

Profiling reading comprehension in Mangere Schools: a research and development collaboration

Lai¹, M.K., McNaughton², S., MacDonald², S., Hall³, A., MacDonald³, B., McKee³, D., Nicholls³, J., Reeves³, J., Swann³, J., Valgrve³, D., Weir³, P., Farry², S & Warren⁴, S.

Affiliations:

1. Mangere AUSAD School Improvement Initiative
2. Woolf Fisher Research Centre
3. Mangere schools
4. Ministry of Education

This research and development collaboration was made possible through the partnership of the affiliated groups and government support.

Over recent years there has been a strong focus on understanding the developmental features of early literacy and characteristics of effective instruction at the early stages of learning to read and write at school. In the United States, the National Research Council released a report by a panel of experts called Preventing Early Reading Difficulties (Snow, Burns & Griffin, 1999), and in New Zealand a Literacy Task Force (1999) and a Literacy Experts group (1999) reported to government, both focusing on instruction in the early years at school.

A major goal in the New Zealand case has been to understand and reduce disparities in achievement for Maori and Pasifika students, and children in low decile schools in the first four years of school. There has been a number of national and local initiatives since 1998 (Ministry of Education, 2002) and there is evidence for increased effectiveness of early literacy instruction. The National Education Monitoring Project's (Flockton & Crooks, 2001) second cycle of assessments of reading showed reduced numbers of children in the lowest bands at year four in reading accuracy. A recent renorming of standardised assessments at year 1 (6,0 years) conducted in 2000 suggests knowledge of letters and sounds is higher (Clay, 2002). Research-based interventions in low decile schools have demonstrated teachers can raise rates of progress significantly across a broad band of early literacy measures including accurate decoding of text (Phillips, McNaughton & MacDonald, in press).

But a major challenge has been created by these advances. In part this is an issue of sustainability, succinctly identified by The RAND study group, set up in 1999 by the US Department of Education's Office of Educational research and Improvement to identify the most pressing needs for reading research. The authors of the report say:

We have made enormous progress over the last 25 years in understanding how to teach aspects of reading. We know about the role of phonological awareness in cracking the alphabetic code, the value of explicit instruction in sound – letter relationships, and the importance of reading practice in producing fluency. ...The fruits of that progress will be lost unless we also attend to issues of comprehension. Comprehension is, after all, the point of reading.' (Sweet & Snow, 2003, p xii)

The second part of the challenge has been repeatedly noted by Michael Pressley (2001) as a failure to turn what we know about comprehension and the teaching of comprehension into generally more effective teaching in schools. The issue Pressley identified was one of knowledge transfer. It is how to add the knowledge about comprehension that has been developed in recent research to what is practiced in schools in a way that enables schools to develop effective and sustainable practices.

The situation in New Zealand is critical. The snapshots provided by the National Educational Monitoring Project show that despite the gains in decoding, there are wide and increasing disparities in achievement on comprehension tasks for Maori and Pasifika children and particularly in low decile schools (Flockton & Crooks, 2001). The AsTTle project collected national data also in across multiple dimensions of comprehension and

found large differences between Maori and Pasifika children which were stable across decile levels despite significant trends across deciles (Hattie, 2002).

The present report is the result of a cluster of seven decile one schools developing a partnership with researchers to meet the RAND group's challenge of building sustainability and Pressley's challenge of research-based applications of best practice. They have implemented effective instructional programmes for beginning literacy instruction. These school-based projects have indicated what an exercise would look like to understand teaching and learning needs, and the conditions for sustainability (McNaughton, Phillips & MacDonald, 2003; Timperley, 2003). One is that systematic data on both learning and teaching would need to be collected by the schools and analysed by the schools. That assessment data would need to be broad based in order to know about the children's patterns of strengths and weaknesses, to provide a basis for informed decisions about teaching, and to clarify and test hypotheses about how to develop effective and sustainable practices.

Previous assessments collected individually by schools using their batteries of standardised achievement tests suggested generalised low levels of comprehension on these tests. Several general hypotheses are possible for the low levels, which can be checked and refined in the early stages of a research and development exercise in schools. One is that children's comprehension levels are low because of low levels of accurate and fluent decoding (Tan & Nicholson, 1997). A second is that children may have learned a limited set of strategies, for example they may be able to recall well but are weaker in more complex strategies for drawing inferences, synthesising and evaluation or children may not have been taught well enough to control and regulate the use of strategies (Pressley, 2002).

Other possible contributing reasons might be more language based – that children's vocabulary may be insufficient for the texts used in classrooms tasks (Biemiller, 1999), or that they are less familiar with text genres. Well known patterns of 'Matthew effects' may be present in classrooms, where culturally and linguistically diverse children receive more fragmented instruction focused on decoding or relatively simple forms of comprehending, or they receive relatively less dense instruction, all of which compounds low progress (McNaughton, 2002; Stanovich, West, Cunningham, Cipielewski, & Siddiqui, 1996). There is also a set of hypotheses around whether the texts, instructional activities and pedagogy of the classroom enable cultural and linguistic expertise to be incorporated into and built on in classrooms (Lee, 2000; McNaughton, 2002).

