.56
Q8	.38	.61	.54	.49	.52	.68	.62
Q9	.74	.56	.69	.46	.48	.42	.60
Q10	.56	.40	.44	.39	.49	.37	.48
Alpha	.581	.632	.582	.557	.659	.632	.726

Cronbach Alpha reliability coefficients were adequate, though not as high as for the other scales.

Scale 3: NEGATIVE WHY QUESTIONS

Once again the pattern of higher correlation for the group as a whole emerged.

Correlation did not show a consistent improvement with age, indicating that the younger children demonstrated greater abiiity in dealing with a linguistically more complex question presented in the negative form than the 12 year group. This may be due to younger children being more exposed to less formulated language forms through television, or demonstrating more flexible thinking skills due to improved schooling. However it is interesting to note that all but the 9 year group showed difficulty with one item at least in this section.

For the 7year group, Question 5, (Why should you never ride a bicycle with a puncture?) showed a particularly low correlation, showing some improvement for the 9 and 12 year group, and good correlation for the 8, 10, and 11 year groups. The low correlation for the 7year group may have been related to the content of the question, which had to do with punctures and bicycles. The children may not have been able to draw on world knowledge in the absence of real ability to analyse the situation, hence the lower correlation for that item.

Table 5.4
Item-scale correlation per age group and for the total group for the scale negative why questions

	Age 7 N=51	Age8 N=51	Age9 N=50	Age10 N=48	Age11 N=45	Age12 N=47	Total Group N=292
Q1	.59	.29	.44	.46	.35	.21	.50
Q2	.43	.56	.50	.54	.58	.61	.61
Q3	.63	.64	.51	.50	.48	.53	.62
Q4	.50	.31	.39	.55	.52	.51	.55
Q5	.18	.54	.29	.54	.59	.30	.45
Q6	.36	.06	.38	.24	.58	.53	.48
Q7	.42	.32	.48	.57	.23	.39	.47
Q8	.40	.62	.46	.42	.42	.42	.46
Q9	.46	.57	.57	.60	.57	.27	.60
Q10	.43	.47	.40	.58	.45	.38	.59
Alpha	.518	.582	.545	.669	.603	.471	.717

For the 8 year group, the low correlation of .06 showed a frequency distribution of 88% of children scoring 0 for Question 6 (Why hasn't the ambulance arrived?). This question also posed a problem for the 10 year group, and required an understanding of the role of ambulances at the scene of an accident, and the relationship between severe/fatal injury, and mere damage to cars. There was a tendency, especially for the 10 year group to provide an answer related to the inaccessibility of the road, which was not relevant for the picture stimulus.

The 11 year group showed a lower correlation for Question 7 (Why will the boy not go to hospital?), with 49% of the children scoring 2, and 38% scoring 0. This answer again combined some world knowledge about children's injuries, and a more complex understanding of the level of injury one takes to the hospital. This distribution did not indicate variability in thinking about this question, hence the low correlation.

The 12 year group showed a low correlation for Question 1 (Why do they not pay for the food they are eating?), with 59% of children scoring 2 or 3, and 21% of children scoring 0%. With the other age groups there was a greater tendency for more children to score 2 or less. Hence within this low correlation indicates more of the 12 year group presented better answers.

The 9 year group showed no particularly low correlation on any one item and frequency was fairly well spread over the different scores.

Cronbach Alpha reliability coefficients were adequate, though lowest for the 12 year group.

Scale 4: DETERMINING SOLUTIONS

Although the total group correlation was still adequate, correlation for the scale Determining Solutions was the lowest, with a large number of questions scoring 0 in all age groups except the 12year group, where this trend was weaker. Thus formulating thoughts that require children to offer solutions, was generally more difficult than providing causes to problems. Considering that in everyday social discourse, the frequency of the concrete, where,who,when,what questions is followed by why questions, even rural Zulu-speaking children may have some life skill in answering causal questions. Determining a solution requires active creation of schemata for something that has not yet happened and conceptual theorizing as to what the possible outcome of the solution may be. Hence it appears to provide a greater challenge to these children.

Question 4 (The painter does not like the colour he was told to use. What should he do about it?) was particularly difficult for the 7 year group, with 92% of children scoring 0. This group of children found it difficult to differentiate role and function of a painter and authority, stating that the painter could independently change the paint if he so wished. Surprisingly, a similar pattern emerged for the 10year group with 75% of the children scoring 0%, and 66% of the 9 year group scoring 0%.

Table 5.5
Item-scale correlation per age group and for the total group for the scale determining solutions

	Age 7 N=51	Age8 N=51	Age9 N=50	Age10 N=48	Age11 N=45	Age12 N=47	Total Group N=292
Q1	.50	.60	.38	.45	.58	.25	.56
Q2	.48	.37	.36	.48	.41	.44	.57
Q3	.33	.61	.24	.37	.36	.44	.51
Q4	.16	.30	.27	.18	.34	.50	.32
Q5	.44	.52	.67	.56	.41	.40	.54
Q6	.32	.37	.27	.26	.64	.38	.44
Q7	.42	.30	.16	.47	.39	.37	.40
Q8	.66	.46	.40	.50	.53	.44	.54
Q9	.50	.26	.50	.61	.45	.28	.53
Q10	.55	.36	.58	.50	.44	.53	.56
Alpha	.562	.487	.368	.515	.568	.388	.661

Question 7 (One of the boys brought R5:00 to school, but now he only has R3:00. What should he do about the money he has lost?), was very poorly correlated with other items for this scale for the 9 year group. The low correlation for this item could be due to the fact that a fairly high number of children scored at both ends of the distribution, with 16% scoring 0 and 14 % scoring 4, with 30% and 32% respectively scoring 2 and 3. Although the distribution of scores was not dissimilar in other age groups, a clear peak at which the majority of children scored, was evident in the other age groups. Once again, the children seemed to process half the question (What should he do about the money?).

Cronbach Alpha reliability coefficients were adequate, with the 12 year group once again showing lower levels of reliability.

