

ROB091559

Outcomes of the National Accelerated Literacy Program in the Northern Territory, 2004 - 2007.

William Tyler, (willtyler@msn.com.au); Gary Robinson, (gary.robinson@cdu.edu.au);
Claire Bartlett, (Claire.bartlett@cdu.edu.au), School for Social and Policy Research,
Institute of Advanced Studies, Charles Darwin University.

Abstract

The National Accelerated Literacy Program (NALP) aimed to improve educational outcomes for students, particularly Indigenous students in the Northern Territory. The National Accelerated Literacy Program (NALP) commenced in the Northern Territory in late 2004, when the Northern Territory Department of Education (NTDET) committed to implementing the program in 100 schools, training 700 teachers and reaching 10,000 students in predominantly remote locations by the end of 2008.

This paper presents major findings of the evaluation of the NALP in the Northern Territory to 2008. The purpose of this paper is to discuss the challenges faced by the evaluation of an ambitious intervention of the scale and scope of the NALP, to describe the research methods adopted and to present the main learning outcomes of the NALP in the Northern Territory (2006-2008). The findings of multilevel modelling of student outcomes point to significant proportions of Indigenous and non-Indigenous students showing accelerated gains in some, mainly urban contexts, but very low rates of accelerated gain in other contexts, particularly for very remote Indigenous students with low reading age at initial assessment. Reasons for and responses indicated by this “bifurcation in outcomes” are set out in discussion of findings and the conclusions to the paper.

Outcomes of the National Accelerated Literacy Program in the Northern Territory, 2004 - 2007.

Introduction

From late 2004, Accelerated Literacy (AL) was adopted by the NT Government as the mainstay of reforms intended to improve literacy teaching outcomes in the NT. The National Accelerated Literacy Program (NALP) was based on a partnership between the NT Department of Education and Training (NTDET) and Charles Darwin University (CDU). NT DET was responsible for program implementation and CDU for continuing development of the pedagogy, provision of professional development and teacher training, systematization and standardization of resources and materials and for program evaluation¹.

The scale of the NALP in the NT was significant. The program was charged with implementing AL in 100 schools and with training of 700 teachers in order to reach a target of 10,000 students. To meet these targets it was necessary to simultaneously bring a large number of schools into the program and to rapidly train new teachers and teacher trainers (centrally based and school-based coordinators) in the AL approach in order to achieve improved learning outcomes for the population of participating students.

Evaluation of Accelerated Literacy

Before the decision to expand implementation in the NT, Scaffolding Literacy, as AL had until then been known, had been trialled in approximately 30 schools in Western Australia, South Australia and Queensland and later in six schools in the NT. Student outcomes from these pilots, including the six NT schools were said to be promising. Gray & Cowey (2005, p. 22) report that in a period of two years (1999-2001) 'unprecedented literacy gains in some of the most challenging educational contexts in Australia' had been achieved as a result of the implementation of NALP and claimed that, (2005, p. 6) 'students involved in this project are in the process of achieving a major shift in their developmental progression. This kind of developmental shift is unprecedented in the field of Indigenous education'.

Claims of unprecedented accelerated progress were also reported in a largely qualitative evaluation of the Scaffolding Literacy project conducted by the Australian Council for Educational Research (ACER) which concluded that:

'The changes in observable levels are not uniform – the populations of these schools are too diverse socially and geographically for such to be the case, and the circumstances of the students and their schools vary widely as well. But the

¹ NALP also included schools in WA, SA, QLD, and, later, NSW. These schools were able to access training and resources provided by NT DET and CDU but were not included in the evaluation program. The findings of the evaluation are reported in Robinson, G, et al., 2009a & 2009b.

changes are always solid, always upward, and often spectacular'
(Cresswell, 2002, p. 27)

Additionally, Gray & Cowey (2005a, p. 6) contended, 'The indicator chosen for this section of the report (percent of students reading unseen text at or above benchmark grade level) provides strong evidence that a significant and quite revolutionary shift is occurring in the outcomes for the students in the (pilot) study.... It is a perfectly reasonable presumption given the extremely low initial reading levels that virtually none of these students would have progressed without the intervention provided by this program.'

However, the validity of the pilot evaluation findings and subsequent student progress reports (cf. CDU, 2006) were questioned both within the program and by some literacy educators who criticized the use of non-standard assessment instruments and the "unsystematic in-house process" of reporting progress. Moreover, since IL can only assess the progress of students who can read at least 90% of a transition or higher level text, "non readers" were not included in the analyses (cf. Gray, 2007). Thus progress was systematically overstated because of the exclusion of a very large group of students. The Interim Evaluation Report by CDU (Lea et al, 2006) acknowledged these difficulties and indicated that steps would be taken to overcome them.

Griffin, Woods and Nguyen (2005) describe *evidence* as the evaluation of methods used by practitioners to demonstrate the effectiveness of interventions, which are in turn fed back to policy makers to influence future practice. They point to a tension between the complex social and cultural act of teaching which relies on professional judgement and wisdom, and evidence based research findings based on empirical data collection and analysis. They contend that, 'Evidence is more than assessment data...the evidence underpinning policy decisions should be based on professional judgment tested against empirical data' (p. 44). The approach to evidence taken by Gray & Cowey (2005) fits more or less with this description, in that student assessment data, practitioner wisdom and qualitative studies are deployed intuitively to show promising gains. There is no rigorous evaluation of program *effectiveness*, much less *efficacy* (Flay, 1986) no analysis using any advanced statistical techniques, much less randomized or matched control groups to investigate outcomes of delivery under controlled or at least known conditions of implementation.