The significance of collecting and analysing data, rather than making assumptions about what children need (and what instruction should look like) without careful analysis, recently was underscored by Buly and Valencia (2002). Policy makers in the State of Washington had mandated programmes without actually analyzing profiles of low progress students identified by test scores from fourth grade National Assessment of Educational Progress (NAEP) scores. The assumption underlying policies and interventions was that poor performance reflected students' difficulties with more basic decoding abilities. Yet, there was little data about this assumption or to know whether

focusing on such skills would improve the comprehension at 4th grade. Using a broad band of measures Buly and Valencia identified five groups of low progress readers, some of whom did indeed have limited fluency and accuracy in decoding. However, mandating phonics instruction for all students who fell below the proficiency levels had missed the needs of the majority of students, whose decoding was strong but who struggled with comprehension or language requirements for the tests. This finding highlights the need for research-based applications of best practice based on analyses of student needs.

This present study is Phase One of a three-year research and development partnership between schools and researchers to raise student achievement in reading comprehension through the development of research-based applications of best practice. In this phase, we examine achievement in reading comprehension including patterns of strengths and weaknesses and observe classroom practice to understand teaching and learning needs.

Methods

Participants

The overall partnership involves schools in the Ministry of Education Mangere AUSAD (Analysis and Use of Student Achievement Data) school improvement initiative, the initiative leaders, the Woolf Fisher research Centre and Ministry of Education representatives. Baseline data were collected from 1975 students in seven Mangere schools, all of which were decile one¹ schools. The students involved in the analysis were from years four to nine and consisted of equal proportions of males and females (50% and 50% respectively) from 14 ethnic groups. Four main ethnic groups made up 87% of the sample. These groups were Samoan (33%), Maori (20%), Tongan (19%) and Cook Island (15%) ethnic groups.

Tools

Data on reading comprehension were collected using the revised Progressive Achievement Tests in Reading (reading comprehension section only) (Reid & Elley, 1991) and the Supplementary Tests of Achievement in Reading (Elley, 2001). The tests were what schools had decided as a group to collect to measure reading comprehension because they provided a recognized, standardized measure of reading comprehension which could be reliably compared across schools.²

The revised Progressive Achievement Tests (PAT) in Reading measure both factual and inferential comprehension of prose material in years four to nine. Each prose passage

¹ A school's decile indicates the extent to which the school draws its students from low socio-economic communities. Decile 1 schools are the 10% of schools with the highest proportion of students from low socio-economic communities. A school's decile does not indicate the overall socio-economic mix of the school.

² At the time of this decision-making, not all schools had introduced AsTTle - Assessment Tools for Teaching and Learning (Ministry of Education, 2003) - into their schools and/or had not yet received sufficient professional development to use it effectively for these purposes.

consists of 100-300 words and is followed by four or five multi-choice options. The prose passages are narrative, expository and descriptive, and different year levels complete different combinations of prose passages. The proportion of factual to inferential items per passage is approximately 50% - 50% in each year level.

The Supplementary Tests of Achievement in Reading (STAR) were designed to supplement the assessments that the teachers make about students' "close" reading ability in years four to nine (Elley, 2001). In years four to six, the test consists of four sub-tests measuring word recognition (decoding of familiar words through identification of a word from a set of words that describe a familiar picture), sentence comprehension (complete sentences by selecting appropriate words), paragraph comprehension (replace words which have been deleted from the text in a 'Cloze' format) and vocabulary range (find a simile for an underlined word). Only the paragraph comprehension sub-test is not multi-choice and consists of 20 items, 10 more than the rest of the sub-tests. In years seven to nine, students complete two more sub-tests in addition the four sub-tests described above. The sub-tests are the language of advertising (identify emotive words from a series of sentences) and reading different genres or styles of writing (select phrases in paragraphs of different genres which best fits the purpose and style of the writer) In years seven to nine, there are 12 items per subtest except for paragraph comprehension which consists of 20 items. Both tests have high reliability and validity (Elley, 2001; Reid & Elley, 1991)

The correlation between the two tests was $.62 (P < .01)^3$. In the test manual, Elley (2001) reported correlations between 0.70 and 0.78 for year four to eight students. This, as Elley suggests, indicates that the tests measure similar but not identical facets of reading comprehension.

Procedure

Representatives from seven schools formed a Senior Assessment Team (SAT) as part of the Ministry of Education Mangere AUSAD initiative to work with the researchers, the Ministry of Education and the leaders in the initiative on developing an intervention to raise student achievement. Decision-making on the research is collaborative and any ethical requirements brokered through this team.

At the beginning of the year, SAT developed an intra-school standardized process of administering the test and moderating the accuracy of teacher scoring. This involved standardizing the week and time (morning) of testing and creating a system of randomly checking a sample of teachers marking for accuracy of scoring. Accuracy of scoring was further checked by the data entry team from Woolf Fisher Research Centre during data entry and during analysis. The STAR and PAT were administered as part of schools' normal assessment cycle at the beginning of the school year.

³ Correlations were only performed on students who had sat both tests. (n=1339). Thirty two percent of students did not sit either the PAT or STAR test. The high proportion was due to incomplete data from one school.

Analysis

Data was analysed in several phases and through several forums. Area-wide data was analysed in collaboration with SAT in two meetings, then analysed by senior managers and senior teachers with support from the first author. Four of the seven schools with support from the first author conducted further analysis sessions with teachers. Analysis sessions involved examining and evaluating both researchers and practitioners' theories of the patterns of achievement, rather than privileging either researchers' or practitioners' theories.