SCALE 5: AVOIDING THE PROBLEM

Avoiding the problem was another scale in which a more complex form of thinking was required, as it required children to recognize the implications of a problem first, and then to inference as to how the problem could be *avoided*. The correlation, again tended to be consistent between the different age groups and the group as a whole, with the 10year group showing the highest correlation, once again indicating that younger children may present consistently better answers than older children.

Question 9, (What should the children do so that the ball will not get caught in the tree in the future?), showed a particularly poor correlation with the scale score. 67% of 7year olds scored 0 for this item, which was linguistically challenging, and may have accounted for such a large no of children scoring 0. By 8 years only 49% scored 0, by 9 years this was further reduced to 34%, showing that the older children were coping better at this level of linguistic complexity.

Table .5.6
Item-scale correlation per age group and for the total group for the scale avoiding the problem

	Age 7 N=51	Age8 N=51	Age9 N=50	Age10 N=48	Age11 N=45	Age12 N=47	Total Group N=292
Q1	.36	.32	.56	.26	.35	.37	.35
Q2	.42	.27	.31	.43	.52	.37	.51
Q3	.40	.48	.50	.32	.49	53	.52
Q4	.51	.32	.30	.47	.40	.30	.42
Q5	.63	.40	.41	.45	.27	.42	.51
Q6	.49	.35	.54	.52	.30	.52	.54
Q7	.61	.65	.30	.55	.65	.37	.54
Q8	.51	.24	.28	.64	.48	.47	.50
Q9	.07	.48	.61	.61	.49	.55	.54
Q10	.49	.38	.42	.53	.46	.58	.63
Alpha	.555	.399	.505	.624	.530	.519	.678

Question 1, (After everyone had gone, the visiting team found that they had forgotten the team's ball behind. What should they have done to prevent this?), shows relatively low correlation within the scale across the age groups. The critical issue that action should have taken place before the departure, requiring temporal organization of events, appeared difficult for the children to conceptualize.

Overall, this item analysis confirms the validity of the items in the TATE-ZC for evaluating thinking skills, in these five scales. We are also alerted to the fact, that, firstly, contrary to expectation, there is no clear developmental progression, with younger children performing relatively better than older ones in some instances. Secondly, larger socio-economic and socio-political events may be impacting on measures we make of the thinking skills of rural Zulu-speaking children, and hence on cognitive measures in general.

Cronbach Alpha reliability coefficients were adequate.

5.2.1.3 . Inter- scale correlation

Inter-scale correlation using Pearson's Correlation Co-efficient was calculated for each age group for each scale and for the test as a whole (see Appendix K), and the sample as a whole (see Table 5.7).

Table 5.7
Pearson inter-scale correlation for the group as a whole (N=292)

Thinking Skill	Explaining Inference	Determining cause	Negative why question	Determining solution	Avoiding the 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 the problem	0.64630*	0.70248*	0.66447*	0.65873*	1.000

p<.0001

The Pearson correlation coefficient indicated that when the group as a whole, i.e. the primary school population as a whole is considered, the significant correlation between each pair of thinking skills confirms, that all these thinking skills are contributing to the overall construct of ability to explain. There was also a consistently significant correlation between means for each thinking skill and means for the test as a whole across all age groups, indicating that all scales do measure the construct of thinking skills and ability to explain.

The review of the correlation per age group has implications for a sequence in the development of thinking skills and will be presented in section 5.3.4.2.

5.3. RELIABILITY OF TEST PROCEDURE

5.3.1 Reliabily of Translators

A Friedman non–parametric analysis of variance was performed. The following results were computed on scores, which reflected whether there was concordance or difference in the language of scripts translated and transcribed from Zulu audio tapes into English. No significant difference between the translations of the 3 RA's was found, confirming reliability of the translations used. Table 5.8 reflects, the means, Standard deviations and the Friedman test statistic.

Table 5.8
Results from the Friedman Procedure for inter-translator reliability

Translators (T)	Mean scores	Std Deviation	Friedman Test Statistic	p value
T1	47.7272	1.6181	0.59	0.7442
T2	48.2727	1.3484		
T3	47.7272	1.6181		

$p < 0.05$

The critical issue here, is whether the translation of the original answer impacts on the score obtained by the child. Although this was not found to be a problem for this research, the subtleties of language difference and translation of material is revealed through a qualitative analysis of the differences that did emerge. Table 5. 9 describes some of these differences.

The use of translators for research in language has been shown to be reliable in this research. It is however, a time consuming and costly procedure, and the use of Zulu mother-tongue speech pathologists is undoubtedly the preferred option. In the absence of such researchers in the short term, accurate translation will be an integral part of research involving African languages.

Table 5.9
Qualitative analysis of translator differences

Question	Translator 1	Translator 2	Translator 3	Difference
After everyone was gone, the visiting team found that it had forgotten its ball behind. What should they have done to prevent this?	They should have came back to fetch their ball and go with it	They should have come back to fetch their ball and then leave with it	They should have gone back to fetch their ball and take it away	Translation cannot always be literal, hence variability in translation
Why do these children not jump down? (They are high in the tree-their ladder has fallen down)	They will get hurt and break themselves	They will get hurt and get fractured	They'll get hurt and get a fracture	Vocabulary, and language rules for intransitive verbs in English
How do we know it is time for the evening meal?	We know because they are cooking	We see because it is cooked	We can see they are cooking	Interchangability of verbs in Zulu,*see/know* Translator 2 has changed the syntax of the sentence though the content is still accurate in terms of scoring criteria

5.3.2 Reliability of scoring procedures

A Pearson's Correlation Coefficient was calculated. Two calculations were computed.

- Scorer 1 scored the same set of 30 scripts with an interval of 6 weeks between scorings- intra-scorer reliability.
- Scorer 1 and Scorer 2 each scored 26 scripts and results were compared- inter-scorer reliability.

Table 5.10 describes these results.