Rowe, Devine, Knight, Loudon, Lovett, Myer, Ramsay, Rice, Scalfino & Smith, K. (2005a, p. 85) define evidence-based research as:

...the application of rigorous, objective methods to obtain valid answers to clearly specified research questions. It includes research that: (1) employs systematic, empirical methods that draw on observation and/or experiment designed to minimise threats to validity; (2) relies on sound measurement; (3) involves rigorous data analyses and statistical modelling of data that are commensurate

with the stated research questions; and (4) is subject to expert scientific review.

No previous evaluation of Accelerated Literacy or the antecedent programs has met these criteria. The evaluation reported here and in other papers takes some significant steps in this direction, but, as will be outlined, encountered some important limits in research design.

From both policy and evaluative standpoints, the evaluation of the NALP needed to answer two seemingly straightforward questions. Was, or, rather, to what extent was AL actually implemented in NT schools? What were the learning outcomes that could be attributed to it? However, these were not simple tasks in practice.

The evaluation design had to flexibly and pragmatically respond to the realities of uneven program implementation *post hoc*. AL was not fully implemented from a single point in time, but rather cumulatively, as different elements of capacity could be brought together: teaching resources written and disseminated, coordinators trained and schools inducted into the program over time. There was differential production of assessment data by which to measure outcomes, with teachers and coordinators slowly increasing their output of assessments over time, but very unevenly across participating schools, so that some schools were reporting “progress” on the basis of very small numbers of students. There was virtually no capacity to align comprehensive assessment information with any measurable process of implementation of AL in classrooms. Both any kind of pre-post comparison of outcomes using baseline data, and any attempt to use comparison groups for a significant sample of schools were ruled out.

These important constraints on the evaluation had the effect of forcing a dual strategy. Firstly, it could not be assumed that the implementation effort would lead to the required changes in teaching practice and the actual teaching of AL in classrooms to the standards required. Therefore, an empirical investigation of AL teaching in a significant sample of participating schools by direct observation was required. Observational data were then interpreted using a mix of quantitative and qualitative methods drawing on a survey of all practitioners and focus group interviews with teachers and coordinators. The findings of these investigations are reported elsewhere. Secondly, a quantitative analysis of system-wide outcomes for *all* participating schools was required. This would use all available assessment information and apply advanced statistical techniques to estimate the contribution of AL and of key contextual determinants to measured progress in literacy learning by participating students. For the time being, it is not possible to measure learning outcomes of students taught by teachers observed.

This paper presents the findings of our investigation of the second of the two main evaluation questions. In effect it tests a simple hypothesis: that the implementation of AL teaching practice in schools led to accelerated student learning. However, the NALP implementation program was complex and multi-layered, and not a single intervention or treatment, so that measured outcomes almost certainly reflect

change mechanisms based on complex interactions between many determinants of the effectiveness of teaching practice in a given school and community context (Pawson & Tilley 2000). These complexities are inherent in the characteristics of the key elements: the *teacher*, which includes prior training and experience, including experience in a given context or with a specific population group, AL-specific training and experience, and the teacher's retention in a given school; *systemic* factors, including provision of training and support and managerial requirements; the *school*, in terms of leadership, capacity, location and organizational-contextual influences on its operation; and *teaching practice*, which is multi-dimensional, consisting of various interacting elements, skills and strategies which are no doubt variably deployed by individual teachers and not all of which are known in terms of their relationship to teaching outcomes; finally, the *context*, including community-contextual influences and characteristics of students which shape students' readiness to engage with academic learning or with the teaching as provided.

The analysis of system-wide outcomes needed to be able to investigate the effect of these many contextual factors on the achievement of the desired outcome, accelerated student learning.

Research hypotheses and design

A systematic investigation of outcomes in all participating schools addressed 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 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 the NALP as estimated by:
 - a. the proportion of students who demonstrate accelerated rates of progress in either literacy assessment
 - b. the effect of the 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 assessment²
3. That variation in rates of accelerated reading progress can be explained in terms of statistically significant differences between students, schools, sectors, regions and other sub-sample characteristics.

² This hypothesised effect recognizes 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.

The research design was based on a multilevel framework, whereby the main outcome variable (the accelerated annualized rate of progress for each assessment) was predicted by selected student, school and contextual variables within and between each level of analysis. Recent 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 1 sets out the relationships between levels (clusters) and outcomes (student progress) as they are mediated or influenced by contextual factors or properties (co-variables) specific to each level. This framework, if matched to appropriate analytical procedures, is able both to isolate the clustering effect of success or non-success and to make context-specific influences open to more precise statistical estimation.