Data were analysed using descriptive and inferential statistics to examine achievement in reading comprehension across the year levels, including content analysis to determine student strengths and weaknesses. The latter involved analysing student scores on factual and inferential questions in the PAT, and analysing subtest scores in STAR. In STAR, this also included qualitatively coding the types of errors students made on the cloze passage according to the types of errors reported in the STAR manual (Elley, 2001). Four raters were trained to code the errors. These raters subsequently discussed how to code the errors and collectively rated a sample of tests to determine reliability of coding. The coding was subsequently checked and inter observer agreement on 10% of students' subtests (across ages) was 90.5%. Gender and ethnic breakdowns are also included in this paper.

Classroom Observations

Systematic classroom observations were carried out by the second author. These were designed to provide a sample of how features of teaching and learning might map on to the achievement data. Our argument was that a full problem analysis was needed to examine classroom instruction otherwise assumptions about what was or was not being taught would be unchecked. Observations were carried out in 16 classrooms in seven schools from year four/five through to year eight, (including one bilingual Samoan classroom). Classroom instruction was observed for 40 – 60 minutes during the usually scheduled core reading session within which the teaching and learning of reading comprehension occurred. Class sizes generally ranged from 21 – 26. A combination of diary and audio recording of specific activities occurred. Discussions with teachers also provided an important level of professional reflection. Specific details from these observations are reported further in the paper when the student achievement profiles are discussed.

Observations revealed that the general programme was similar in most classes. A whole class activity typically was followed by group work, the (2 – 5) groups organised by ability level. In several classes the whole class activity involved the teacher introducing and sharing a narrative text. The activity often took place over 20 minutes. For the following (up to) 40 minutes, small group activities occurred. These included text study and analysis (such as study of plot or character in narrative texts and extracting and using information in informational texts), specific group or paired forms of instructional / guided reading (such as 'reciprocal teaching'), and individual or group project work (such as developing taxonomies of categories in science topics). Typically, the teacher

worked with two groups over this time period and held conferences on the run with other groups.

The general programme appeared to work well in terms of levels of engagement, which were high, with routines well established and frequent instances of teacher-student and student-student interactions. The general organisation meant whole class activities occurred on 3-5 days per week and small group work with one or two groups often daily, so that each group had at least one session but up to three sessions with direct teacher guidance each week. However, the variation in frequency of contact with each group was quite marked between schools, and in some classrooms few specific interactions with individual children per session. When not with the teacher, groups engaged in a range of activities. Some had developed to the point of being able to operate just with peer guidance in reciprocal teaching. In most classrooms worksheets were used, which contained questions about a text and often contained sentence, word or sub word studies. Texts and topics were well planned to be engaging and at appropriate text reading levels.

Results

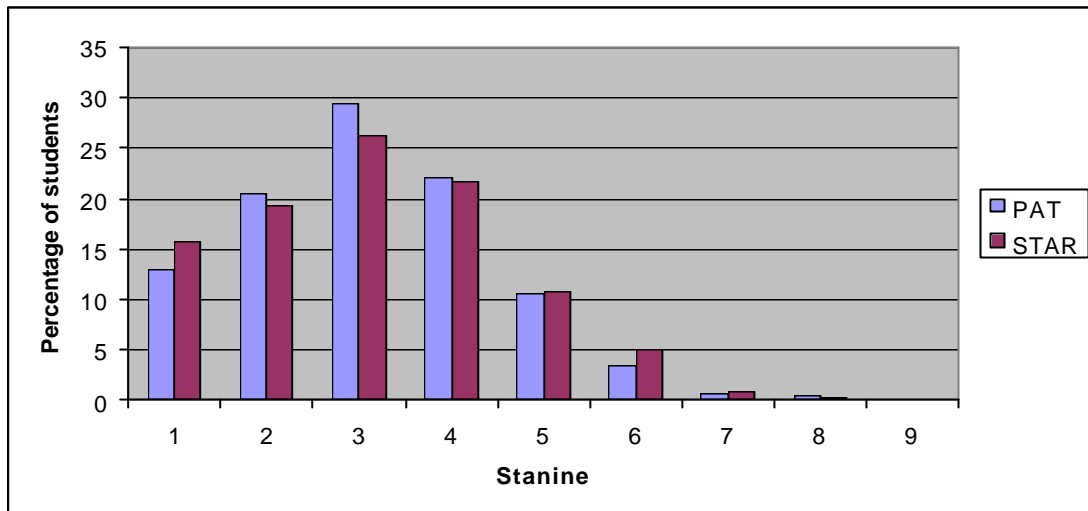
The results are presented in three sections: The general profile of reading comprehension, content analysis of PAT and STAR, and gender and ethnic group breakdowns.

General profile of reading comprehension

The stanine⁴ distribution of both tests indicate that the average student experienced difficulty on these measures of reading comprehension. Figure 1 shows the stanine distribution in both tests across all year levels. The average student in both tests scored in the “below average” (stanine two and three) band of achievement (PAT mean=3.10, SD=1.38; STAR mean=3.10, SD=1.50). This indicates that the average student was just below the “average” band (stanines four to six) and was two stanines below the expected average of stanine five. However, nearly 25% of students were within the “average”, “Above average” and “superior” bands of achievement (stanine five to nine).