The high levels of significance across all scales and for the test as a whole indicates that the scoring criteria, the 5 point scale and examples for each score that were devised are reliable for evaluating thinking skills using the TATE-ZC. In addition, the subtleties in translation described in Table 5.8 confirm that scoring for linguistic accuracy as had been done in the TOPS (1985) is problematic when the researcher works in a language that is not her mother-tongue.

Table 5.10
Pearson's Correlation Coefficient for reliability of scoring

Scale (Thinking skill)	Scorer 1, on 2 occasions	Between Scorer 1 and 2
Explaining Inferences	.97647	.90999
Determining cause	.96135	.87106
Negative why	.98030	.94183
Determining solutions	.92099	.85075
Avoiding the problem	.82800	.88230
Total Test	.97551	.93922

$p<.0001$

These scoring criteria (See Appendix G), may therefore form the basis of a criterion-based system of evaluation, which has been suggested as being more applicable for non-mainstream population groups. Table 5.11 represents skills in relation to explaining and reasoning, the associated TATE-ZC scores, TATE-ZC scoring criteria and the approximate educational level within a rural context, to which these skills relate.

In terms of these criteria, it is most disconcerting to note that the 12year group achieved a mean score for the test as a whole of 56.1, placing them at a level of thinking skill appropriate for entry into literacy. The 7year group achieved a mean score of 30.1, indicating skills adequate for entry into a readiness programme.

Table 5.11
Criterion-based evaluation indicators from the TATE-ZC

TATE-ZC score	Categories of TATE-ZC scoring criteria	Related skills or criteria	Related educational level	Age
0-25	Answer vague and imprecise	Demonstrates some understanding of the question. Must be able to engage in a question/answer discourse	Entry into preschool	3 years
26-50	1 concrete factor presented. Precise vocabulary used	Concrete reasoning presented. Able to derive an answer from a picture and express it verbally	Entry into formal education- Grade R (reception class equivalent to a pre-school preparatory year)	5-6 years
51-75	2 concrete factors or 1 abstract factor	Able to see multiple reasons. Abstract reasoning presented. Accurate verbal expression	Entry to formal learning and literacy	6-7 years
76-100	Complete cause and effect reasoning. Creative answers	Ability to draw on world knowledge and previous schemata. Language reflecting an identification of the critical issue for that question. Competent verbal skills.	Entry to class 5. Emphasis is now shifted from developing oral language and acquiring literacy, to using language and literacy for learning	10 years

5.4 THE RESULTS OF THE ADMINISTRATION OF THE TATE-ZC

The results represented in Figures 5.2-5.9 form the basis of the analyses of the five thinking skills (Explaining Inferences-EI, Determining Cause-DC, Negative Why- NW, Determining Solutions-DS &Avoiding Problems- AP) to follow.

Figure 5.2 demonstrates the clustering of the 7 & 8, 9 & 10 and 11 & 12 year age groups. It also shows that while Explaining Inferences tends to have the highest scores, Negative why and Determining solutions appeared to be the most difficult.

Figure 5.3 indicates that while the trend is not absolutely consistent, there is a pattern in scores within each age group, as to the thinking skills that appear to be easier as opposed to those that are more difficult.

