3

System-wide Outcomes of NALP in the NT 2004 to 2007

This chapter reports the findings of an analysis of system-wide outcomes of the National Accelerated Literacy Program in the Northern Territory.[12] It commences with an overview of the main findings, followed by an account of the analytical steps taken to derive them. This is necessary because of the constraints of this evaluation. The absence of benchmark or baseline comparison data, and the absence of control groups necessitated a strategy based on the careful exploration of internal variations between subsamples of the target population and the contribution of numerous variables to the outcomes observed. Further, the analysis draws some evaluative conclusions about the appropriateness of assessment measures used for the program to date, with implications for review of their further use.

General Objectives of Quantitative Analysis of Student Outcomes

The two key general functions[13] of the evaluation process have been expressed as:

1. Evaluation of program effectiveness with regard to the specific characteristics of its implementation as teaching practice in schools.

2. Provision of reliable information on school and teaching performance and their outcomes at multiple levels – practitioners, schools and system coordination levels – to support continuous improvement of practice.

For the purposes of the quantitative analysis of system-level outcomes, these objectives are expressed in three operational hypotheses which establish the framework for the data-analytic component of the present evaluative strategy.

[12] Analysis conducted by Dr William Tyler, Adjunct Principal Research Fellow, School for Social and Policy Research, CDU; Dr Tyler's full report is presented in Volume 2.

[13] Lea, T. et al., *NALP Evaluation – Draft Interim Report*, SSPR/CDU, November 2006, p. 75, Section 6.3.

Operational Hypotheses

The objectives of this component of the evaluation procedure may be expressed in terms of three major hypotheses:

1. That the tests (IL and ToRCH) employed in the assessment procedures for the Accelerated Literacy Program are valid and reliable instruments for the measurement of accelerated rates of student reading (i.e., progress >1 yr per four terms of program exposure) for the participating population of students.

2. That measurable rates of accelerated progress may be directly attributable to the levels of a student's exposure to NALP as estimated by:

    a) the proportion of students who demonstrate accelerated rates of progress in either literacy assessment

    b) the effect of the level of a student's exposure to the program (i.e., number of terms per sequence and the number of sequences recorded)

    c) the degree of catch-up indicated by relatively higher rates of progress among lower-scoring students on the initial assessment[14]

3. That variations in rates of accelerated reading progress can be explained in terms of statistically significant differences between students, schools, sectors, regions and other subsample characteristics.

This formulation of the hypotheses (particularly hypothesis 3) leaves open the option of differential treatment of subsamples, particularly if the assessment instrument might be shown to be invalid or inappropriate (hypothesis 1). In fact, the need for a strategic differentiation of the evaluation subsamples was found in the intersection of two major "fault lines" in the evaluative process within and between regions, after comparison of outcomes as measured by IL and ToRCH to two main outcome measures. There fault lines conincide with differences in required entry levels for IL and ToRCH assessments, and with well-documented gaps in average reading levels between remote Aboriginal students and urban and non-indigenous counterparts.

Thus, after analysing the distribution of reading gains based on regional location of schools and students, it was decided to differentiate the process into four separate (though interconnected) evaluation subsamples. Differences between regions were also reflected in access to test type, since larger, urban and high schools were overrepresented in ToRCH sequences, while IL sequences covered a much wider spectrum of regions and grade levels (60% of the IL sample were in the very remote regions compared with only 16% of the ToRCH sample).

[14] This hypothesised effect recognises the need to discount any regression to the mean i.e., the statistical tendency of extreme values on a first measure to converge towards the average on the second.

Research Design

The research design was based on a multilevel framework, whereby the main outcome variable (accelerated annualised rate for measured reading progress for each test) was predicted by selected student, school and contextual variables within and between each level of analysis.

Table 13: Multilevel research design: evaluating student progress in NALP

Clusters	Contexts	Outcomes (Student Progress)
Regions	remoteness	**Test Type** • IL • ToRCH **Main outcome variable:** *Accelerated Reading Progress* (annualised rate of progress > 1 year, aggregated to student level)
⇧ Sectors	primary, secondary, government/non-government, school cluster	
⇧ Schools	school size, teacher replacement rate, student replacement rate – both within and between years	
⇧ Teachers	not available for matching to classes or students	
⇧ Students	age, gender, indigeneity, grade level, attendance record/program participation	
⇧ Assessment Sequences	"pre" and "post" reading levels, year of second or "post" test, number of sequences taken, number of terms per sequence	

⇧ = "nested within"

Table 13 sets out the relationships between levels (clusters) and outcomes (student progress) as mediated or influenced by contextual factors or properties (covariates) specific to each level. Such a framework, if matched to appropriate analytical procedures, can both isolate the clustering effect of success or non-success and make context-specific influences open to more precise statistical estimation. In the past decade or so, research literature in estimating contextual effects on reading progress has been dominated by *multilevel modelling*. A multilevel framework has been useful in evaluating literacy programs in Australia and is now one of the standard approaches used by the Australian Council for Educational Research in the area of teaching and learning methods (Rowe 2003, Rowe & Hill 1998, Rowe & Rowe 1999). Table 13 shows this approach adapted to the NALP evaluation in the NT at system level. Each measure nests within successively higher levels of context, each of which is treated as analytically distinct. There is no attempt to reduce all measures to a common unit of analysis, for example, by making the sequence, student or school the basis of calculating all averages.

Description of Main Variables

Outcome Variables

For the purposes of this evaluation, the annualised rate of progress for each assessment sequence was the same as in the Interim Evaluation Report, that is, by calculating the ratio of the change in reading year levels measured by IL and ToRCH to years of program exposure (allowing four terms per year). Since one reading year level per year is considered normal, any rate of progress greater than one year may be considered accelerated. This was adopted as the principal outcome variable. In summary:

- progress rates, as described above, were calculated on the ratios of the years of reading improvement divided by the annualised period of participation in the program

- "trimmed" progress rates were used throughout; these excluded all those rates that seem to be out of range i.e., those with progress rates of 10 or more years of either reading loss or reading gain

- for purposes of comparison and explanation, there was no initial adjustment to progress rates for any effect of length of exposure on rates of gain; instead, length of exposure (average number of terms per sequence and number of sequences) was added to the list of explanatory variables as sequence effects

Explanatory Variables

Explanatory variables were defined as:

Student Effects

For the purposes of description and evaluation, the individual student was the level at which sequences were aggregated, with each student then treated as the primary unit of analysis. No attempt was made in this stage of evaluation to identify sequences where students may have changed classrooms, teachers or schools between the first and second assessment. Data were not available whereby students could have been matched with individual teachers.

School, School Grouping and Regional Effects

For these purposes:

- school values were based on the aggregated (averaged) values across years 2005–2007

- for descriptive and comparison purposes, school values (averaged) were attributed to each sequence and aggregated to the student level; for multilevel analysis (i.e., including mixed models), schools were treated as a sample in

their own right

- student length of stay is calculated across years only, based on term three data
- average number of teachers per school does not include executive grades
- because of the small numbers in school correlates (only two sectors, three types, three regions), these variables were treated as fixed effects

Analytical Procedure

Testing the hypotheses involved four analytical steps in the development of a quantitative evaluation of accelerated reading progress:

Description **–** univariate descriptions of the distributions of outcome and explanatory variables.

Comparison – based on two- and three-variable associations between accelerated progress and selected variables (Indigenous status and region).

Prediction and statistical inference – based on results of multivariate regression techniques in order to identify:

- the statistically significant predictors of accelerated reading progress both within and between the levels of analysis as set out in the above table (using fixed-effect modelling which treats schools as individual effects)
- the unique combinations of values which are associated with accelerated rates of reading progress. This analysis was intended to reveal the *fault lines* in the student samples based on student and contextual factors
- the effects of the school context in a multilevel model which combined both the row and column dimensions of Table 13, treating schools as a sample in their own right rather than as individual effects

Explanation – based on the results, a further probing was carried out to identify the dynamics of reading progress under NALP, in the light of the three hypotheses.

- Are the testing instruments appropriate for all participating students? If not, what are the consequences for evaluation?
- Does exposure to the program provide hope of catch-up to average reading levels for all participating students?
- What might be the main factors associated with a student's chances of achieving accelerated reading literacy?

These steps form the main divisions of the following presentation.

Description: Data, Outcomes and their Distributions

Sample Profiles

For the IL, there was a total of 6,269 valid sequences, which were aggregated to produce a sample of 3,167 students spread over 52 schools. This yielded an average of 1.98 sequences per student. For the ToRCH, there were 1,573 valid sequences, aggregated to 941 students spread over 32 schools (all of which had participated in the IL). The great majority of the participating schools were in the government sector, distributed over three ARIA regions: *provincial* (Darwin only), *remote* (Alice Springs and Katherine) and *very remote* (all CECs and remote communities, including Tennant Creek).