Table 1: Multilevel research design: evaluating student progress in the NALP

Clusters	Contexts	Outcomes (Student Progress)
Regions	remoteness	Assessment Type <ul style="list-style-type: none"> • IL • ToRCH Main outcome variable: <i>Accelerated Reading Progress</i> (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, indigenous status, 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”

The analysis enabled identification of levels and predictors of accelerated reading gains of students who participated in the NALP in the Northern Territory during 2005–2007. Accelerated reading gain was defined as a student’s rate of reading progress as an annualized score greater than one reading year over four school terms according to either of two assessment measures: the Individual Level test (IL), n = 3,166 students and the Test of Reading Comprehension (ToRCH, ACER, 2003), n = 941. Each assessment was applied at specific minimum entry levels: Transition level for IL; Year 4 for ToRCH. The analytical strategy was based on bivariate, multivariate, data-mining and multilevel procedures for estimating the effects of blocks of explanatory factors grouped by sequences, students, schools and regions³.

The use of two major instruments with different guidelines for use to measure reading gain by students participating in AL meant that the evaluation in effect investigated two distinct but overlapping samples of students, one for each assessment type. Indigenous students attending schools in very remote communities were 57.3% (n = 1,657) of IL-assessed students, compared with only 9% (n = 144) in the ToRCH-assessed sample. The main sample characteristics are summarized in Table 2.

Table 2: Characteristics of samples for IL and ToRCH assessments

	<i>IL</i>	<i>ToRCH</i>
By Size		
Valid Sequences	6219	1573
No of Students	3167	951
No of Schools	52	32
By Socio Demographics		
Indigenous	80%	44%
Very Remote	58%	15.4%
NESB (home language)	69%	38%
High School	14%	52%
Average School Year Level	4.84	6.73

³ For each assessment type and for items on teacher, school and student characteristics, data were assembled from various sources and provided in Excel format by the NALP data manager, Mr Bruce Dunn.

Overview of Findings

There was wide variation in average rates of accelerated progress for the two evaluation samples. The ToRCH samples outperformed the IL samples by .45 reading years per year with an overall rate of progress of 1.4 reading years per years, with 47.8% of students showing accelerated progress (progress greater than 1 reading year per year) and 24% showing no progress in reading scores. For the whole IL sample, the average rate of improvement was .95 reading years per year. In the IL sample, 31.8% of all students showed accelerated gains. However, only 19% of very remote Indigenous students recorded accelerated gains, as against an average of almost 46% for the other groups in the IL sample. In the IL sample, 38.9 % of students recorded no progress in reading scores and most of these scored zero on both assessments.

The following figures show the distribution of rates of accelerated progress for the two samples.

Figure 1: Bar chart distribution of IL student level progress scores in years (2005–2007)

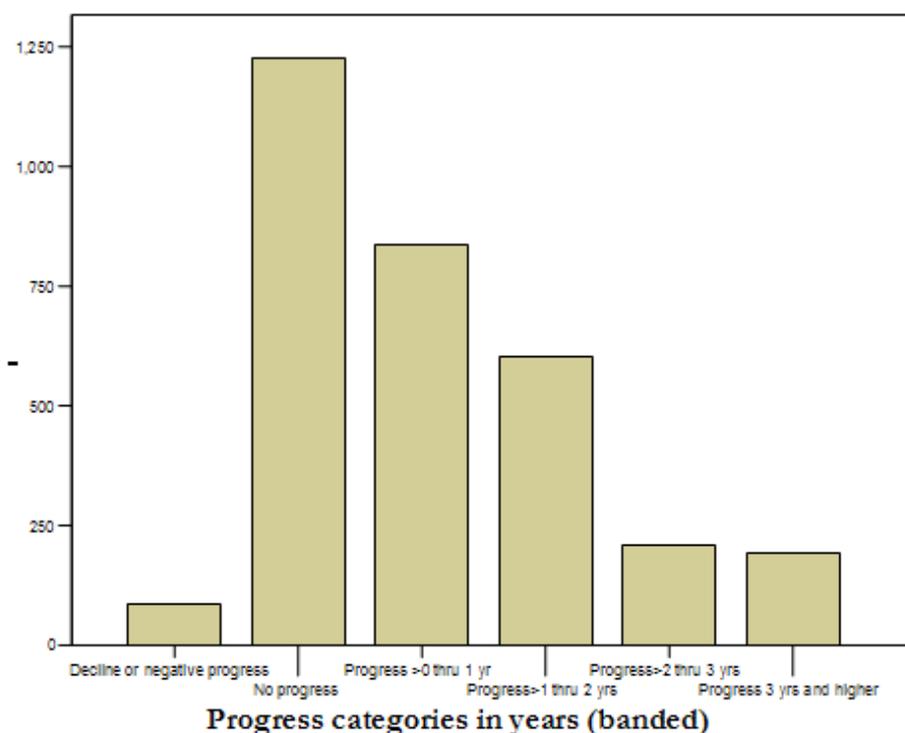
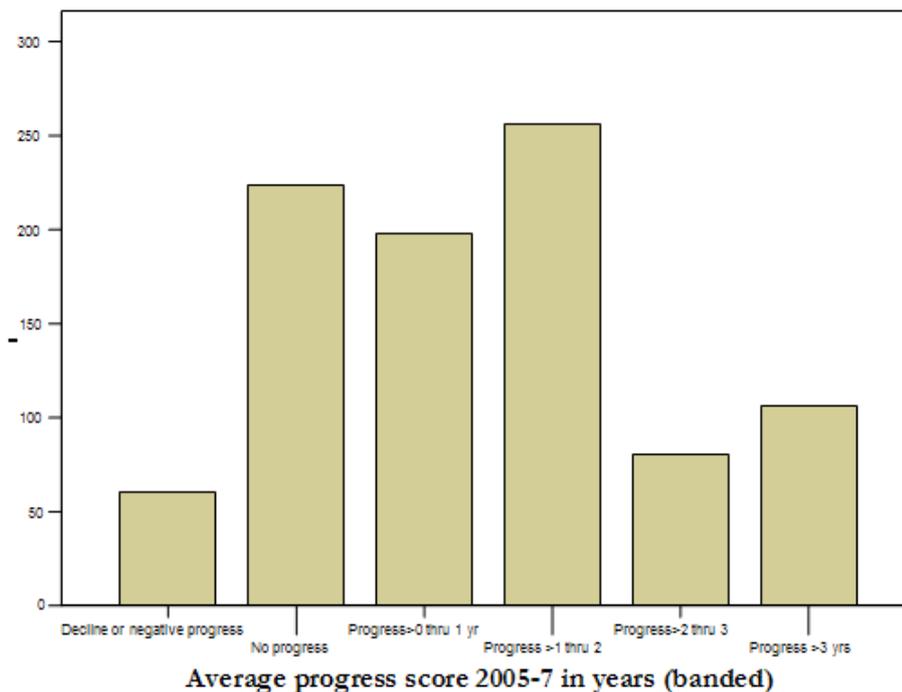