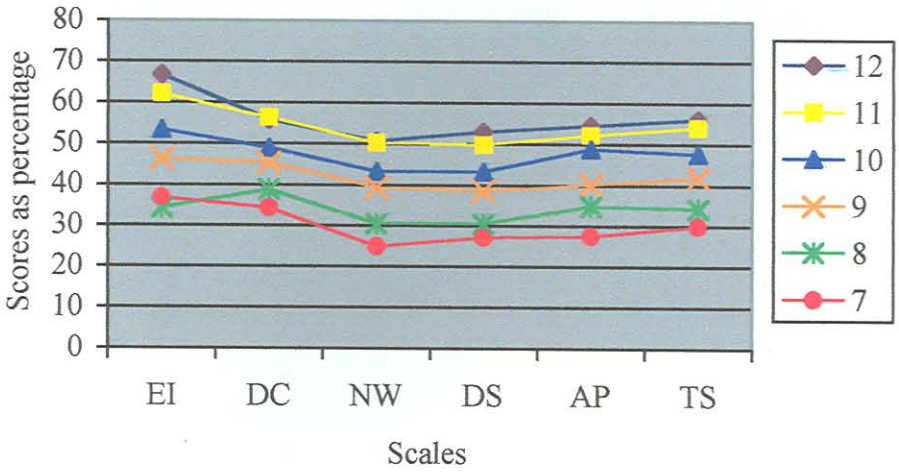

Figure 5.2 Mean scores per age and scale

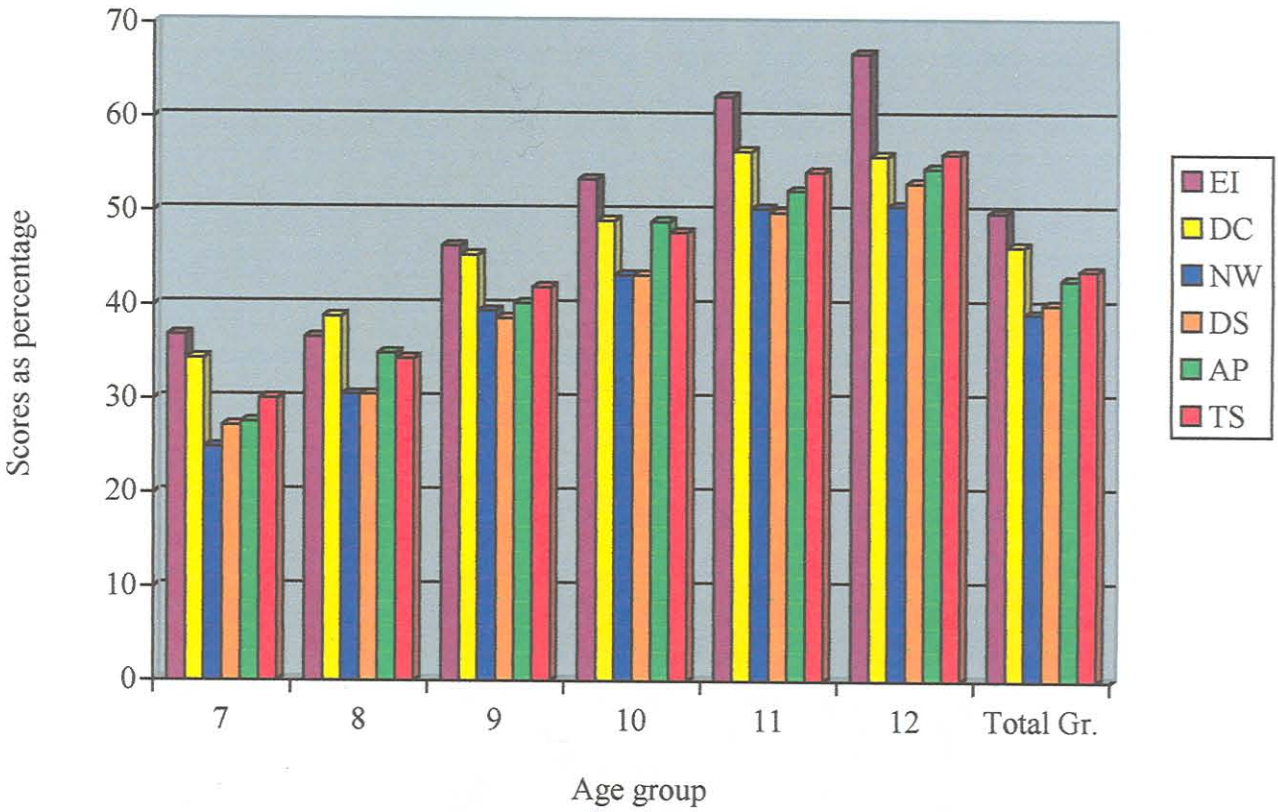

Figure 5.3 Scores per age and for the total group

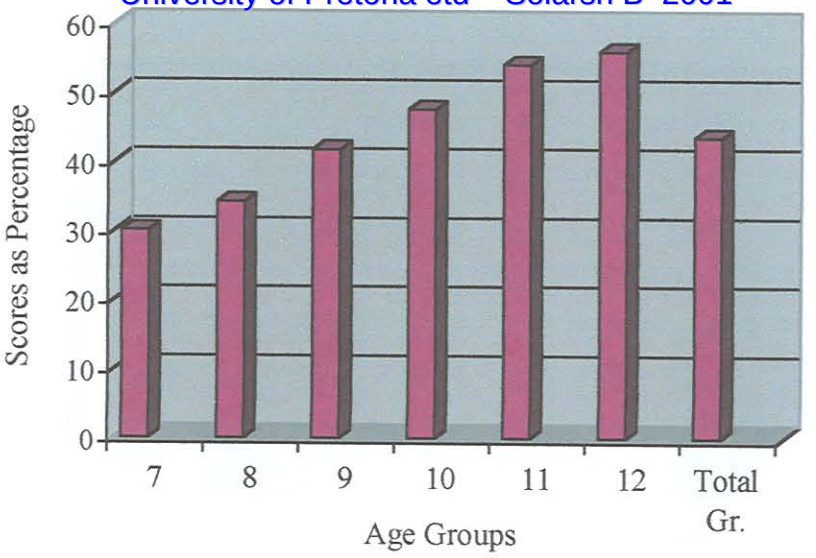

Figure 5.4 Explaining Inferences- age and score

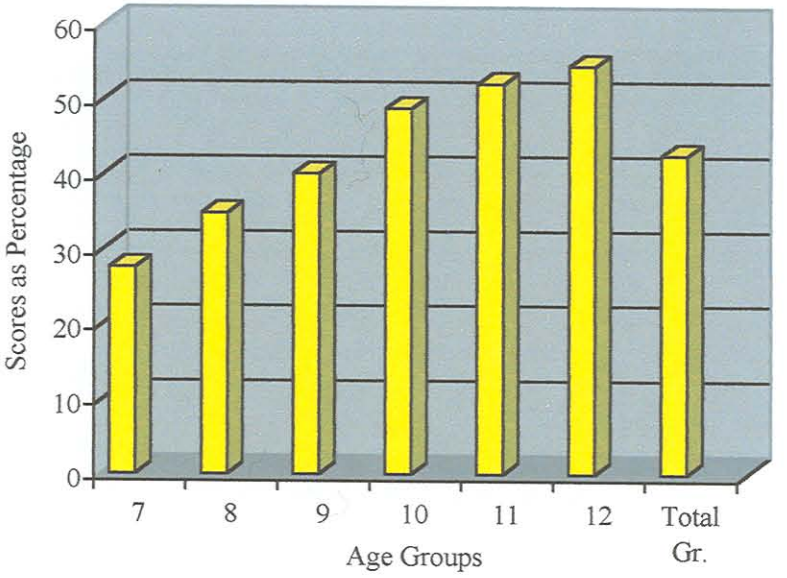

Figure 5.5 Determining Cause – age and score

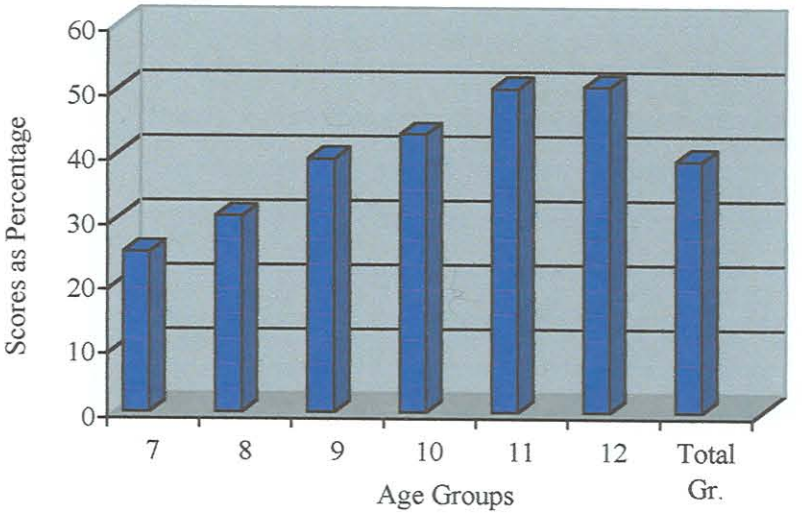

Figure 5.6 Negative Why – age and score

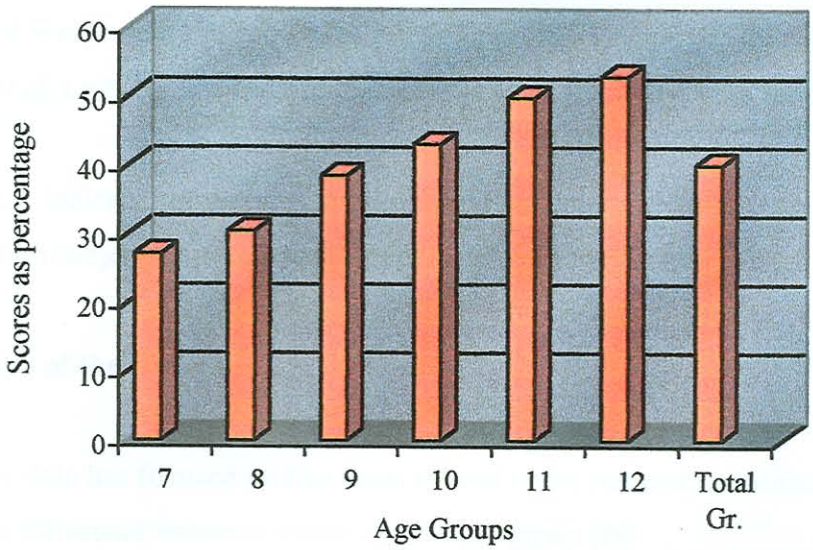

Figure 5.7 Determining Solutions – age and score

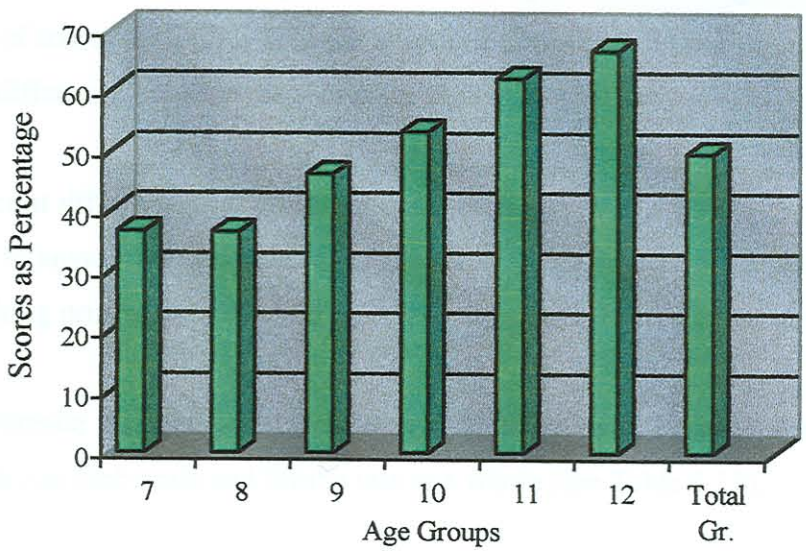

Figure 5.8 Avoiding the Problem – age and score

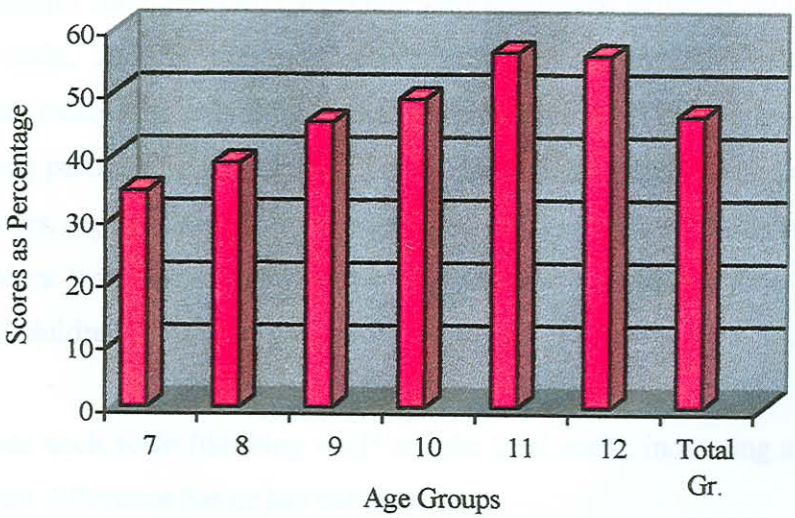

Figure 5.9 Total Score- age and score

Figures 5.4 – 5.9 represent the scores for the different age groups per thinking skill or scale, and for the test as a whole. It is evident that there is an increase in scores for all.

Figures 5.4-5.9 indicate progressive improvement in thinking skills across the ages. The following section subjects this data to statistical analysis for validation.

5.4.1 Analysis of the data

Analysis of the data has focused on five areas related to the research questions in sub-aim 3:

- Significant difference between scores at each age group (iv)
- Correlation between mean scores for each scale and for the total score (v)
- Identification of a developmental process in the emergence of thinking skills (vi)
- Correlation of scores for TATE-ZC with academic performance (vii)
- Significant difference between the scores for the two gender groups (viii)

5.4.1.1 Significant difference between scores at each age group

The motivation to answer the question of how thinking skills develop from year to year in the rural Zulu-speaking primary school child has formed the basis for the whole research project.

Significant difference was computed from mean scores obtained for each age group and the group as a whole, on each scale and for the test as a whole (see Table 5.13).