Schools were located in seven *clusters*: Arnhem, Central Storm, Darwin City, Desert Oaks, Northern Suburbs, Palmerston and Rural (which included some very remote locations such as Pularumpi and Peppimenarti), and Rivers (Gulf basin). The tables following set out the sample characteristics of government schools for each assessment sample. The full list of table values for individual schools is given in the full multilevel analysis in Volume 2 of this report.

Contrasting Samples: IL and ToRCH

The characteristics of the samples of the two assessment regimes present an interesting regional contrast in their distributions in that 60% of the IL sample of students is in the very remote region compared with only 16% of the ToRCH sample. This difference no doubt reflects the difference in entry level to each assessment regime (Transition level for IL, Year 4 level for ToRCH), as well as the heavy concentration of ToRCH assessments in high schools and the more urbanised clusters generally.

This distinction presents both a challenge to and an opportunity for the evaluation strategy. There may indeed be two different criteria for estimating program effectiveness emerging from the two tests, based on modal geographical location of assessed schools and test difficulty. However, since all ToRCH-assessed schools were also participating in IL assessments, there is opportunity for instructive comparison.

Progress Rates Comparing IL and ToRCH

Outcome values were expressed in terms of the annualised rate of progress (trimmed of outliers with values exceeding plus or minus 10 years), and in six bands: those that showed an actual decline in reading scores, those that showed no progress, those that fell respectively into the intervals >0 through 1 year, >1 through 2 years, >2 through 3 years, and >3 years. The figures show large gaps in progress rates between the two samples, with ToRCH samples outperforming IL by .45 yrs for IL and .36 yrs for ToRCH assessments. Comparing the distribution of both assessment outcomes suggests that most of this gap in average progress may be attributable to the differences in the spread of scores – the IL being concentrated at the lower end of the scale

with a modal score of “no progress” and a ToRCH concentration within the >1 through 2 years band.

The contrast between the two distributions is dramatic. While just under half (47.8%) of the student-level ToRCH assessments showed signs of accelerated progress under NALP (an annualised reading progress greater than one year), less that a third (31.2%) of the IL assessments met this standard. Most of this gap would appear to lie in the high proportion of IL students who showed “no progress”, the bulk of whom scored 0 years on both the pre and post assessments.

The question that follows must therefore be, whether these gaps in progress outcome for each test type can be attributed to differences between the characteristics of students and schools for each assessment sample.

Figure 33: Bar chart distribution of IL student-level progress scores in years (2005–2007)

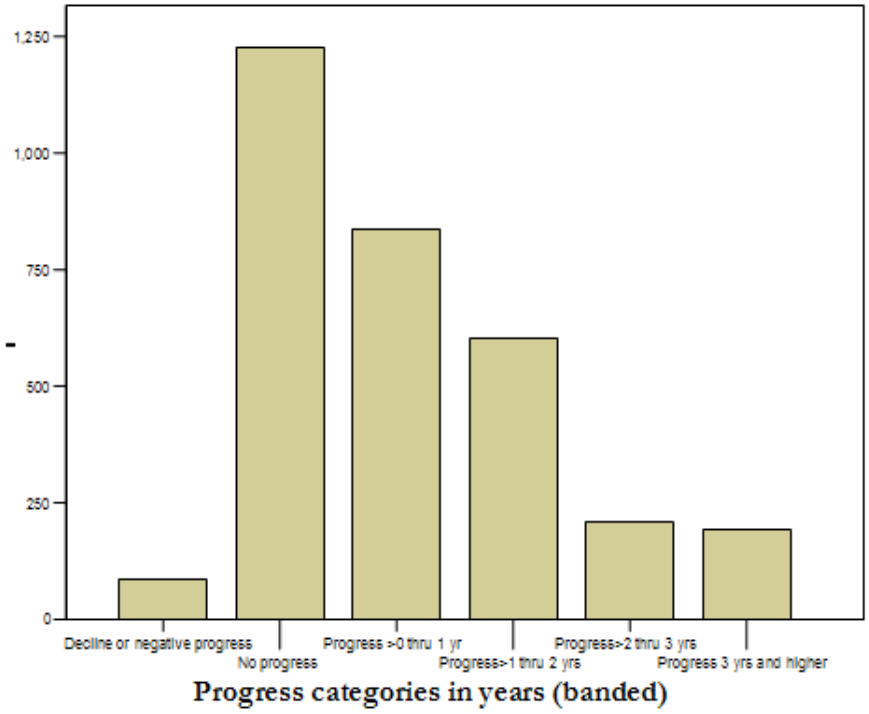


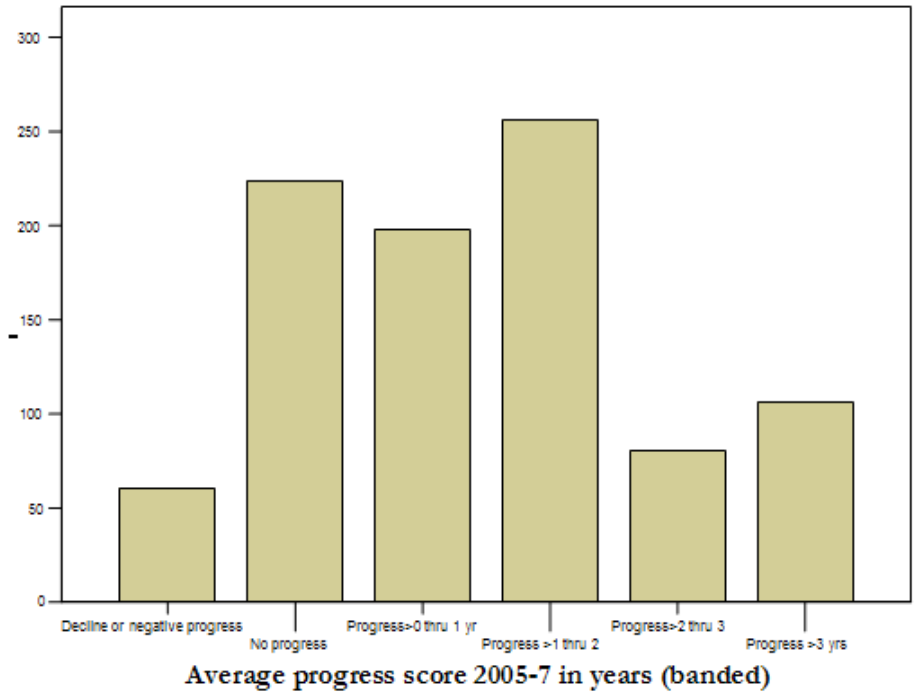
Table 14: IL sample – school cluster by ARIA region (NT DET schools only)

IL Sample: School Cluster by ARIA Region (Govt Schools Only)					
Statistic	**cluster**	**provincial**	**remote**	**very remote**	**total**
average no. of students	Arnhem			59.00	59.00
	Central Storm		66.00	80.88	79.22
	Darwin City	45.67			45.67
	Desert Oaks		73.67	31.50	49.57
	Northern Suburbs	93.50			93.50
	Palmerston and Rural	108.33		49.17	68.89
	Rivers		17.00	48.63	45.11
average no. sequences per student	Arnhem			2.02	2.02
	Central Storm		1.89	1.77	1.78
	Darwin City	1.51			1.51
	Desert Oaks		1.87	1.29	1.54
	Northern Suburbs	1.64			1.64
	Palmerston and Rural	1.90		1.91	1.91
	Rivers		1.00	1.54	1.48
total no. of students		836.00	304.00	1752.00	2892.00

Table 15: ToRCH sample – school cluster by ARIA region of school (NT DET schools only)

ToRCH Sample: School Cluster by ARIA Region of School (Government Only)					
statistic	**cluster**	**provincial**	**remote**	**very remote**	**total**
no. students	Arnhem			4.00	4.00
	Central Storm		58.00	112.00	170.00
	Darwin City	69.00			69.00
	Desert Oaks		63.00	2.00	65.00
	Northern Suburbs	294.00			294.00
	Palmerston and Rural	270.00		6.00	276.00
	Rivers		8.00	21.00	29.00
average no. seqs per student	Arnhem			2.25	2.25
	Central Storm		1.69	1.63	1.64
	Darwin City	1.29			1.29
	Desert Oaks		1.36	1.50	1.43
	Northern Suburbs	1.57			1.57
	Palmerston and Rural	1.91		1.25	1.64
	Rivers		1.00	1.44	1.37
total no. of students		633.00	129.00	145.00	907.00

Figure 34: Bar chart distribution of ToRCH student progress scores in years (2005–2007)



Comparing Mean Progress Rates: Effects of Indigenous Status and Region

Because the samples differ so markedly in terms of both region and Indigenous representation, there is a need to investigate the possibility of major fault lines in AL outcomes, both between and within each assessment regime. For this analysis, error bar charts were used to compare the mean accelerated progress rate, that is, the proportion of students demonstrating annualised progress rates greater than one year. Error bars in the figures show confidence intervals at the 95% level of confidence of where the mean accelerated progress rate could lie, with the width of the interval being largely a function of the size of the base group. Six groups were compared by associating two categories of Indigenous status (Indigenous/non-indigenous) with three levels of ARIA region: *provincial* (Darwin), *remote* (Alice Springs and Katherine) and *very remote* (Tennant Creek and remaining remote communities). These were generated for each assessment type.