Figure 2: Bar chart distribution of ToRCH student progress scores in years (2005–2007)

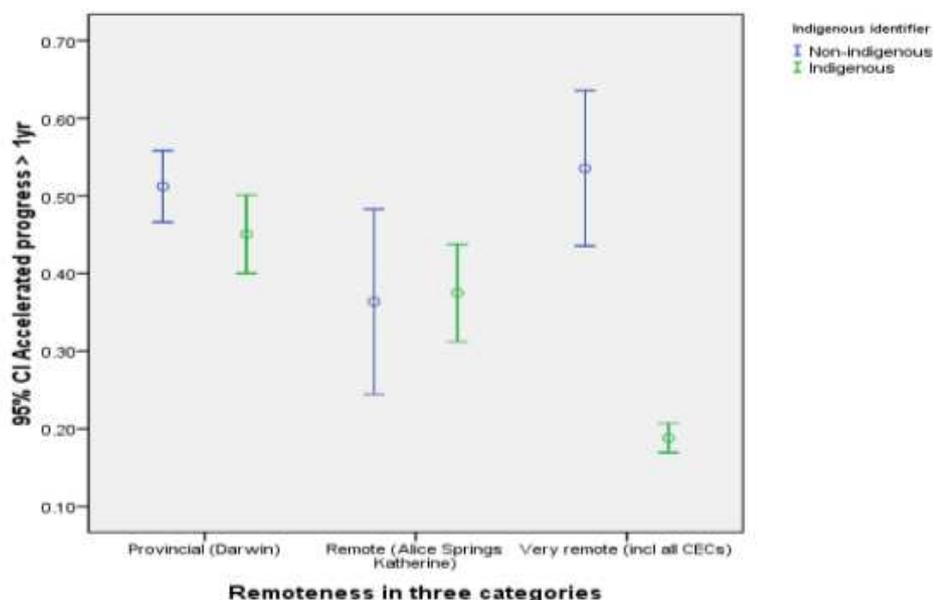


Clearly the rates of progress for IL assessed students is dominated by students showing no progress and progress less than 1 reading year level per year, while these groups are a much smaller proportion for ToRCH assessed students.

The major explanatory factors for IL students were language other than English spoken at home and remoteness of region, with some influence of attendance rates, with Indigenous status less significant; while for ToRCH assessments, Indigenous status and language other than English spoken at home were the only significant effects, with no significant variation across regions. School differences, when these other factors were held constant, were very small for the IL, and not statistically significant at all for ToRCH assessments. The diversity of student characteristics for each assessment type thus coincided with a major bifurcation in outcomes for different sub-populations of participating students. This bifurcation was particularly marked for IL assessments.

These outcomes are represented by the following figures which show contrasting rates of progress for the major subgroups of students by region.