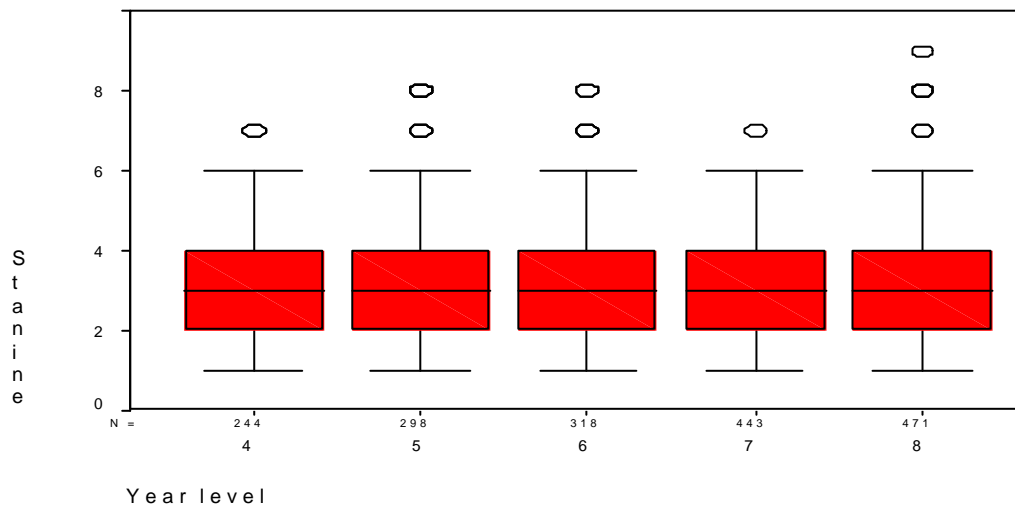
⁴ Stanines are normalized standard scores having a mean of five and a standard deviation of about two (Reid & Elley, 1991). They are expressed as a scale of nine units with a low of one and a high of nine. In the PAT manual, stanine nine is described as “superior”, stanine seven and eight as “above average”, stanine four to six as “average”, stanine two and three as “below average” and stanine one as “low” (Reid & Elley, 1991, p. 23). The nine stanine units may be considered as nine categories of reading attainment, making it “highly suitable for interpreting performance on the PAT:Reading” (Reid & Elley, 1991, p.23).

Figure 1
PAT and STAR stanine distribution across all year levels



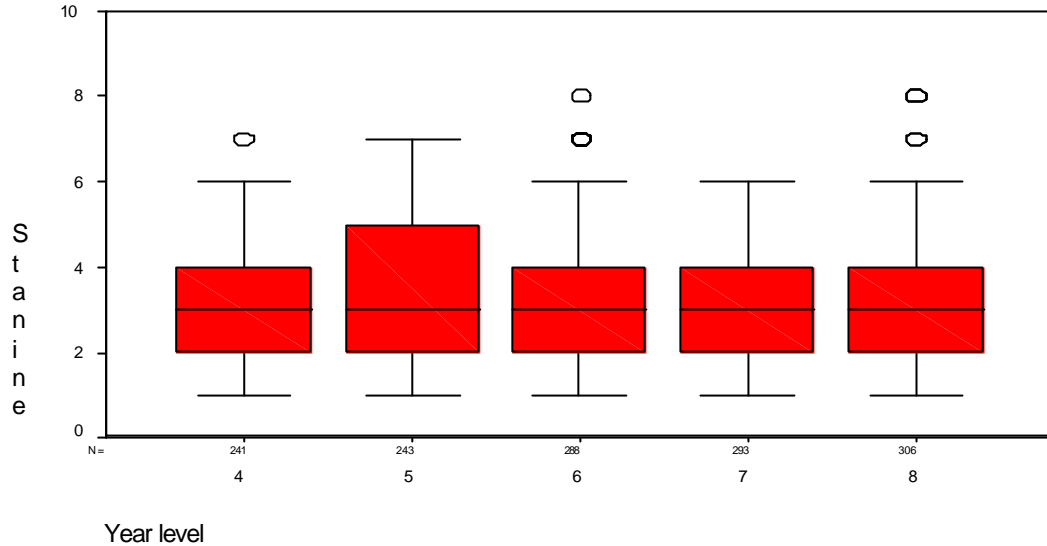
Across the year levels, the pattern was the same in both tests, with the average student in every year level scoring at stanine three⁵. This is displayed graphically in Figures 2 and 3. The range of achievement was large, from stanine one to nine in the PAT and one to eight in STAR.

Figure 2
Stanine Distribution for PAT in years four to eight



⁵ For reasons of confidentiality, year nine students were not included in this, or any like, analysis, as there is only one school with year nine students.