The two error bar charts compare the confidence intervals of the mean for each of the six groups for each of the accelerated progress outcomes, IL and ToRCH. They tell a different story of the effect of each of the main explanatory variables. For the IL chart, it is clear that the very remote Indigenous group occupies an entirely different space, well below any chance of overlap with the confidence intervals of the other groups (a conservative test of mean differences).

Figure 35: Error bar comparison of means of accelerated progress for IL (95% CI) by Indigeneity of student and remoteness of school

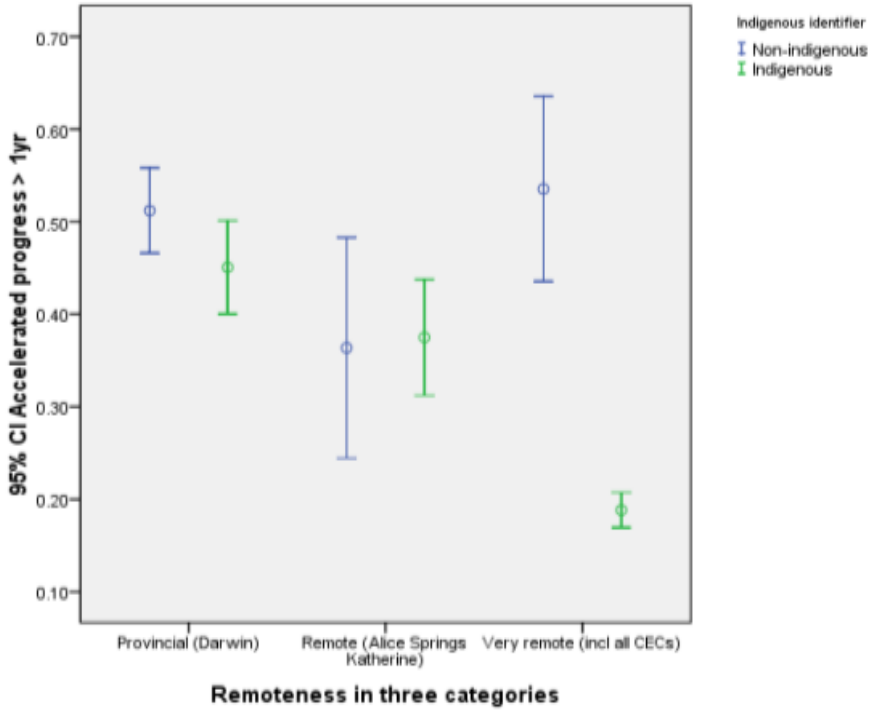


Figure 36: Error bar comparison of means of accelerated progress for ToRCH (95% CI) by Indigeneity of student and remoteness of school

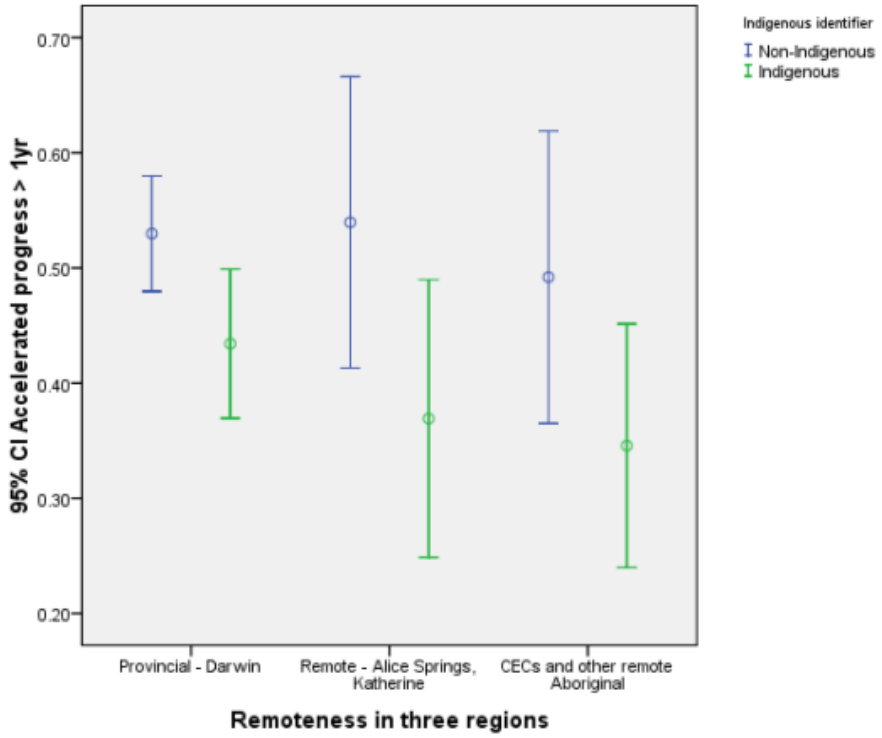
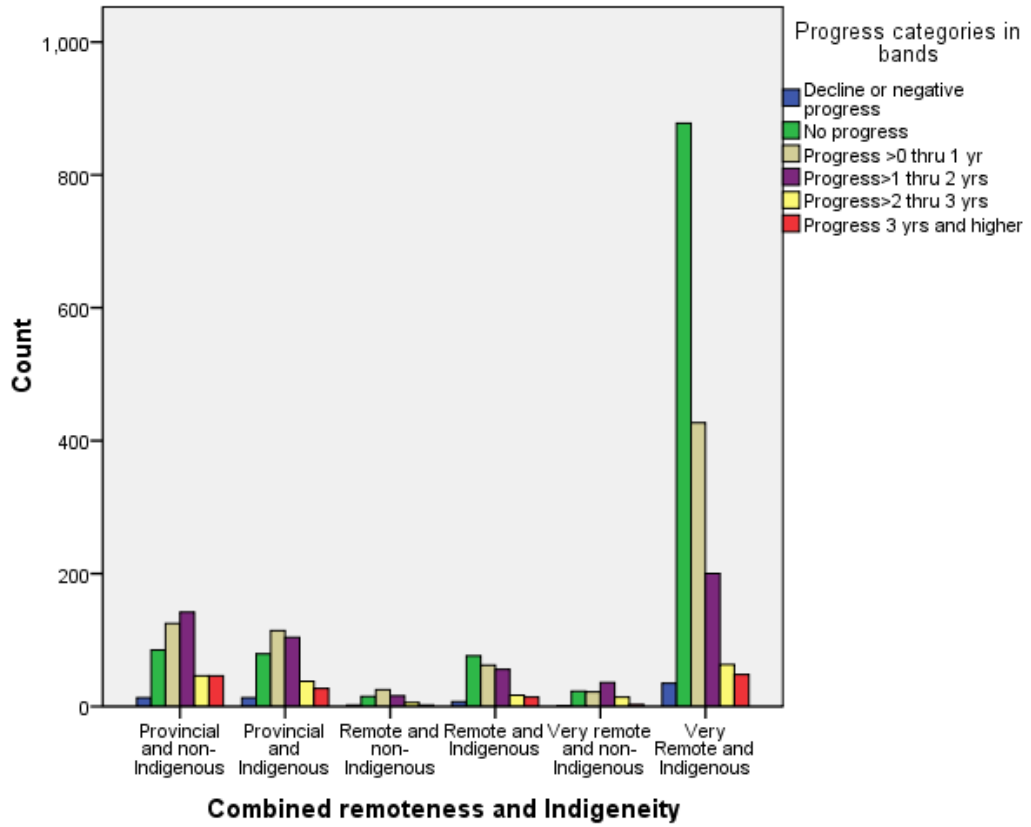


Figure 37: Bar chart of IL progress band by region and Indigenous status groupings (frequencies)



For the ToRCH-assessed students, the Indigenous "gap" appears to be the most important source of variation in mean acceleration scores, with only marginal differences obvious between the regions – in fact the very remote Indigenous group here approximated very closely to that of the remote group (Alice Springs and Katherine). For the IL sample, however, the proportions are dramatically illustrated in Figure 37, which shows numbers of students in each progress band in years. This shows the great overrepresentation of the very remote and Indigenous group in the non-accelerated groups, particularly in the very high proportion that fall into the no progress category.

Teasing out Main Effects

The apparent disadvantages of very remote location for Indigenous students for the IL invites speculation as to what may be the most important determinants of this association, whether in environment or in linguistic background or cultural and pedagogic experiences. This sort of question was explored by multivariate analysis, using regression and analysis of variance (ANOVA) models to identify the most powerful predictors of accelerated progress for each assessment type.

Predicting Accelerated Literacy under NALP: Multivariate Approaches

The predictive part of this evaluation aims to identify the most powerful and significant predictors within the variations in literacy progress graphically displayed in the previous section. While these analyses are important for identifying main predictive effects, they do not fully reveal the unique combinations of these variables that might also contribute to our explanatory models.