An ANOVA, using Scheffe's test, a post hoc test for pair-wise comparisons, was applied to these results. Measures for significant difference were calculated between means for each age group for each scale, and for the total score. These results indicated that over all the comparisons made, there were only 3 instances in which significant difference (improvement) in performance in a particular thinking skill was evident on an annual basis, i.e. each year. *In the majority of cases, significant difference was noted every two years and in a few instances, significant difference was noted over 3 years*. This result is a cause for great concern in the education of rural children in South Africa today.

Table 5.12 presents each scale (thinking skill) and the total score, indicating ages and grades at which significant difference has or has not occurred.

Table 5.12
Ages and grades reflecting significant difference for the five scales and total score

Scale	Ages/grades No significant difference		Ages/grades Significant difference		Overview
	Age	Grade	Age	Grade	
Total Test	7 >8 9>10 10>11 11>12	2>3 4>5 5>6 6>7	7>9 **8>9** 9>11 10>12	2>4 **3>4** 4>6 5>7	Significant difference for 8>9year or grade 3>4. No significant difference between other consecutive academic years. 2 year steps noted
EI	7>8 9>10 10>11 11>12	2>3 4>5 5>6 6>7	7>9 **8>9** 9>11 10>12	2>4 **3>4** 4>6 5>7	Significant difference between 8>9 or grade 3>4. No significant difference between any other consecutive years. 2 year steps noted
DC	7>8 8>9 9>10 10>11 10>12 11>12	2>3 3>4 4>5 5>6 5>7 6>7	7>9 8>10 9>11 9>12	2>4 3>5 4>6 4>7	No significant difference between 2 consecutive academic years.7-10years 2 year steps noted No significant difference in last 3 years
NW	7>8 9>10 10>11 10>12 11>12	2>3 4>5 5>6 5>7 6>7	7>9 **8>9** 9>11 9>12	2>4 **3>4** 4>6 4>7	Significant difference between 8>9 or grade 3>4. No significant difference between any other consecutive years No significant difference in last 3 years
DS	7>8 8>9 9>10 10>11 11>12	2>3 3>4 4>5 5>6 6>7	7>9 8>10 9>11 10>12	2>4 3>5 4>6 5>7	No significant difference between 2 consecutive academic years. 2 year steps noted
AP	7>8 8>9 9>10 10>11 10>12 11>12	2>3 4>5 5>6 5>7 6>7	7>9 9>11 9>12	2>4 4>6 4>12	No significant difference between any consecutive years No significant difference in last 3 years

The data is based on a statistical significance at the 5% level of agreement: $p < 0.05$
(EI-Explaining Inferences, DC- Determining Cause, NW-Negative Why, DS- Determining Solutions, AP-Avoiding the Problem)
7>8 = no / significant difference from 7 to 8 years

These results indicate that there were only 3 instances out of all the all the age groups, on all scales and for the total test, in which statistically significant development took place from one class level to the next. On the test as a whole, and in two thinking skills only, Explaining Inferences and Negative Why Questions statistically significant development took place from

grade 3 to grade 4 over 1 year. Although this is not a generalized trend even for the 8 year group, it could be explained in terms of indications that education may be improving.

The finding that for three of the thinking skills, the children of 10 – 12 years, did not show significant improvement over 3 years has critical implications. It implies that there was an even greater limitation in the development of Cognitive and Academic Language Proficiency (CALP) in the second phase of the primary school, during which there is a heavy emphasis on language for learning, than the first. The poor item-scale correlation noted for the 12 year group (5.2.1.2) indicates a very wide range of ability in 12 year children who are in grade 7 and have not failed. This may account for the lack of significant development in some thinking skills over a 3 year period.

This is further endorsed by the finding that for the test as a whole and for the specific thinking skills, children are only showing significant improvement every two years. By the end of the primary school phase, they have therefore only progressed to a class 5 level of thinking skill at the most. This has great implications for their ability to use language for learning and to access information independently, as well as for their ability to access literature and make appropriate inferences from reading materials. It also has implications for high school, with children entering the next academic phase with CALP appropriate for a grade 5 pupil.

Table 5.13 shows mean scores expressed as percentages, for the total group and each age group per thinking skill to give an indication of absolute values required for significant improvement.

Table 5.13
Mean Scores as percentages for the total group and per age group for each thinking skill and the test as a whole.

	Total Group	7years (Gr 2)	8years (Gr 3)	9years (Gr 4)	10years (Gr 5)	11years (Gr 6)	12years (Gr 7)
EI	49.8	36.8	**36.6**	**46.3**	53.4	62.2	66.8
DC	46.3	34.3	38.8	45.3	49.0	56.4	55.9
NW	39.1	24.9	**30.5**	**39.5**	43.2	50.3	50.6
DS	40.1	27.2	30.5	38.6	43.2	49.9	53.0
AP	42.7	27.6	34.9	40.2	48.9	52.2	54.6
Total Score	43.7	30.1	**34.3**	**41.9**	47.7	54.2	56.1

From the above mean percentage scores, it can be seen that for the individual thinking skills an improvement of 9% was required before the difference became significant With the total scores a difference 7.5% showed significant difference. This converts into a 4 point difference on the basic test score out of 40 required for statistically significant improvement per scale, and a 15-16 point difference for the basic test score out of 200. Where there was lack of statistical significance from one age group to the next, there was therefore less than a 4 point difference in scores from one year to the next per thinking skill.

This result has serious implications, which will be further discussed in section 5.3.5. It will also be discussed in the light of findings in section 5.3.4.3.

5.4.1.2 Identification of a particular thinking skill showing high correlation with the test as a whole

This research question aimed to identify whether one particular thinking skill correlated very highly with the total score and could be used as an accurate indicator of thinking skill as a whole. Implications here are that a shorter test or screening test could be devised.

Mean scores for the group as a whole (N=292) for each thinking skill were correlated with the total scores (see Table 5.14). Determining Cause showed a minimally higher correlation than the other scores, but not enough to conclude that in all cases, for all age groups, a score for Determining Cause is sufficient to measure overall thinking skill. Contrary to this, looking at this result as well as the high inter-scale correlation, one could also conclude that performance on any one of the scales could give a fairly accurate estimate of thinking skills. However further research and analysis would be required.