Figure 3
Stanine distribution for STAR in years four to eight

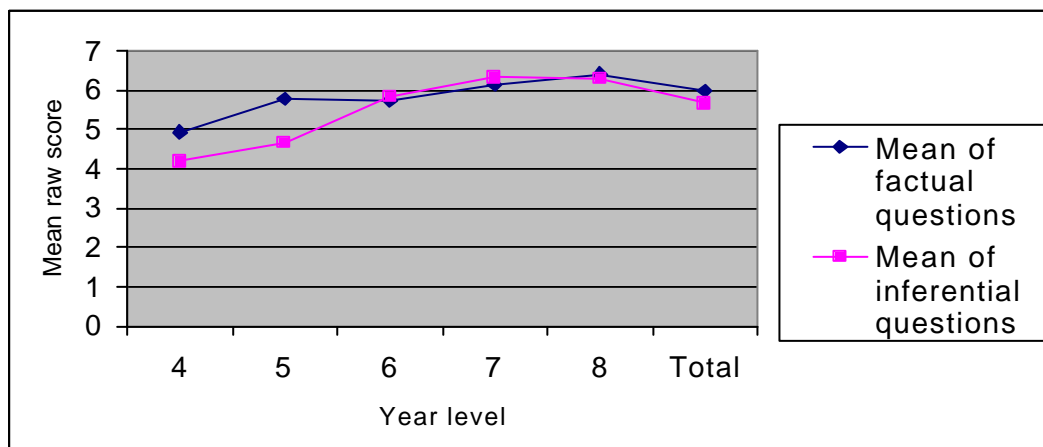


Content analysis

PAT content analysis

Mean scores on factual and inferential questions in the PAT were virtually identical across all the year levels (see Figure 4). Note that the maximum raw score for both factual items and inferential items was approximately 20 (Reid & Elley, 1991). This suggests that students experienced similar difficulties in answering factual and inferential questions. There was a significant correlation between factual and inferential items ($r = 0.61, p < .01$).

Figure 4
Mean raw scores on factual and inferential questions by year level



This pattern was constant across the year levels, suggesting that students of different year levels experienced similar difficulties in answering factual and inferential items. Table 1 presents the breakdown of factual and inferential questions for PAT reading comprehension across the year levels.

Table 1
Means (and standard deviations) of factual and inferential questions across the year levels

| Year level | N | Factual questions | Inferential questions |
|------------|------|-------------------|-----------------------|
| 4 | 245 | 4.93 (2.80) | 4.18 (2.36) |
| 5 | 298 | 5.76 (3.32) | 4.64 (2.76) |
| 6 | 319 | 5.75 (3.04) | 5.85 (2.76) |
| 7 | 390 | 6.14 (2.82) | 6.31 (2.67) |
| 8 | 471 | 6.41 (3.10) | 6.30 (3.12) |
| 9 | 144 | 6.57 (3.30) | 5.92 (2.51) |
| Total | 1867 | 5.96 (3.08) | 5.65 (2.90) |

Content analysis on the STAR subtests

Analysis of the subtests of STAR revealed consistent patterns across the subtests at each year level. Figures 5 and 6 show the average percentage obtained in each subtest. At every year level, students scored highest on subtest one (Word recognition) and lowest on subtest three (paragraph comprehension), indicating that students in all year levels experienced more success in decoding words than comprehending a paragraph. All the sub-tests of STAR were significantly correlated ($p < .01$). A series of paired t tests between sub tests averaged across years revealed that at both age groupings, the means for sub test 1 were significantly higher than the means for the other sub tests (t values > 18.0 ; $p < .000$), and sub test 3 means were significantly lower than each of the other sub tests (t values > 12.0 , $p < .000$). In addition, in the age group y4 – 6, all sub tests were significantly different from each other. In the older age group, subtest 2 and 5, 2 and 6 and 5 and 6 were not significantly different ($p > .05$).

Figure 5
Average percentages obtained in each subtest (STAR) for year levels four to six