The analysis of so-called *interaction effects* allows the researcher to mix and match values within the predictors. For example, male gender may by itself have a negative influence on reading progress, but when combined in a "segment" along with a particular grade level, linguistic background and pattern of program exposure, may exert a positive influence. This data-mining method (sometimes called *segmentation analysis*) therefore allows the researcher to identify *fault lines* in the samples according to the proportion of students showing accelerated progress. Based on the results of this segmentation analysis, as well as on the apparent cleavages based on region and Indigenous status in the samples, a decision was taken to *split* the IL sample in order to better compare and contrast the characteristic predictive profiles of Indigenous and very remote students.

For the final analysis, a multilevel ANOVA model was applied to each of these three assessment samples: two samples for IL and one for ToRCH. In contrast to the fixed-effects model of the first analyses, schools and their characteristics were now treated as second-level effects. This type of modelling is increasingly used in educational evaluation and results in a much more satisfactory and accurate estimation of the effects of contextual factors on reading progress.

Multivariate Regression with Fixed School and Regional Effects

To identify the main predictors of AL when all the other explanatory variables are held constant at their mean values, two applications of regression analysis were conducted, based on the multilevel research design set out above. The first set of analyses explored the predictive effects of each set of variables from region down to sequence characteristics; the second explored the effects of selected variables in a combined model. The analytical technique employed here was binary logistic regression within the SPSS Advanced Statistics module, with "accelerated (annualised progress rate >1 year)/non-accelerated progress" as the dependent variable.

Within Levels: Results of Regression Analysis

The results of the within-rows regression analyses are set out in Table 16, starting with the lower levels of sequences and program participation patterns to identify within-level effects.

Table 16: Summary of within-level effects on accelerated progress IL and ToRCH, 2005–2007

Summary of Within-level Effects of Accelerated Progress IL and ToRCH Assessments 2005–2007 (Student Aggregated)						
Explanatory Variable	IL			ToRCH		
	Significance (p<=.05)	Sign	Effect Size	Significance p<=.05	Sign	Effect Size
Timing of Sequences						
Second assessment year	Significant	Negative	**High**	Significant	Negative	**High**
Second assessment term	Not significant			Significant	Negative	Moderate
First assessment year	Significant	Positive	Moderate	Significant	Positive	Moderate
First assessment term	Significant	Positive	Low	Significant	Positive	Low
First assessment score	Significant	Positive	Low	Significant	Negative	Low
Program Participation						
Number of terms per sequence	Significant	Negative	Low	Significant	Negative	Low
Total no of periods offered	Significant	Positive	Very low	Not significant		
Percent periods attended	Significant	Positive	Very low	Not significant		
Number of sequences per student	Significant	Positive	Low	Borderline	Positive	Low
Student Characteristics						
ESL Indicator	Not significant			Significant	Negative	**Moderate/high**
Male gender	Significant	Negative	Moderate	Not significant		
Indigenous identifier	Significant	Negative	**Very high**	Significant	Negative	**Moderate/high**
Year (grade) level second assessment	Significant	Positive	Low	Not significant		
Language other than English (home)	Significant	Negative	**Moderate/high**	Not significant		
School Characteristics						
Teacher retention rate across years	Significant	Negative	**High**	Not significant		
Number of teaching staff	Significant	Positive	Very low	Borderline significant	Positive	Low
Teacher retention within years	Significant	Positive	**Very high**	Not significant		
Student replacement rate	Not significant			Not significant		
Student length of stay	Significant	Positive	Low/moderate	Borderline significant	Positive	Low/moderate
Type, Cluster and Region				Not applicable		
High school vs other	Not significant			(collinearity)		
Arnhem cluster	Not significant					
Central Storm cluster	Significant	Positive	Moderate			
Darwin City cluster	Not significant					
Desert Oaks cluster	Not significant					
Northern Suburbs cluster	Not significant					
Palmerston and Rural cluster	Borderline significant	Negative	Moderate			

Sequence and Student Characteristics

The summary of within-level differences reveals in detail the convergence and divergence of explanations of accelerated progress between the IL and ToRCH samples. At the level of assessment timing (top of the table), for example, both types have high negative values for the later assessment year, and moderate positive values for a later first assessment year. At first sight, this might suggest that there is a possible advantage for shorter, more compressed sequences of exposure to AL since the stronger effect appears to occur early.

As we go down the table, the divergence between the patterns of explanation for the two assessment regimes increases. The exception here is the common finding of the high negative effect of Indigenous identifier; this remains even when ESL and NESB (language other than English spoken at home) identifiers are held constant.

The lack of effect of language other than English spoken at home for ToRCH sequences is difficult to explain, except in terms of the composition and location of much of the sample for this assessment type in urban schools. Male gender has a moderately negative effect for IL, but is not significant for ToRCH. Findings suggest an interaction between language group, ethnicity and gender that requires further exploration.

School Environment Effects

The divergence between the causal or explanatory background of the two assessment types continues to increase at the levels of school characteristics, school type, cluster and region, with variations between similar variables for the two assessments. For IL, teacher retention rates *across* years have a high negative effect on acceleration, while the teacher retention *within* years is highly positive. Neither effect is significant for ToRCH. The first effect for IL may indicate the need to attract more effective teachers into remote schools offering AL, the second effect indicating the need to retain teachers for a minimum of a school year to reduce discontinuity of teaching delivery. These inferences must remain speculative.

Student retention over a number of years may be a contributing factor to acceleration for IL at least, although the non-significant effect of student replacement rate over a shorter period should be noted. School size as indexed by the number of teaching staff has a positive effect on acceleration for both assessment types. This variable may, however, be masking environmental and regional effects, including covariates such as school resources, literacy teacher specialisation and PD. As the program rolls out to the smaller schools, compensatory adjustment for these factors may be called for.

Across Levels: Defining the Fault Lines

This section looks at defining fault lines between the IL and ToRCH assessment samples. We address the following questions that have emerged from the within-level analyses of the predictors of accelerated progress:

1. What are the main factors determining accelerated progress under each assessment regime *across* (rather than within) the various levels of analysis?

2. Do these factors indicate such deep divisions (fault lines) within the assessment samples that they might justify separate or distinct forms of evaluation for modelling the causal background of reading progress under NALP?

Between-level Regression Analysis

This analysis tests whether the apparent divisions in the two assessment samples, based respectively on remoteness of region and Indigenous status, retain their predictive strength when other important background variables are held constant.

To test their independent effects, accelerated progress was regressed with seven other variables selected from across the student-aggregated levels of the previous section: first assessment score; number of terms per sequence; percent periods attended; number of sequences assessed; year (grade) level of the second assessment; gender; Indigenous status; language other than English spoken at home; and remoteness of school region (by the three ARIA categories).

The results of the cross-level analysis tend to confirm a conclusion that remoteness of school region is a central factor for explaining accelerated progress in the IL sample, while for the ToRCH sample, the main effect appears to be Indigenous status and its correlates. This interpretation can be countered by arguing that remoteness of region is simply a proxy for Indigenous status in the IL sample, and that regional variation in the ToRCH sample was small (particularly among very remote schools). But the pattern of error bar intervals shown above, which accounts for the numbers in each subgroup, and the pattern of covariate effects, suggests a deeper source of division.

Indeed, a close reading of the pattern of the significant covariate effects in the above table suggest that the one column of effects for each test regime is almost a mirror image of the other. The explanatory power of the group of predictors for each sample also varies considerably, with 22% of variance explained for IL assessments (Nagelkerke R-squared estimate) as against only 8% for ToRCH, while the correct classifications for accelerated progress were 71.5% for IL and 59.1% for ToRCH (influenced by the original proportions). Indigenous status, perhaps the most important effect in the error bars, and the major effect for ToRCH, is simply not statistically significant for IL students when the other variables, notably attendance, remoteness, number of sequences and grade level, are held constant. For ToRCH, the only variables that reach significance in common with those of IL are language other than English spoken at home and first assessment score (positive for IL but negative for ToRCH).

Contrasting Rates of Catch-up

The mirror image effect just described is best illustrated in the sign and size of the effect of the first assessment score on the chances of accelerated progress (where grade or year level has been held constant, along with the other predictors). Here it appears from the positive sign for IL that the higher the first score, the higher the chances of accelerated progress, while for ToRCH the opposite effect holds. This has implications for the possibility of a catch-up effect, particularly for the IL sample.