Table 5.14
Correlation between scores per scale and for the total test for the whole group (N=292)

	Total Score
EI	.86394*
DC	.87106*
NW	.86720*
DS	.85203*
AP	.85403*

p<.0001

5.4.1.3 Identification of a developmental process in the development of thinking skills.

A non-inferential analysis of mean scores as well as patterns in the inter-scale correlation per age group, provides the data for this section.

Table 5.13 provides the mean scores for the different ages and for the group as a whole, for each scale. An attempt was made to identify whether a particular thinking skill emerged earlier by ranking the scores for each thinking skill from 1to 5 for each age group and the group as a whole (see Table 5.15).

Table 5.15
Ranked order of mean scores for the different thinking skills

Ranked order	7year group	8year group	9year group	10year group	11year group	12year group	Total Group (n=292)
1 (highest score)	EI	DC	EI	EI	EI	EI	EI
2	DC	EI	DC	DC	DC	DC	DC
3	AP	AP	AP	AP	AP	AP	AP
4	DS	DS	NW	DS	NW	DS	DS
5	NW	NW	DS	NW	DS	NW	NW

Table 5.15 shows that in the majority of instances Explaining Inferences scored the highest, followed by Determining Cause and Avoiding the problem. This was followed by Determining Solutions and the Negative Why. Although this order is not 100% consistent, it is relatively consistent and may therefore provide a basis for further investigation into a developmental order, for the development of thinking skills in rural children (see Figure 5.3).

The review of inter-scale correlation (see Appendix K) produced inconsistent results making conclusive statements difficult. Indications for developmental order are presented in Table 5.16.

Table 5.16 provides some support for the ranked order in the development of thinking skills (Table 5.15). Table 5.17 represents an integration of the two sets of results.

Table 5.16
Inter-scale correlation for developmental order for thinking skills

Thinking Skill	Age	Significant Correlation with:	No Significant Correlation with:	Interpretation
Explaining Inferences (EI)	7	DC, DS, AP	NW	EI correlated significantly with 3 other scales for the 7 & 12 year group. At these two ages, children performed fairly uniformly on all scales. For the 8 &11 year groups, children performed differently on the scale EI, from all other scales, which may be significant developmentally
	8		NW, DC, DS, AP	
	9	NW	DC, DS, AP	
	10	DC, NW	DS, AP	
	11		DC, NW, DS, AP	
	12	NW, DS, AP	DC	
Determining Cause (DC)	7	EI, DS, AP	NW	DC showed significant correlation with other scales over all age groups. This was less applicable to the 11, and especially the 12 year groups. DC did not correlate significantly with EI over a number of ages, which may be significant developmentally.
	8	NW, DS, AP	EI	
	9	NW, DS, AP	EI	
	10	EI, NW, DS, AP		
	11	NW, AP	EI, DS	
	12	DS	EI, NW, AP	
Negative Why Question (NW)	7	DS	EI, DC, AP	NW did not correlate significantly, for most scales for the 7year group, i.e.all scales at a similar level of difficulty for this age group. Significant correlation with DS for 7year groups indicates these two scales are of a similar level of difficulty No significant correlation was particularly evident for the 7, 11 & 12year groups, i.e. it is discriminating for these ages. Significant correlation with most scales from the 8,9 and 10 year groups indicates this scale is not discriminating for the middle of the primary school phase i.e. uniform performance
	8	DC, DS, AP	EI	
	9	EI, DC, DS, AP		
	10	EI, DC, DS, AP		
	11	DC, AP	EI, DS	
	12	EI, AP	DC, DS	
Determining Solutions (DS)	7	EI, AP	NW, AP	DS showed no correlation for most of the scales for the 8 and 11 year groups. DS showed correlation for most scales in the 9 and 10 year group. There was no correlation between DS and EI for 5 age groups.
	8	AP	EI, DC, NW	
	9	DC, NW, AP	EI	
	10	DC, NW, AP	EI	
	11	AP	EI, DC, NW	
	12	DC, AP	EI, NW	
Avoiding the Problem (AP)	7	EI, DC	NW, DS	AP appears to be discriminating for the 8,9 and 12 year group, showing particular difference from EI, NW and DS scales, which may be significant developmentally
	8	DC	EI, NW, DS	
	9	DC	EI, NW, DS	
	10	DC, NW, DS	EI	
	11	DC, NW	EI, DS	
	12	EI	DC, NW, DS	

Table 5.17
Integration of ranked order and inter-scale correlation for developmental order in thinking skills

Thinking Skill	Ranked Order	Inter-scale correlations
EI	1	EI showed non-significant correlation with most other scales. This may indicate that it is different because it is easier and emerges earliest. Item-scale correlations were consistently high for the 7 and 12 yr group (5.2.1.2). This may be due to the fact that this skill emerges earliest, and for the 12 year group which did not perform as well as expected, all scores were depressed including EI.
DC	2	DC showed significant correlation with a number of other scales until 11 years, when it became more discriminating. Thus it seemed to become more challenging
AP	3	AP EI, NW and DS scales showed particular difference from AP. This may indicate it is harder tan EI, but easier than NW and DS
DS	4	DS was not significantly correlated with EI, confirming it is more difficult. The 9 & 10 year groups found DS as difficult as the all other scales. For the 12 year group, DS did not correlate with EI which is easier, and it did not correlate with NW which was harder
NW	5	NW seems to be most discriminating for the younger and older age groups. It is not discriminating for the middle age groups. Mean scores indicated it was the most difficult thinking skill.

Scores and correlation coefficients have been scrutinized in the attempt to identify a developmental process. The above developmental order appears to have merit, but requires further research to propose such a developmental order with confidence.

5.4.1.4. Correlation between TATE-ZC scores and academic performance.

The attempt to collect accurate marks for academic performance at the rural schools proved to be problematic. This was due partly to poor administrative infrastructure in the schools, and partly due to a lack of consistency in allocating grades.