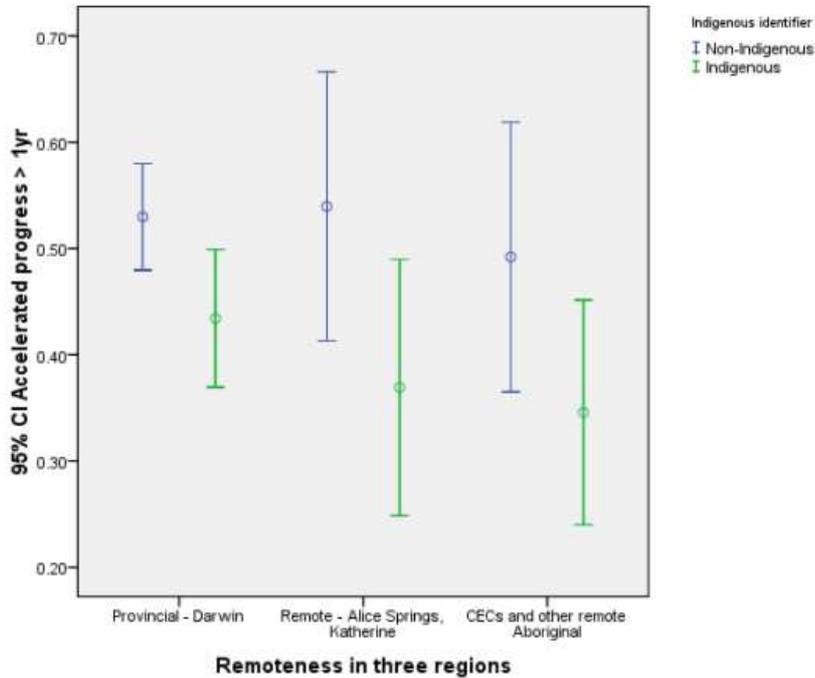
**Figure 3: Error bar comparison of means of accelerated progress for IL (95% CI)
by Indigeneity of student and remoteness of school**



While proportions of Indigenous students showing accelerated progress are substantial in major centres (and for non-Indigenous students in very remote areas), for very remote Indigenous students in the IL sample, less than 20% of students show evidence of acceleration. The outcomes in major centres are generally comparable for ToRCH assessed students – with a reading age over 4 years and more likely to be in a high school – with somewhat higher numbers of Indigenous students showing accelerated progress in very remote schools than in the IL sample.

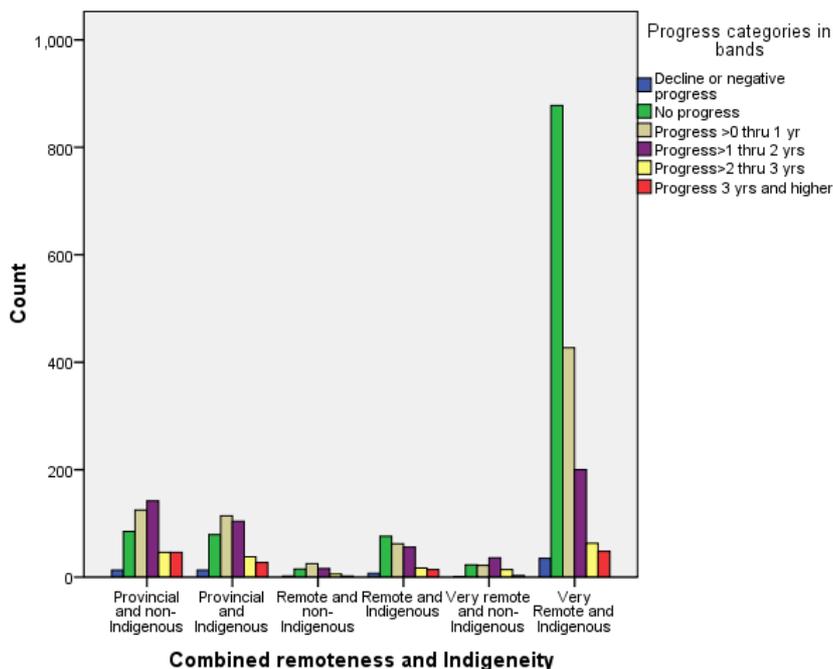
For the ToRCH-assessed students, the Indigenous “gap” appears to be the most important source of variation in mean acceleration scores, with only marginal differences obvious between the regions – in fact the very remote Indigenous group here approximated very closely to that of the remote group (Alice Springs and Katherine).

Figure 4: Error bar comparison of means of accelerated progress for ToRCH (95% CI) by Indigeneity of student and remoteness of school



For the IL sample, however, the proportions are dramatically illustrated in Figure 5 which showed numbers of students in each progress band in years across the regions. This shows the great over-representation 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.

Figure 5: IL progress band by region and Indigenous status groupings (frequencies)



The low rates of accelerated progress for Indigenous students from very 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). There was a very high proportion (53%) of average rates of zero progress across sequences and a strong relationship between lower initial or pre-program scores and overall rates of progress.

Explanation of Outcomes

There were considerable differences in the segments of the study samples defined by rates of acceleration, according to regional remoteness, Indigenous status and linguistic background. For example, 37% of female students aged 11 through 14 years in very remote regions achieved accelerated progress. This was well above average for the full IL sample and points to the possibility that there may be different growth rates at key periods of student learning in some groups. Indigenous students in Alice Springs and Katherine have higher average rates of acceleration than their non-indigenous peers, a finding favourable for the hypothesis of a catch-up of Indigenous and non-Indigenous populations in these centres. Non-indigenous students in very remote regions have the highest rates of all six groups identified by Indigenous status and region, outperforming their non-indigenous peers in Darwin schools. This works against the idea of remoteness by itself explaining poor outcomes and points to the need for further analysis of causes of difference between sub-populations, in this case, the distinctive characteristics or experiences of non-indigenous learners in remote communities.