Table 17: Multivariate prediction of accelerated progress – cross-level effects

Multivariate Prediction of Accelerated Progress: Cross-level Effects#						
	IL			ToRCH		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.
First assessment score	0.072	1.075	0.008	-0.157	0.854	0.023
Number of terms per sequence	-0.104	0.901	0.002	-0.069	0.933	0.415
Percent periods attended	0.012	1.012	0	0.01	1.011	0.234
Number of sequences taken (per student)	0.293	1.34	0	0.01	1.01	0.927
Year (grade) level of second assessment	0.167	1.182	0	0.072	1.075	0.278
Male gender	-0.245	0.782	0.023	-0.094	0.91	0.632
Indigenous identifier	-0.119	0.888	0.442	-0.74	0.477	0
Language other than English spoken at home	-0.405	0.667	0.001	-0.519	0.595	0.016
Remoteness of school location (three categs.)	-0.534	0.586	0	-0.169	0.845	0.217
Constant term	-1.378	0.252	0.001	0.196	1.216	0.864

#Odds ratio Exp(B) 1 = no effect <1 negative effect, > 1 positive effect)

Splitting Differences: A Segmentation Analysis

One of the most widely-used methods for segmentation/data mining is the CHAID (Chi-square automatic interaction detection) method which generates "splits" in a sample according to the relative predictive power of each explanatory variable.[15] This data-mining method was applied to both data sets to identify the composition of the segments for both IL and ToRCH samples using these predictors: remoteness (region), gender, Indigenous identifier, age (five groupings), language other than English spoken at home, and ESL indicator. For IL assessments, this procedure generated 12 significant segments at a depth of three levels of splitting with region as the first or root division. However, the same procedure when applied to the ToRCH sample, produced only one split or two segments, Indigenous and non-indigenous.

The comparative complexity of the IL sample displayed in the next table does not lessen the importance of regional divisions in the IL sample between provincial and remote on the one hand and very remote locations on the other. From the segmentations, it is clear that, with the exception of female students aged 9 through 10 years, all segments falling below the IL sample average of 31% for rates of acceleration fall within the very remote regions, complicated by factors such as language other than English (NESB), lower primary year levels and male gender. This finding would support the case that for the IL assessment sample, remoteness, Indigenous status and linguistic difference can combine to produce a kind of "triple whammy" of inhibitions to accelerated progress. Acting jointly, these factors can reinforce one another to produce a definite fault line in rates of accelerated progress within the total population of students participating in AL. This can be further explored in terms of the interaction of reading age at first assessment, remoteness and Indigenous status.

[15] The result of segmentation analysis is a tree structure whose end points identify mutually exclusive and exhaustive segments. The segments generated are then ranked according to the percentage of the dependent variable that they, in combination, predict.

Table 18: IL segmentation analysis predicting accelerated progress

Table 4 IRL Segmentation Analysis predicting Accelerated Progress(Student Averaged 2005-7)					
Segment Combination	*size*	*% of all*	*Number Accelerated*	*% All Accelerated*	*% Segment Accel.*
Age 9 thru 12 yrs, Provincial (Darwin)	339	10.8	188	18.8	55.46
ESL, Age 13 yrs thru highest 4, Provincial	238	7.6	124	12.4	52.1
Non-ESL, Age 13 yrs thru highes, Provincial	272	8.6	115	11.5	42.28
Age lowest thru 8 yrs, Provincial	240	7.6	98	9.8	40.83
Female, Age 11 thru 14 yrs, Very remote	298	9.5	110	11	36.91
Remote (Alice Springs and Katherine)	302	9.6	111	11.1	36.75
Female, Age 9 thru 10 yrs, Very remote	203	6.4	51	5.1	25.12
Male, Age 11 thru 14 , Very remote	296	9.4	73	7.3	24.66
Age 15 yrs thru highest, Very remote	291	9.2	66	6.6	22.68
NESB, Age lowest thru 8 yrs, Very remote	170	5.4	30	3	17.65
Male, Age 9 thru 10 yrs, Very remote	206	6.5	28	2.8	13.59
NESB, Age lowest thru 8 yrs, Very remote	293	9.3	8	0.8	2.73
Total No	3148	100	1002	100	30.9%

Modelling School Effects: A Multilevel Analysis

The consistency of the division in the IL sample between Indigenous and very remote students and other students indicates a strong case for a differentiated analytical strategy in which the reading progress of very remote and Indigenous students is submitted to a more appropriate assessment strategy by way of multilevel modelling. The multilevel approach to social explanation is an efficient and statistically valid method for estimating the effects of school environment on reading progress, whether accelerated or not. By treating school environments as a random variable rather than each school as a unique fixed effect, multilevel models enable the researcher to make inferences relating to school differences outside those observed in the immediate sample. The fixed-effect modelling strategy in previous sections of the report was helpful for exploratory purposes, but could be less than fair if used to rank the performance of higher-level effects, such as teacher, classrooms, schools or clusters.

Estimating School Effects: Three Assessment Samples

This section uses the multilevel strategy to answer the following question:

> Do the differences between schools exert a statistically significant influence on the rate of accelerated progress of their students involved in the AL program, after adjustment has been made for the their participation patterns and individual characteristics?

The results of the three mixed-model analyses are displayed in Table 19 (effects that are statistically significant at the .05 confidence level are highlighted in ***bold italics***; effects just above that level are highlighted in *italics*). The data demonstrates the

variegated pattern of effects of schools, remoteness and student characteristics across the three assessment regimes, and confirms the findings of the cross-level regressions:

- effects for the full IL sample which are, as before, shown to be far more complex than those yielded for the ToRCH assessment, a probable reflection of the greater ethnic and geographical diversity of the former

- importance of language background and remoteness in the prediction of accelerated progress for the full sample

- salience of both Indigenous status and linguistic difference for the ToRCH assessments, rather than remoteness, age and gender

- negative effects of sequence length for full IL, but a definite positive effect for the number of sequences taken

By dropping very remote Indigenous students from the IL sample, most of the major effects that marked the full sample profile disappear or become severely attenuated. Language differences lose their force, as one might expect, but so does Indigenous status, which is, if anything, a positive finding, when compared with its enduring effect on the ToRCH (urbanised) sample's rate of accelerated progress. In fact, for the reduced IL sample, the two most significant effects are both unexceptional: age appears to still be an important influence, but so does the number of sequences taken.

The problematic group across the effects and samples would appear then to be the younger, linguistically different, remote students with poor attendance assessed. To experienced educationalists, this may not be such an astounding finding. However, it has three consequences: it identifies the main barriers to accelerated progress and raises questions about the specific skills required for remote area instructors and about the appropriateness of the IL assessment as a major instrument of program evaluation.

Between-school Comparisons

What then of the school effects, once these student-level and regional factors have been held constant? The only significant effect for between-school differences (and these include all aspects of the school environment, its organisation and the other factors listed in the school level or layer in the previous sections) was found for the full IL sample, and here the effect is very small indeed. Clearly the most important differences that schools exert on rates of acceleration are either embedded in the range of covariates or explanatory factors controlled for in this model, or are to be found at the teacher and classroom levels. The evidence therefore does not support any league table ranking of schools if adjustments are not made for these factors.

In sum, analyses using a multilevel or mixed-effects model indicate that once the factors that make a difference to reading progress can be isolated, the "school" is reduced to nothing more than an organisational site for delivering the program. Many factors within schools are yet to be identified, such as the quality of local leadership, school ethos and professional support, that are either connected with process out-

comes such as the rate of student attendance, or lead to the attraction, retention and support of effective staff. The results point however, to the dangers (a) of ranking schools according to their raw (unadjusted) average rates of accelerated reading progress on the one hand, and (b) of assuming that what may be true of a school as a whole is also true of every element.

The overwhelming evidence of school effects is that there is far more variation within these institutions (probably associated with teacher quality) than between them. This has implications for the way data have been reported to date: the inclination to use raw unadjusted scores as an index of performance of the school should be avoided.

Table 19: Results of mixed-effects model predicting accelerated progress, 2005–2007