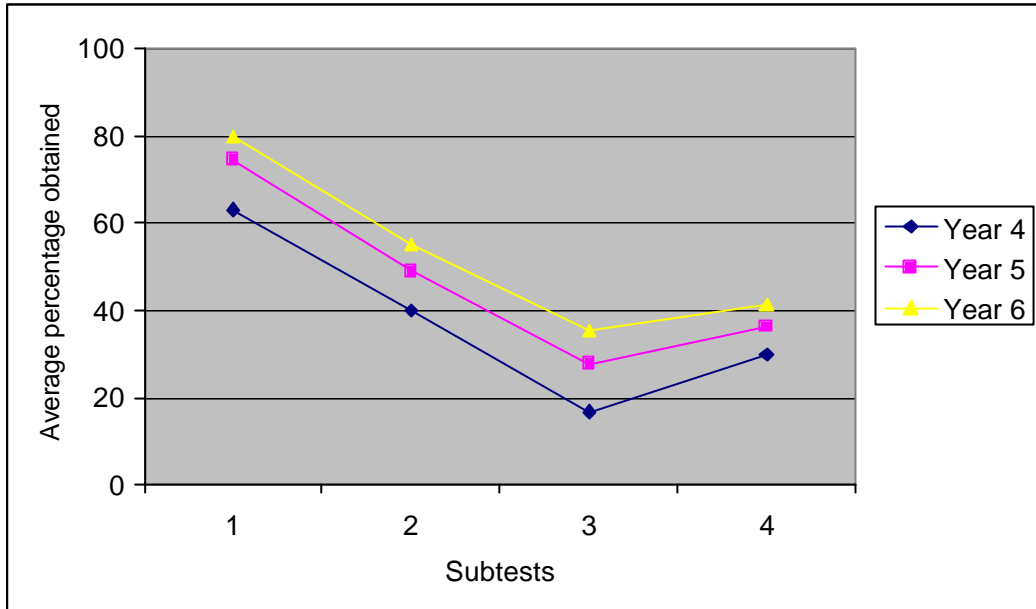
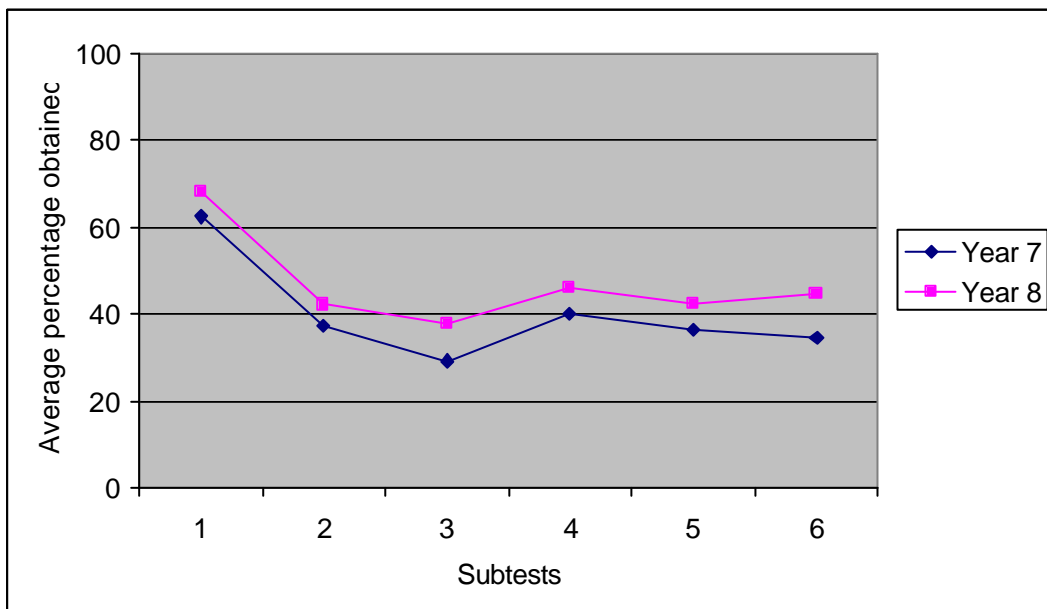


Figure 6
Average percentage obtained in each subtest (STAR) for year levels seven and eight



Error analysis of subtest 3 (paragraph comprehension – Cloze)

Errors on the third sub test (paragraph comprehension) were analysed according to the coding in the STAR manual based on common errors made by students such as “irregular plurals” and “tense problems”. However, nearly half (46%) of all possible errors across all year levels did not fit any of these categories. These errors appeared to be of two new types. One involved mistakes that sometimes made sense in the pre sentence context but resulted in nonsense or illogical sentences. The following examples show student responses in italics and the correct response/s is in brackets):

- ‘All they *did* (could) afford was a tiny room in a *shoe* (cottage/house/shop) in a village by a river.’
- ‘He grabbed frantically, and felt his *head* (hand(s)/finger(s)) closing around the branch of a tree.’
- ‘Suddenly, round a sharp bend in the *head* (road/path/track), he fell again, missed his *self* (footing/step) and plunged over the *ugly* (cliff/rock) face.’

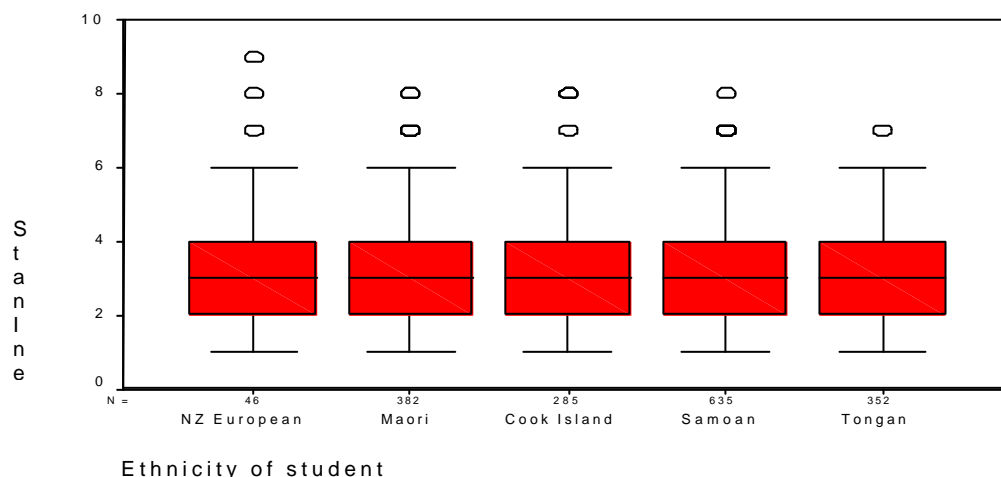
There were a few instances of a second type of error, where relatively close synonyms, not listed by the manual, were used.

- ‘It browses entirely on tall *branches* (trees/plants) especially the foliage of mimosa and acacia trees.’