In some instances marks for some classes were not available because the teachers had taken the mark books home for safekeeping and the teachers were not available on the day of the data collection. In other cases, particularly in the junior classes, teachers did not have a defined mark for Zulu or Language. It was therefore decided that a mark for literacy and numeracy would be interchanged with Zulu and maths. Due to the confusing changes with education at present with the Outcomes Based Education (OBE) approach being introduced then withdrawn, and traditional marking schemas being changed to symbols, there is little consistency within the schools. Some grades within one school are on previous traditional

programmes, some are on OBE, some grades define academic performance by marks, others by symbols 1-5 or A-E.

The following information was gleaned from attempts to co-ordinate an objective measure of academic performance.

- At one of the schools:

  Pass conditions up to grade 4: maths and 1 language

  Pass conditions for grades 5 & 6: maths, 1 language and any 2 core subjects
- Foundation phase symbols at one of the schools-

  A and B = Very good

  C =good

  D =weak

  E = very weak
- Key to overall performance at another school

  1=Few skills and very little or no knowledge and values are demonstrated

  2=Some of the knowledge, skills and values are demonstrated

  3=Much of the knowledge, skills and values are demonstrated, but with some minor limitations

  4=Outstanding ability is continuously demonstrated

  This is followed by a Progression key:

  RP- Ready to progress

  NRP- Not ready to progress

  NRC- *Not ready to progress but condoned.*
- A further grading system at another of the schools:

  Symbol code: Grade5-6

  A+ = outstanding (90-100)

  A = excellent (80-89)

  B = very good (70-79)

  C = good (60-69)

  D = very satisfactory (50-59)

  E = satisfactory (40-49)

  F = weak (below 40

Scale: Grade 2 & 3

A =very good

B =good

C =satisfactory

D =weak

E = very weak

Despite the total lack of consistency in the recording of academic performance, the researcher decided to persevere with the attempt to correlate the TATE scores and academic performance. This was done by using actual marks if available, and converting symbols to an equivalent mark allocated by the researcher. This was considered to be the best option as the Spearman Correlation coefficient used for this analysis, computes ranked scores.

Table 5.18
Correlation of TATE-ZC scores with academic performance using the Spearman Correlation Coefficient

Age Group	Correlation between TATE-ZC and numeracy/maths	Correlation between TATE-ZC and Literacy/Zulu	Correlation between Numeracy/maths and literacy/Zulu
7year (N=43)	.22811	.21482	.92056*
8year (N=38)	.46849	.46431	.84189*
9year (N=38)	.19392	.10827	.68526*
10year (N=44)	.25499	.33748	.36211*
11year (N=33)	-.07650	.04826	.69973*
12year (N=39)	.15549	-.05962	.62133*

p<.0001

Results indicated that there was no significant correlation for any of the age groups between academic performance as measured by teachers at the schools, and thinking skills as measured by the TATE-ZC (see Table 5.18). There was significant correlation between the literacy and numeracy skills for all ages.

Children are passing each year, but are not showing statistically significant increases in certain cognitive skills as measured by the TATE-ZC, required for academic progress. Although demographically this sample is small, it does provide some statistical evidence for the fact that rural schools have a low standard of education, and that children are reaching the end of the primary school phase without adequate skills for secondary education.

5.4.1.5 Significant difference between scores for gender

An ANOVA using Sceffe's procedure was performed in which gender was related to means for each scale or thinking skill and the total score, and levels of significant difference calculated. Table 5.19 indicates the significant difference between the genders for most thinking skills.

Table 5.19
ANOVA measures of significant difference for gender

Scales	p value	Mean Values
Total Score	P=.0076*	M=45.3 F=42.1
EI	P=.2942	**M=50.7** **F=49.1**
DC	P=.0003*	M=48.9 F=43.8
NW	P=.1438	**M=40.6** **F=38.3**
DS	P=.0447*	M=41.7 F=38.7
AP	P=0170*	M=44.7 F=40.8

p<0.05 M=Male F=Female

The result of significant gender difference, in which males do better than females in 4 out of the 6 measures, is unexpected. Equal performance appears to be demonstrated for the two thinking skills at either end of the developmental order identified in section 5.3.4.2- Explaining Inference appears to develop earliest, and Negative Why appears to be the most challenging and appears last.

5.5 OVERVIEW OF RESULTS

The following results were obtained:

1. The test instrument was shown to be valid and reliable.
2. The test procedure was shown to be reliable in terms of reliability between the three translators in the translation of scripts.
3. The test procedure was also shown to be reliable in terms of scoring, particularly in terms of scoring criteria identified, the 5 point scale and inter- and intra-scorer reliability.

4. A high level of inter-scale correlation indicated each of the thinking skills in the test represents a valid aspect of thinking skills, and all are testing the same theoretical construct.
5. There was no indication that one of the thinking skills in particular reflected the construct of thinking skills better than any of the others.
6. Statistically significant development in thinking skills in rural African children was shown to occur every 2 years in the majority of cases, and every 3 years in 3 instances.
7. Non-inferential statistics and an inter-scale correlation indicated there was a pattern in the emergence of thinking skills, and that some thinking skills tested were more challenging and developed later that others.
8. No correlation was found between scores on the TATE-ZC and academic performance for any of the age groups.
9. Some statistical differences in scores between the genders were shown to be present.

5.6 SUMMARY

In this chapter the results of eight research questions (sub-aim3 (i)-(viii)) were presented. The results confirmed reliability and validity of the test instrument and reliability of the test procedure. An analysis of the results obtained from administering the Test Of Ability to Explain to 292 rural Zulu-speaking children was presented. Unexpected findings that statistically significant improvement in thinking skills did not occur on an annual basis, and that gender differences in scores are evident, will be further discussed. The difficulties described in accessing an objective measure of academic performances raises issues as to the quality of education in rural schools. An indication that there may be a developmental process in the development of thinking skills tested in this research, may provide constructive indicators for intervention.

The chapter was concluded with an overview of results obtained. An integrated discussion of these points follows in Chapter 6.