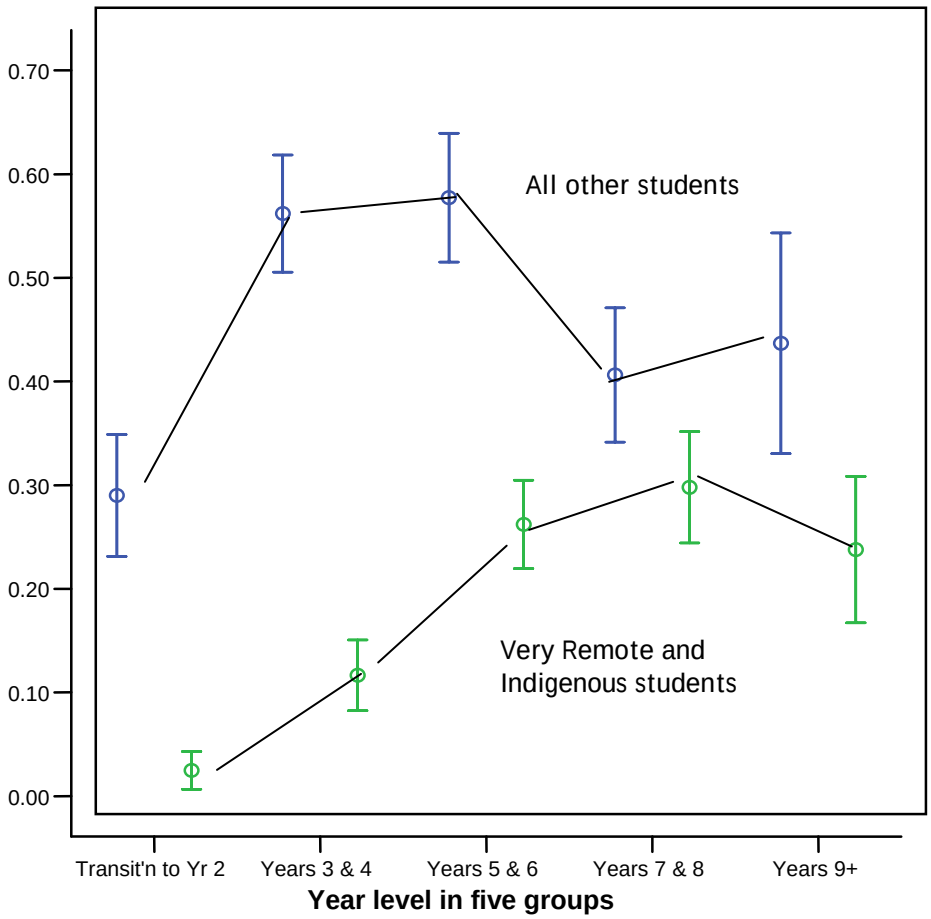
Results of Mixed-Effects Model Predicting Accelerated Progress (y_{ij}) **Student-aggregated data 2005–2007**						
	IL Full		**IL Reduced#**		**ToRCH**	
Explanatory variables (Equation 1 terms in parentheses)	**Estimate**	**Sig.**	**Estimate**	**Sig.**	**Estimate**	**Sig.**
FIXED EFFECTS						
Intercept (β_0)	-0.255	0.001	-0.064	0.740	0.177	0.573
Level 2 (remoteness)						
provincial (Darwin & surrounds) ($\beta_8 x_{8j}$)	***0.199***	***0.000***	0.122	0.319	0.104	0.120
remote (A.S and Katherine) ($\beta_9 x_{9j}$)	*0.108*	*0.060*	0.007	0.960	*0.155*	*0.058*
very remote (comparison category)	0[a]	.	0[a]		0[a]	.
Level 1 (student characteristics)						
Indigenous identifier ($\beta_1 x_{1ij}$)	-0.046	0.120	-0.040	0.311	***-0.156***	***0.002***
length of seqs (no of terms) ($\beta_2 x_{2ij}$)	***-0.017***	***0.001***	-0.006	0.563	-0.016	0.408
age as at 1st July ($\beta_3 x_{3ij}$)	***0.036***	***0.000***	**0.033**	**0.001**	0.014	0.322
male gender ($\beta_4 x_{4ij}$)	***-0.049***	***0.008***	-0.029	0.405	-0.021	0.646
language other than English ($\beta_5 x_{5ij}$)	***-0.058***	***0.023***	-0.054	0.170	*-0.091*	*0.070*
number of sequences taken ($\beta_6 x_{6ij}$)	***0.045***	***0.000***	***0.034***	***0.033***	0.012	0.647
percent periods attended ($\beta_7 x_{7ij}$)	***0.002***	***0.000***	0.001	0.557	0.002	0.322
RANDOM VARIATION						
Level 2 (school effects)						
residual (e_{0ij})	0.172	0.000	0.241	0.000	0.240	0.000
school ID (u_{0j})	***0.005***	***0.035***	0.013	0.249	0.000	Na

a. This parameter is set to zero because it is redundant.
#Excludes very remote Indigenous students because of the high proportion of pre-scale scores

Year Level, Indigenous Status and the Acceleration Gap

Since initial reading scores are correlated with length of schooling, it is necessary to explore the distribution of acceleration rates across year levels. Comparing these rates might also help explain the wide acceleration gap between very remote and Indigenous students and all the other groups shown in Figure 38. Figure 38 compares acceleration rates for these two groupings in the IL sample across year/grade levels.

Figure 38: 95% CI Error bar comparison of proportions showing accelerated progress (0-1.00), very remote and Indigenous students and all others, IL sample



This comparison of different rates of accelerated progress across the year or grade level groups opens another dimension to the deep divisions noted between very remote and Indigenous students and for all the other groups already noted. These divisions here show the relationship between differences in rates of acceleration and student academic progression from Transition through to middle school.

Patterns of Divergence and Convergence

Not only are the rates of progress much lower for the very remote Indigenous group in every case, but the patterns of increase and decline are quite divergent in a number of significant ways. For the very remote and Indigenous students there is a clear positive linear relationship between year level and acceleration rate, but significantly lower than all other groups at every year level. The positive effect of year level for very remote and Indigenous students hits a peak at Years 7 to 8. For the other groupings, there is very rapid early growth in the proportions of students showing accelerated progress. However, beyond Years 5 and 6, exposure to AL is linked to a decline of about 15% in the proportions of students who are accelerated learners, albeit from a relatively high proportion of around 60–65% of accelerated learners between Years 3 to 4 and Years 5 and 6.

For both groups, the first level (Transition to Year 2) shows the lowest proportion of accelerated learners. For Year 9 and higher the effects tend to diverge, with very remote and Indigenous showing a slight decline, while for the other groups the proportion of accelerated learners again increases. It is not possible to say to what extent this divergence reflects program effects, such as regional selection practices into the program – whole school exposure in the remote schools as against a more targeted extraction policy for slower readers in the many urban middle and high schools. The gap in accelerated proportions dramatically emerges at the second level, at Years 3 and 4, so that one might reasonably infer that the origins of the acceleration gap between very remote Indigenous students and others is located mainly, though not exclusively, in the different rates of acceleration in the early primary years.

Closing the Gap Strategy and Year Level

Though there are many discouraging aspects to this analysis for prospects of closing the gap in reading levels between very remote and Indigenous students and other groups, there are also some findings for improving the effectiveness of such a strategy. The dimensions of this challenge must first be addressed. These are illustrated in Figure 39 which compares these two groups by their first IL assessment reading ages. Here we see an initially increasing and thereafter persistent disparity of about three years.

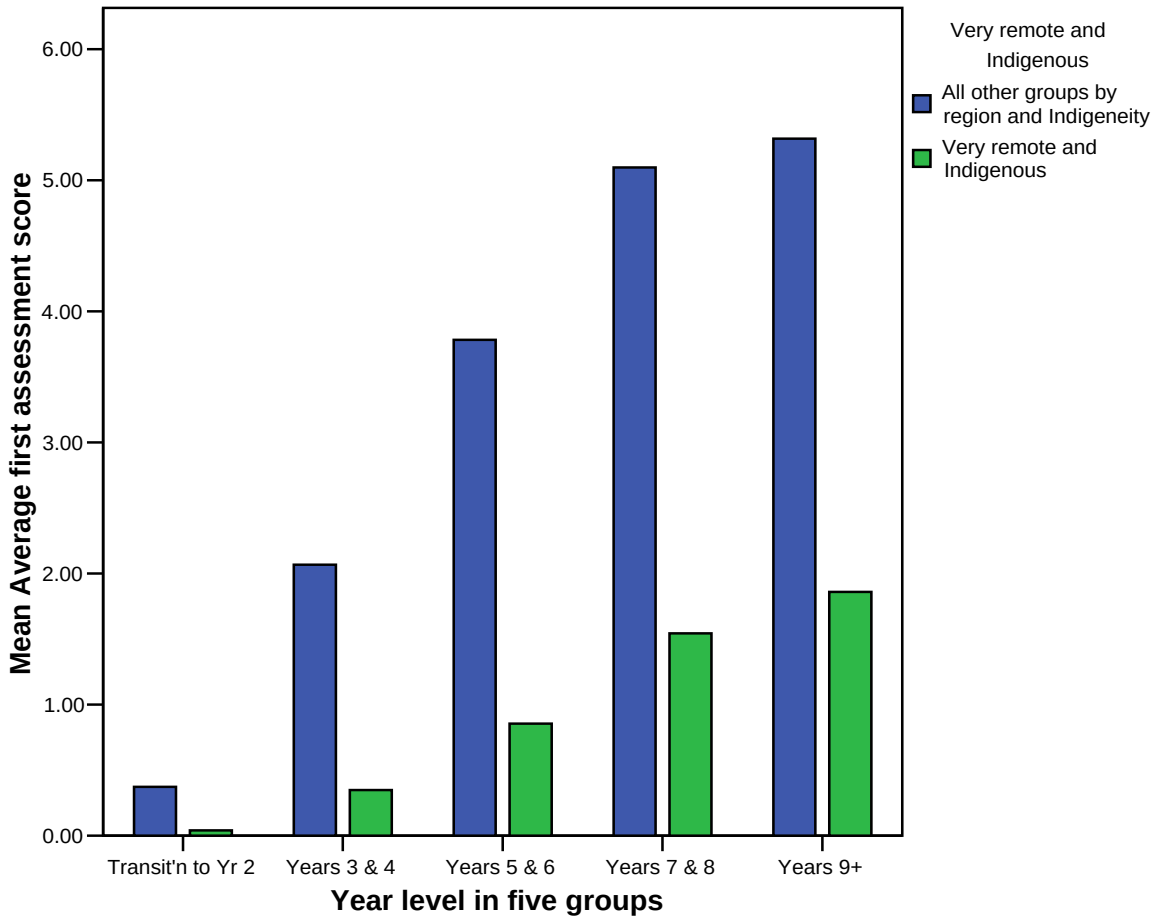
In light of both the analysis of accelerated reading progress and the size of the gap in initial assessed IL reading ages, these observations appear relevant to consideration of the effectiveness of NALP in closing the gap in literacy levels across the NT:

- Since the narrowest gap in the proportions of accelerated learners appears to fall in the Years 7 and 8, these years of schooling may currently yield the best prospects for bridging the gap between the two populations.