Ethnicity and gender

Analyses were also conducted by ethnicity and gender. For the PAT (see Figure 7) the pattern of achievement was consistent across the major ethnic groups. Students, on average, scored at stanine three with 50% of students scoring between stanine two and four.

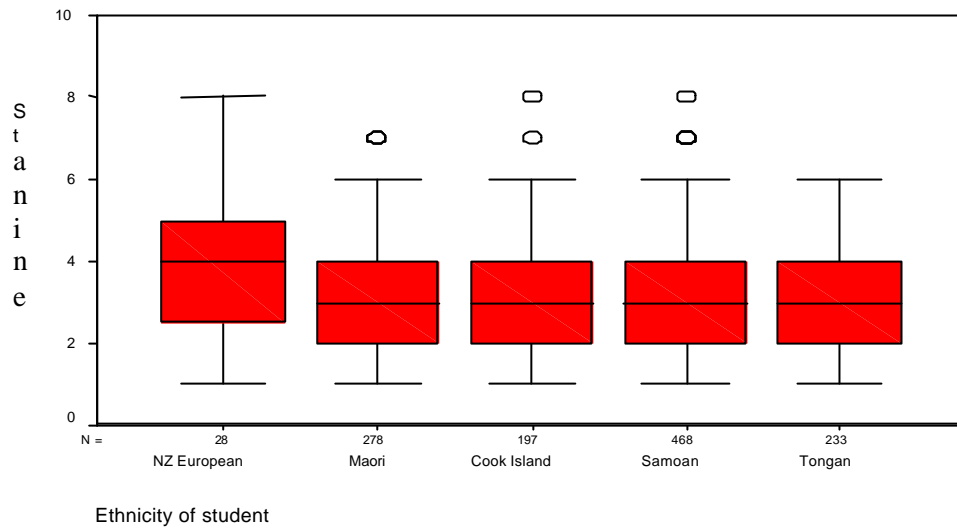
Figure 7
PAT stanine distribution by ethnic group



In the STAR test, however, the pattern of scores was slightly different between NZ European students and the students from the other ethnic groups. (see Figure 8) The NZ

European students scored, on average, at stanine four whilst the other ethnic groups scored, on average at stanine three. The pattern of scores on STAR and PAT, unlike the other ethnic groups, is also different for NZ European students with students, in general, obtaining higher stanines in STAR.

Figure 8
STAR stanine distribution by year level across ethnic groups



Similarly, there were few gender differences between males and females in both PAT and STAR (see Figures 9 and 10). Both males and females, on average, scored at stanine three with 50% of students falling between stanine two and four.

Figure 9
PAT stanine distribution by gender

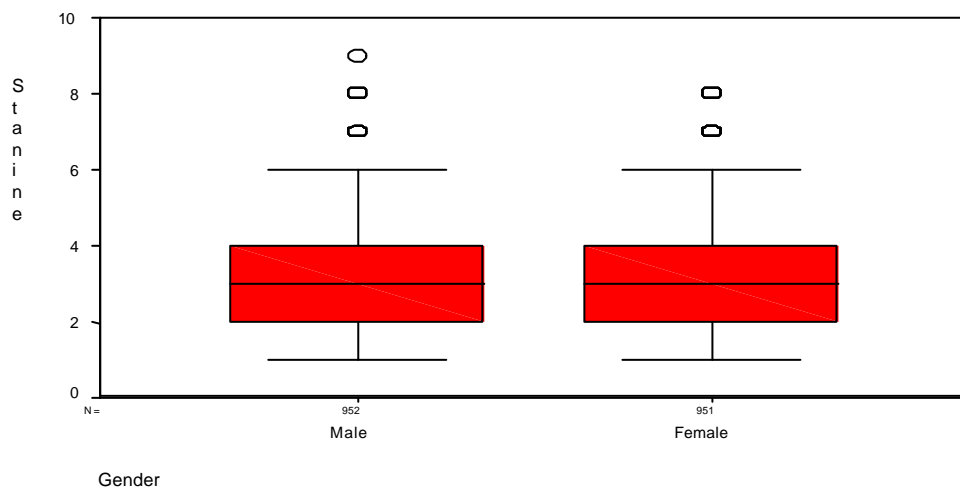
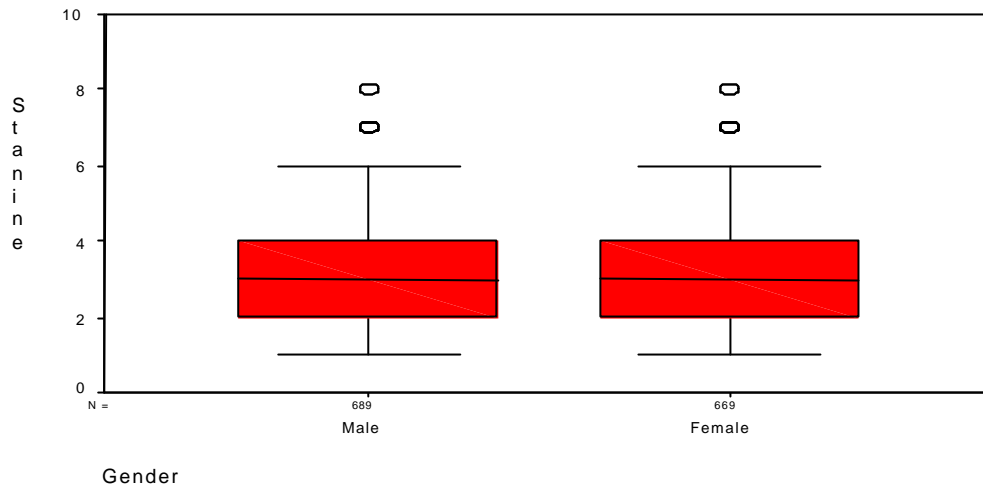