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, when between-school differences are controlled for, it has only a very small influence on acceleration and has no effect at all for the IL sample (excluding those students with a zero IL score on both assessments) or the ToRCH sample. This is an unexpected finding, since rates of attendance are commonly seen as a necessary, if not sufficient, condition for Indigenous literacy progress. While attendance is justifiably pursued as a general policy priority, findings suggest that the priority for AL should be a differentiated approach: to schools in terms of teacher training and adequacy of support, and to students on the basis of their reading level at point of entry to the AL program.

It must be noted that all of the influences in combination explained only a fraction (up to 20%) of the total variation in individual student rates of reading progress. Most of the variation in the distribution of assessed rates of reading progress appeared to be found in differences between students, teachers and classrooms. For this reason, and notwithstanding the need to target efforts to improve teaching performance, it would be misleading and unfair to use the existing outcomes measures as an indicator of the "performance" of schools or to rank schools according to their outcomes using these measures. Even with adjustments of scores

for explanatory factors such as region and Indigenous status, comparison of school performance based on measured outcomes is of doubtful value for evaluation, for policy and management, or for feedback to practitioners.

Indigenous Status and its Correlates

The findings suggest that Indigenous status as an explanatory factor needs to be put into a wider causal context. For both IL and ToRCH assessments, there is little support for the case that negative effects of Indigenous status on reading progress can be explained solely in terms of remoteness or regional disadvantage.

The effects on accelerated progress rates for both Indigenous status and remoteness predictors are expressed in different causal modalities. This analysis shows 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 other findings of this evaluation indicate that very remote schools experience lower levels of support and higher rates of teacher turnover, these factors do not account for the differences in outcomes between very remote Indigenous students and all other groups.

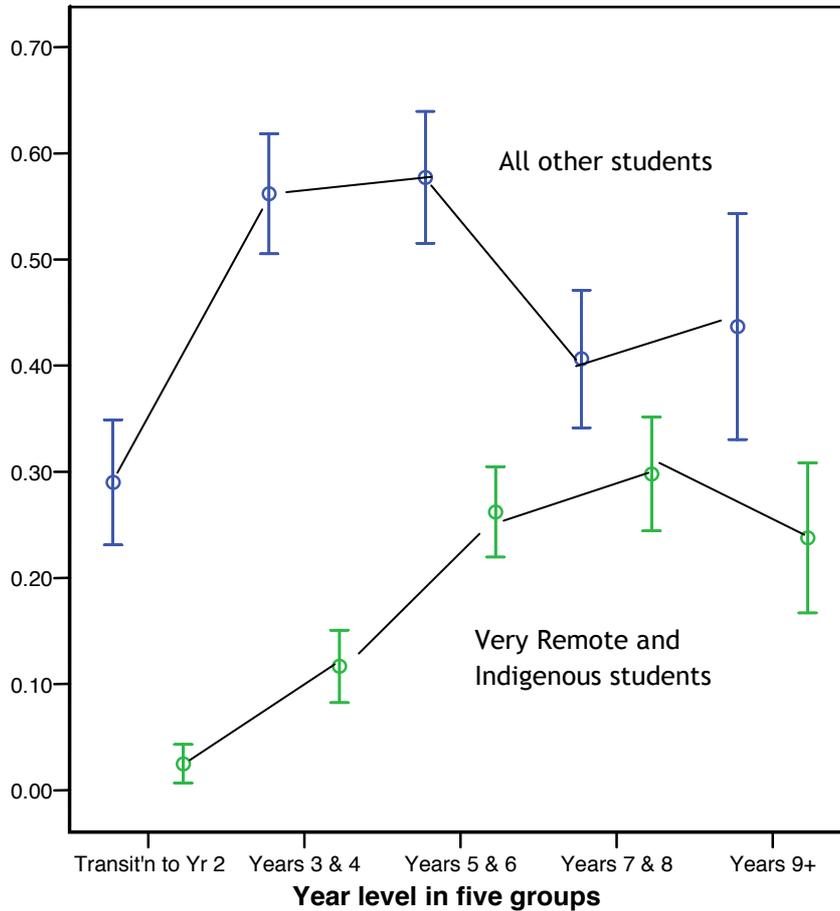
Lower Initial Reading Score and Very Remote Students

The low rates of accelerated progress for the Indigenous students from very remote community schools on the IL test were associated with a low initial reading age score of below Year 1 reading year level, compared with just over Year 3 reading year level for the other subsamples combined. This group also has a very high proportion (53%) of average rates of zero progress across sequences and a strong relationship between lower initial or pre-program scores and overall rates of progress. The factor of reading age interacts with the other factors outlined above to produce a very powerful effect, with the lowest rates of progress among Indigenous students from very remote backgrounds who are younger and have a lower initial reading age at first assessment.