- However, the narrowing of the difference in proportions of accelerated learners between the two groups in Years 7 and 8 must be contrasted with the very wide gaps in their respective initial reading ages (of about three years), which persist from about Year 5 through to Year 9 and above.

Unless the persistent disparity in reading ages between the two groups is addressed in the early stages of schooling, it is unlikely that any growth in acceleration rates in the upper primary and middle school years will be sufficient to meet the considerable challenges of closing the gap.

Figure 39: First assessment age. Comparison of group means, very remote Indigenous students and all others, IL sample



Conclusions: Three Hypotheses Reconsidered

This section of the evaluation of NALP was expressed in terms of three major hypotheses. These dealt with:

1. The validity and reliability of the two assessment procedures (IL and ToRCH) as applied to the population of participating NT students.

2. The degree of accelerated reading progress that may be attributable to participation in the AL program.

3. The effect of a range of explanatory variables on observed differences in rates of accelerated reading progress.

These may now be reviewed in turn in light of the evidence presented so far.

Hypothesis 1

> That the tests (IL and ToRCH) employed in the assessment procedures for the Accelerated Literacy Program will be shown to be valid and reliable instruments for the measurement of accelerated rates of student reading (i.e., progress >1 year per four terms of program exposure) for the participating population of students.

Examining the properties of IL and ToRCH was not primarily aimed at assessing their theoretical or practical validity as the most appropriate and useful measures of literacy acquisition and improvement. However, the highly diverse composition and distribution of the sample demands that they be subjected to some critical scrutiny arising from the statistical examination of their properties and performance as measures for the populations under study.

One of the important signs of the appropriateness, if not the construct validity of a test, is the shape of its distribution. In the case of the distribution of the full IL sample, it appears that the reading progress scores for almost half the students (41.6%) were either a negative (2.7%) or zero (38.9%), nearly all attributable to students who scored at a pre-scale or below Transition level on both assessments. Over three-quarters of the latter category (76% of the no progress scores) were by Indigenous students in very remote regions (mostly CECs and other schools in remote Aboriginal communities).

The inability of the IL test to discriminate elements of these students' progress raises important issues, both pedagogic and methodological. The radical difference in region and linguistic background of these students (85.4% of whom come from a home where a language other than English is spoken) led to the decision to recommend a separate evaluation procedure for this group, although, for comparative purposes, it was included in the full sample for the multilevel procedure.

The contribution of linguistically and culturally different very remote and Indigenous

students to the no progress outcome category indicates that the program has failed to bring these students into the lowest levels of the scale. However, the numbers of students making the transition from no progress to any progress score is a blunt measure of performance that almost certainly does not capture all change or learning that may have occurred for non-readers. The findings therefore indicate a failure of the IL instrument (reading aloud a text) to measure any progress that may have been taking place below, or outside of this skill level, that is, progress in literacy acquisition.

The exclusion of the group of non-readers from the evaluation cannot be justified, since it clearly constitutes a large and important focus within the program's objectives. A supplementary analysis of data from five AL schools (two provincial/urban, three very remote) using an alternative instrument, the LLANs assessment instrument (Australian Council of Education Research, Longitudinal Literacy and Numeracy Survey[16]) was conducted (see Appendix 18c in Volume 2 of this report). This study indicated that just under 40% of students who appeared as "non-improvers" on the IL assessment (88% of whom were pre-scale on all assessments), showed at least some improvement when assessed by LLANS. However, the students re-classified as "improvers" under the LLANS were concentrated in the two urban schools, Ludmilla Primary (average improvement rate 70%) and Gillen Primary (58%), compared with much lower rates in the two remote CECs, Ngukurr (24%) and Ramingining (21%).

While indicating the need to identify an appropriate instrument for measuring early literacy gains for beginning readers, the results of the supplementary study strongly suggest that a switch to the use of LLANS as an evaluative measure for NALP-related outcomes would be likely to accentuate, rather than reduce, the literacy gap between urban and remote schools, as well as between Indigenous or non-indigenous students. It would also undoubtedly confirm the significance of reading age at first assessment and the importance of the early years described above. Nevertheless, for both evaluative and formative reasons, an additional or different mode of assessment is recommended for future adoption by NALP.

The ToRCH test, by contrast, appeared to provide a suitable instrument for the assessment of the largely urbanised upper primary and high school samples. While ToRCH also produces a fairly large percentage of no progress or negative progress scores, it shows a much more balanced distribution of the scores when dichotomised at the level of acceleration (47% showed accelerated progress of one year or more). Clearly there is a need for further exploration of the suitability of this test if and when it is distributed more widely to the very remote regions, as well as closer scrutiny of its application to NESB groups in more urban areas.

[16] See M. Meiers and M. Forster, 1999, The ACER Longitudinal Literacy and Numeracy Study (LLANS), paper presented at the ACER Research Conference, "Improving Literacy Learning – What Does Research Tell Us?" Adelaide, 18–19 October.

Hypothesis 2

> That measurable rates of accelerated progress may be directly attributable to the levels of a student's exposure to the Accelerated Literacy Program as estimated by:
>
> a) the proportion of students who demonstrate accelerated rates of progress in either literacy assessment
>
> b) the effect of the level of a student's exposure to the program (the number of terms per sequence and the number of sequences recorded)
>
> c) the degree of catch-up as indicated by relatively higher rates of progress among lower scoring students on the initial or first assessment[17]

Outcomes and Benchmarks

In the absence of benchmark scores for the sample groups studied above, let alone the existence of control groups of comparable students assigned to mainstream or non-AL classes, it is very difficult to assess directly the effectiveness of the program in either (a) lifting the rates of reading progress, or in (b) providing a means of catch-up for students who are reading below normative levels. The cultural and linguistic diversity of the sampled population of students in the AL program presents some methodological difficulties, as does the fact that different methods are used to select students for the program, with extraction groups in high schools, contrasted with a controlled whole-school roll-out as the mainstream curriculum in primary schools.

The best basis for comparison available was found to be in trends in accelerated progress scores from the present two samples, as the program was extended over the three years of observation. Here the evidence is neither encouraging nor clear. From the cross-level regression analysis for both assessment regimes, there is a strong negative effect for the second assessment year, suggesting that the later the sequence, the lower the rate of accelerated progress. This result is complicated, however, by the effects of the expansion of the program in the recent years to the very remote schools, whose frequency of second assessments for IL more than doubled during 2006, while that for other groups increased by only about 50%.

Length of Program Exposure

The effect of increased exposure to the program is ambiguous. Here the negative effect of the length of sequences (the number of terms in a sequence of assessments) is offset to some extent by the small positive effect of the number of sequences taken. These results suggest that there is greater rate of acceleration, on balance, during the first few terms of participation and that the rate of improvement tends to drop off. The positive effect of the number of sequences, on the other hand, picks up the benefits of long-term program exposure, as represented in the gradual improvements in the average gain per sequence. The number of assessments may be a measure of continu-

[17] This hypothesised effect recognises the need to take into account any regression to the mean or tendency of extreme values on a first measure to converge towards the average on the second.

ity in NALP implementation and delivery in schools, while sequences with longer than a year between assessments may point to some level of discontinuity in delivery.

Without further study the policy implications of these findings are unclear. At present they would not support a variation of the program model to emphasise extraction for participation in an intensive program to achieve catch-up by maximising gains per contact hour. This is unlikely to result in sustained improvement in the long term. Frequency of assessment is clearly important to monitoring student progress, and will be important in gaining a clearer understanding of sustainability of effect.

Contribution of NALP to Catch-up at Population Level

This hypothesis has been tested by means of cross-level regression analysis, where it appeared that the first assessment score had contradictory significant effects on rates of acceleration – positive (very slight) for IL but negative (more pronounced but still low) for ToRCH. Since grade level was held constant for this analysis, it is not likely that the lower first score was merely a reflection of its relationship with age or years of schooling. In fact, for the full IL assessment sample, there is a lowered probability of catch-up, as the lower the first scores, the lower the probability of accelerated progress (see *curve fit* analysis, full report, Appendix 18, Volume 2). The finding is counter-intuitive insofar as it apparently overrides the tendency of first scores to regress to the mean. For ToRCH assessments, the opposite effect applies, indicating a lower chance of acceleration as the value of first assessment score increases. This pattern corresponds to the expected effect of regression to the mean, in that average scores remain more or less within the middle range.

These contrasting effects reflect the difference in entry level scores for each assessment regime (1.85 reading years for IL sample, 4.76 reading years for ToRCH sample) and the different composition of the respective samples across regions, school types and age levels. The divergence in patterns between the two assessment samples represents realistic insights into the relative effectiveness of NALP in contributing to catch-up for the students assessed under each regime under the different criteria for entry to assessment. Subject to reservations about the appropriateness of the two measures used, the findings show that the prospects for AL affecting any kind of catch-up for students who score below a Year 4 level on entry to an assessment sequence, are, on the average, not encouraging. On the contrary, the evidence suggests that they will, in relative terms, fall even further behind. For the ToRCH-assessed student samples, including both non-indigenous and Indigenous students, on the other hand, the prospects for catch-up to normative levels are far more promising.

Close analysis of the relationship between school grade, reading level and proportions of students showing accelerated progress suggests that the gap in accelerated progress between very remote Indigenous students and all other subgroups is established early and persists until middle school. It is associated with a persistent difference in reading age levels of about three years. Although the proportions showing acceleration among the very remote Indigenous students appears to rise from Years 8 and 9 onwards, improved rates of progress among this group are unlikely to be sufficient to redress the gap in outcomes.

Hypothesis 3

> That variations in rates of accelerated reading progress can be explained in terms of statistically significant differences between students, schools, sectors, regions and other subsample characteristics.

Although most of the main effects have been estimated in the regression and analysis of variance models, all of the influences in combination explain only a fraction (between 10% and 20%) of the total variation in individual student rates of reading progress. This is a necessary caution, since, tempting as it is at the system-level to think in terms of collectives, institutions and regions, most variation in the distribution of assessed rates of reading progress appears to be found only at the lower levels of analysis, in differences between students, teachers and classrooms. Even with adjustments of scores for the most obvious explanatory factors such as region and Indigenous status, any kind of ranking of schools according to measured outcomes is of doubtful value for evaluation purposes, let alone for policy or management. The prediction and explanation of effectiveness is a difficult and elusive task, with a great deal of what might "produce results" at the level of the individual student remaining unknown by the tools and data currently available for this evaluation project.

Some of the complexity of factors influencing outcomes is well illustrated by the results of the segmentation analysis. Here we found that, within the fault lines defined by regional remoteness, Indigenous status and linguistic background, there were considerable differences in the segments defined by rates of acceleration, with female students aged 11 through 14 years in very remote regions achieving an above average level of accelerated progress (37%) for the full IL sample. Even within groups defined by these three major determinants, the determining factor or sufficient cause is far from clear in every instance. For the full IL sample, for example, the effect of Indigenous status disappears when linguistic and regional differences are held constant. Again, for the full IL sample, Indigenous students in the remote regions (Alice Springs and Katherine) have higher average rates of acceleration than their non-indigenous peers while non-indigenous students in very remote regions have the highest rates of all the six groups grouped by Indigenous status and region, outperforming their non-indigenous peers in Darwin schools. The former finding is favourable for the hypothesis of a catch-up of urban Indigenous readers with non-indigenous readers while the latter points to the need for further analysis of causes of difference between subpopulations, in this case, the distinctive characteristics or experiences of non-indigenous learners in remote communities.

The negative and counter-intuitive findings are equally interesting. When other explanatory variables are controlled for, it appears that a student's attendance rate does not have a strong or significant effect on average levels of accelerated progress. While attendance is important for the full IL sample, in the multilevel model, which controls for between-school differences, it has only a very small influence on acceleration and has no effect at all for either the reduced IL or the ToRCH samples. This is an unexpected finding, given the central importance given to rates of attendance as a necessary, if not sufficient, condition for Indigenous literacy progress. It implies the need for a differentiated approach to students on the basis of their reading age at point of entry to the AL program and one which does not merely rely on

physical attendance as the only precondition, much less as a measure of a program's effectiveness.

Indigenous Status and its Correlates

The importance of Indigenous status as an explanatory factor needs to be put into a wider causal context. For the IL assessments, the effects of remoteness should not be conflated with those of Indigenous status (the wide gap in means for Indigenous and non-indigenous students in the error bar comparisons for very remote students would discount this interpretation). Conversely, the lack of significant effect for remoteness in the ToRCH assessments (where very remote were 16% of the student sample) does not support the case that any negative effect of Indigenous status on reading progress can be explained solely in terms of regional disadvantage.

The effects on accelerated progress rates for both Indigenous status and remoteness predictors are expressed in different causal modalities to produce the range of the subsample gaps displayed in the error bar comparisons. Both of these effects seem to vary either as separated or as joint effects across the two assessment regimes to produce a complex field for interpretation and policy relevance.

The findings of this analysis show that the most powerful effects determining poor rates of progress are a combination of language other than English spoken at home and reading age at first assessment, with both concentrated among the population of Indigenous students in very remote communities. In short, they suggest heterogeneity in Indigenous populations according to linguistic, cultural and other factors that are relevant to student learning. Although this analysis suggests that, for the IL sample, there are some small differences associated with remoteness of school, and the findings of other sections of this evaluation indicate that very remote schools experience lower levels of AL PD support and higher rates of teacher turnover, these factors do not by themselves, if at all, account for the differences in outcomes between very remote Indigenous students and all other groups.

Summary of Conclusions

These findings are a result of a systematic and exhaustive investigation into the levels and predictors of accelerated reading gains of students who participated in NALP in the NT during the years 2005 to 2007. Accelerated reading gain was defined as the student's annualised reading progress scores greater than one year over four school terms on the Individual Level test (IL, n = 3,166 students) and the Test of Reading Comprehension (ToRCH, n = 941), where each assessment sequence applied its own minimum entry levels – Transition level for IL, Year 4 (NT) for ToRCH. The analytical strategy was based on a variety of bivariate, multivariate, data-mining and multilevel procedures for estimating the effects of blocks of explanatory factors grouped by sequences, students, schools and regions.

Wide variation in average rates of accelerated progress reflected the diversity of student characteristics for each assessment type, in that over half (57.3%, n = 1,657) of IL-assessed students were Indigenous attending schools in remote/very remote

communities, compared with only 9% (n = 144) of the same group in the ToRCH-assessed sample. This diversity of intake was expressed in the outcomes as a major fault line, particularly for IL assessments, with only 19% of the remote Indigenous subsample of students recording accelerated gains, as against almost an average of 46% for other subsamples. The major explanatory factors for IL were language other than English spoken at home, remoteness of region and attendance rates, rather than Indigenous status. From ToRCH assessments, Indigenous status and language other than English spoken at home were the only significant effects, with no significant regional variation. School differences, with these other factors held constant, were very small for the IL, and not statistically significant at all for ToRCH assessments.

The low rates of accelerated progress for Indigenous students from remote community schools on the IL test were associated with a much lower initial reading score (a mean of 1.85 reading years for the full IL sample, below 1 year for Indigenous very remote students, and just over 3 reading years for the other subsamples), a very high proportion (53%) of average rates of zero progress across sequences (average number just over four) and a strong relationship between lower initial or pre-program scores and overall rates of progress.

In conclusion, the major findings of this analysis are as follows:

For those students from all backgrounds who have already achieved a basic level of literacy or reading competence at about the Year 4 level, the AL program appears to be delivering measurable and verifiable accelerated improvement in reading skills in a substantial proportion of students. However, for those who, for reasons of cultural, linguistic and situational factors, have not yet got on the escalator of basic reading competence, the evidence suggests that AL, as evaluated by the two current instruments, has yet to demonstrate the achievement of a general or consistent degree of accelerated gain.

For these reasons, a more differentiated implementation and evaluation strategy is strongly indicated, based on improved program delivery, intensified program evaluation and assessment methods adapted to non-normative student populations, the majority of whom are in remote or very remote schools.

NALP in the NT appears to be delivering accelerated rates of reading progress for urban students who have achieved at least NT Year 4 levels of reading performance. There are thus signs that AL may be able to contribute to closing the gap between Indigenous and non-indigenous students in urban and major centres. However, the effectiveness of AL, for all but a small minority of Indigenous students in remote and very remote communities, has yet to be demonstrated.

The analysis of reading age scores assessed and rates of accelerated progress by school year suggests that a substantial difference in reading between these groups of around 3 reading years is established very early in school – by at least Year 3, and persists throughout the primary years, with some tendency to reduce only in later post-primary years. Judging by these results, the program – at the level and intensity of current delivery in schools – as yet shows few, if any prospects of closing the gap in literacy standards, at a system level, *between* very remote and urban school populations.