The following figure compares acceleration rates for these two groupings in the IL sample from transition through to middle school. For both groups, the first level (from Transition to Year 2) shows the lowest proportion of accelerated learners. From Year 9 and higher, however, 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. For the Very Remote and Indigenous students there is a positive linear relationship between year level and proportion showing acceleration, but 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, after years 5 and 6, exposure to AL is linked

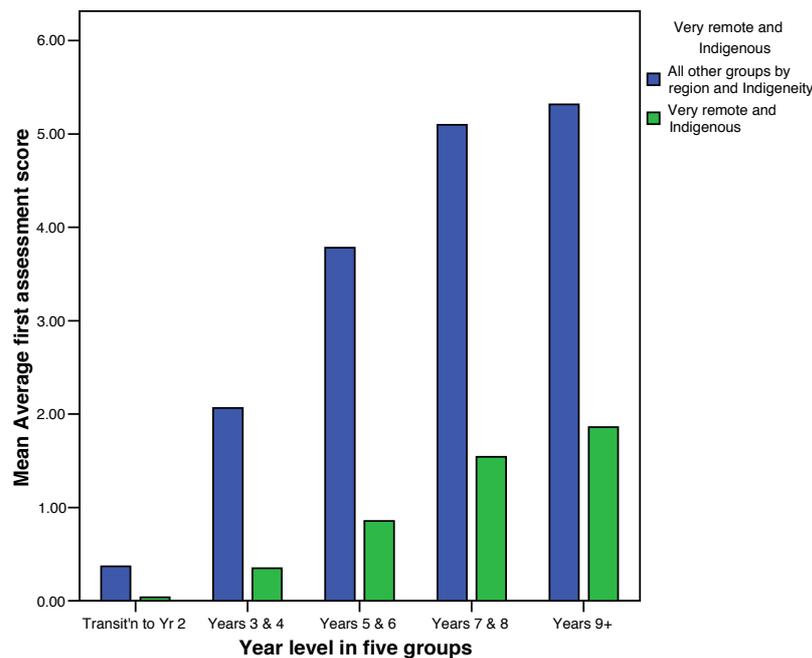
to a decline of about 15% in the proportions of students who are accelerated learners, albeit from a relatively high proportion of around 60-65% of accelerated learners between years 3 to 4 and years 5 to 6.

Figure 6: Error bar comparison (95% CI) of proportions showing accelerated progress (0-1.00), Very remote and Indigenous students and all others, IL sample



Although Figure 6 presents slight evidence that the gap in proportions showing accelerated progress is less pronounced between this group of AL students and others by about year 7, the analysis of initial reading ages shows that this is not sufficient improvement to “close the gap” between outcomes for Very Remote Indigenous students and the remainder of the student population.

Figure 7: First assessment age, comparison of group means, very remote Indigenous students and all others, IL sample



As Figure 7 shows, based on initial assessment scores, a gap of about three reading years is established between very remote Indigenous AL students and all other AL students by about the third school year. This gap (both in reading years and in proportions showing accelerated learning) persists through middle school.

A note on validity of assessments for “non-readers”

Because of the lack of information provided by the IL assessment for the large number of non-readers, supplementary analysis of data from five AL schools (two provincial/urban, three very remote) using the LLANS assessment instrument (Meiers & Forster, 1999; Meiers & Rowe, 2002) was conducted. This study indicated that just under 40% of students who appeared as “non-improvers” on the IL assessment (88% percent 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 compared with much lower rates in the remote schools. The results thus strongly suggest that a switch to the use of a more sensitive evaluative tool such as LLANS for NALP-related progress would be likely to accentuate, rather than reduce, the literacy “gap” between urban and remote schools, as well as between very remote Indigenous and other students. It also confirms the significance of reading age at first assessment and the importance of the early years in establishing the gap between students. In answer to the first hypothesis concerning the validity of evaluation instruments, for both evaluative and formative reasons, the analysis indicated that an additional or different mode of assessment for early readers and non-readers should be adopted.

Conclusions: Between Program Implementation and Learning Outcomes

Within the limits of the research design, the results of multi-level analysis of system-wide outcomes confirmed that the NALP in the Northern Territory is delivering accelerated reading progress of greater than one reading year level per year according to benchmarked assessments for a substantial proportion of students, mainly urbanized, who have achieved at least Year 4 levels of reading performance. That is, there are positive signs that AL is able to contribute to accelerated learning for Indigenous students in the major centres of Darwin, Alice Springs and Katherine. However, with current levels and methods of delivery of AL in NT schools, for most of those students who, for reasons of cultural, linguistic and situational factors, have not yet established basic reading competence, the evidence suggests that Accelerated Literacy has yet to demonstrate the capacity to promote substantial accelerated gain.

These findings must be considered against the findings of direct classroom observation, surveys and interviews reported elsewhere. They showed that the level and quality of implementation achieved in NT schools by the end of 2007 had not been sufficient for by far the majority of school-based coordinators and teachers to achieve standards of teaching consistent with the requirements of the AL method. There was evidence of provision of significantly lower levels of in-school support to the very remote schools, where teachers were less experienced and had shorter tenure at their school. In short, there was less local capacity of many very remote schools to sustain the AL method. Analysis of both survey data and observational outcomes showed that the total number of hours of PD workshops and externally provided support by the DET AL team was the only significant predictor of teaching performance identified. Mean levels of support received by teachers were around 25 hours, with 50 hours of total PD support associated with the standards of teaching performance required. Further, schools in major centres (those who contributed most students to the ToRCH sample for the multilevel analysis) had teachers with more AL experience and longer tenure in schools (significantly in the case of Katherine and Alice Springs).

However, despite the better performance of students in the generally better resourced schools in major centres, the findings of the multilevel analysis suggest that the bifurcation in outcomes between urban and remote contexts will not be addressed simply by increasing the training and support provided to very remote schools, necessary as this may be as a first step. The program's effectiveness seems most plausibly to be demonstrated for students with a reading age at or greater than 4, that is, middle and upper primary levels, and in particular with students from backgrounds where English is spoken at home. Indeed, the early "evidence" for the program's effectiveness was based on exclusion of students under this reading age. After exclusion, it provided simple unadjusted percentages of students showing acceleration above benchmarked year level reading performance for the rest (cf. Gray, 2007).

By contrast, the analysis here includes all participating students and adjusts for all major contextual determinants before attributing outcomes and highlights very

different outcomes for major subgroups of the student population. It shows that to be effective at a population level, the AL method itself may need review and further development for important target groups of students in the NT population.

It might be argued that AL was a victim of its own policy success, in that it was subject to rapid and partial implementation to meet pre-given targets, and was encouraged to expand its scope to include early childhood and beginning readers without accompanying development and testing of the effectiveness of its methods for these groups. It is quite clear, after our investigation, that the potential of AL to contribute to improved outcomes at a population level would need to be supported by much more rigorous evidence of the effectiveness of its strategies for these groups. The maintenance of both practitioner and policy support no less require the continuing development of this evidence to a higher standard than has been the case in the past.

Acknowledgements

The research reported here was funded under contract to the NT Government Department of Education and Training according to the terms of funding agreements with the Commonwealth Department of Education and Workplace Relations for the National Accelerated Literacy Program. Ethics approval was granted by the CDU HREC, Approval no H05073. The authors acknowledge the support of all members of the evaluation team, and colleagues in NT Schools and the NALP.

References

- ACER (2003). *Tests of reading comprehension* (2nd ed.) Canberra: ACER Press.
- Cresswell, J., Underwood, C., Withers, G, 2002, *Evaluation of the University of Canberra Programme for Advanced Literacy Development Scaffolding Literacy Programme for Indigenous Children in School*, Report prepared for DEST, December. Victoria, ACER.
- Flay, B. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.
- Gray, B. (2007). *Accelerating the literacy development of Indigenous students*. Darwin, NT: Charles Darwin University.

Gray, B., & Cowey, W. (2005). Key Elements of Scaffolding Literacy: Pedagogy and Teacher Support. Retrieved from <http://www.canberra.edu.au/centres/schools-community/literacy-research>

Griffin, P., Woods, K., & Nguyen, C. (August, 2005). *An environmental scan of tools and strategies that measure progress in school reform*. [http://www.det.vic.gov.au/edulibrary/public/publ/research/publ/Scan_of Tools Measuring Progress in School Reform 2005-rpt.pdf](http://www.det.vic.gov.au/edulibrary/public/publ/research/publ/Scan_of_Tools_Measuring_Progress_in_School_Reform_2005-rpt.pdf).

Lea, T., Robinson, G., Abu-Duhou, I., 2006, *NALP Evaluation Interim Report*, Darwin, Accelerated Literacy Evaluation Team, Charles Darwin University, November.

Meiers, M., & Forster, M. (1999). *The Longitudinal Literacy and Numeracy Study (LLANS)*. Paper presented at the ACER Research Conference 1999, Improving Literacy Learning, Adelaide

Meiers, M. & Rowe, K. (2002). *Modelling children's growth in literacy and numeracy in the early years of schooling*. Paper presented at the 2002 Annual Conference of the Australian Association for Research in Education, University of Queensland, December 1-5, 2002 [Paper code MEI02176]

Pawson, R., & Tilley, N. (2000). *Realistic evaluation*. London: Sage.

Robinson, G., Lea, T., Rivalland, J., Tyler, W., Bartlett, C., Morrison, P., Cooper, J., Emmett, S., 2009a, *The National Accelerated Literacy program in the Northern Territory 2004-2008, Implementation and Outcomes, Volume 1, Final Evaluation Report*, Darwin, School for Social and Policy Research, Charles Darwin University

Robinson, G., Lea, T., Rivalland, J., Tyler, W., Bartlett, C., Morrison, P., Cooper, J., Emmett, S., 2009b, *The National Accelerated Literacy program in the Northern Territory 2004-2008, Volume 2, Appendices, Final Evaluation Report*, Darwin, School for Social and Policy Research, Charles Darwin University

Rowe, K., Devine, M., Knight, F., Loudon, W., Lovett, T. Myer, Y. Ramsay, G., Rice, A., Scalfino, L., Smith, K. (2005a) *Teaching Reading: Report and Recommendations National Inquiry into the Teaching of Literacy*. Commonwealth of Australia: Canberra

Rowe, K. J. (2003). Estimating interdependent effects among multilevel composite variables in psychosocial research: An annotated example of the application of multilevel structural equation modelling. In N. Duan & S. Reise (Eds.), *Multilevel modelling: Methodological advances, issues and application*, 255-284. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rowe, K. J., & Hill, P. W. (1998). Modelling educational effectiveness in classrooms: The use of multilevel structural equations to model students' progress. *Educational Research and Evaluation*, 4(4), 307-347.

Rowe, K. J., & Rowe, K. S. (1999). Investigating the relationship between students' attentive-inattentive behaviors in the classroom and their literacy progress. *International Journal of Educational Research*, 1(21), 1-138.