Figure 10
STAR stanine distribution by gender



Discussion

The profile of student's comprehension in the seven decile 1 schools confirmed previous descriptions of 'below average' levels in the middle to upper primary school years (Flockton & Crooks, 2001; Hattie, 2002). The profile was the same across age levels and across the two tests used. However, it is important to note the presence of variability within the profile. A quarter of the students were average or significantly above average in their achievement.

What can be determined from the patterns within and across tests? One hypothesis is that students have not developed fast and accurate decoding skills known to be necessary but not sufficient condition for effective comprehending in conventional school texts (Nicholson & Tan, 1997; Pressley, 2002). The findings of this study suggest that widespread problems with decoding skills are unlikely to be the underlying reason for the low PAT and STAR results. At every year level, the subtest "word recognition" in STAR was higher than any of the other subtests. On average, students got between 60% and 80% of the words correct, indicating ability to identify words accurately under timed conditions. These means are between 1.1 and 2 raw scores different from the means reported in the manual for the nation-wide sample, indicating that word recognition skills are similar to the country as whole. (Elley, 2001, states that only a raw score difference of 3 to 4 points can be considered significant.) Anecdotal evidence from senior managers further support this conclusion. Senior managers in the schools reported large numbers of students who score highly on decoding (as measured by running records), but perform poorly on the tests of comprehension such as STAR, PAT and comprehension questions on the running records.⁶

⁶ The disparity between decoding and comprehension scores was one of the primary reasons senior managers began this project with Woolf Fisher Research Centre on reading comprehension.

Further evidence which could support the conclusion that most students did not have problems reading the test was the completion rates of the PAT. The students' rates of completion in the PAT's were relatively high, so that by year eight, the rate of completion was 93%. (For example, in year four, 28.6% did not complete the test and in year six 11.5% did not complete the test.) But it could also suggest inaccuracy and high rates of guessing the questions. Indeed, analysis of the errors on the paragraph completion sub test of the STAR which has a Cloze format, indicates that there were high rates of guessing, or at least not checking answers.

Two further pieces of evidence support this proposition that over use of predicting or guessing without checking may have been occurring. Firstly, classroom observation data suggested students often engaged in 'predicting' during standard classroom reading activities. For example, in all of the guided reading lessons observed, students generated many predictions about narrative or expository content of texts, but rarely checked and were rarely prompted to check the accuracy of these predictions using evidence from the text. Systematic observations recorded prompting to check predictions only nine times in 16 hours of observations. Recent research evidence also has begun to describe a similar pattern that at least some low achieving students have. Dewitz and Dewitz (2003) described a small group of 5th grade readers who were fast, efficient decoders but had low comprehension scores. Error analysis revealed high rates of errors termed "excessive elaborations" (i.e. guessing).

A surprising finding was that the students generally were not better at factual questions compared with inferential questions. Previous research on family literacy practices for Samoan and Tongan children shows many children are experienced in recitation of texts and this might suggest that recall of facts might be a strength (McNaughton, 1995). One possible explanation for this relates to the pattern of guessing noted above, If students' strategies were focused on predicting (guessing) and were not checking answers, low accuracy rates for both types of questions could be expected. A further possibility is that the students do not have the lexical range in English required for these tests. The fourth of the STAR subtests which tests vocabulary range was low. Buly and Valencia (2002), found in their sample of 4th grade low progress children in Washington State that many of the children for whom English was a second language, had difficulties with word meanings on the English tests.

Another surprising result was the lack of gender difference. There are data showing that boys achieve less well on literacy and language-related school assessments. Recent Ministry of Education figures for school entry assessments and on up to national pass rates in secondary schools continue to show this (Ministry of Education, 2003). However, other studies of early literacy development and achievement have failed to find substantial significant differences (Prochnow, Tunmer, Chapman, & Greaney, 2001). The patterns for boys and girls warrant further study. But, given the variation between findings and the substantial overlapping distributions between samples of boys and girls, Prochnow, Tunmer, Chapman and Greaney's claim that the patterns appear to be more related to teaching practices than inherent predispositions of boys and girls, seems likely.

A major issue is the question of the cultural familiarity or appropriateness of the tests. Running through all these results is the potential for students' capabilities in comprehending to have been systematically under-assessed because of the nature of the texts and the tests. Luke, Woods, Land, Bahr and McFarland (2002) have raised this issue. The event knowledge required as background by texts, and the procedural and linguistic knowledge required by testing formats need careful analysis for bias, because background knowledge is such a strong determinant of comprehension (Pressley, 2001). For example, it might be assumed that 'Polynesian' children would perform better on the passages in the PATs that are based on Maori 'myths and legends'. This was not the case, the lowest two mean scores on PAT passages were the Maori legends. There are at least two reasons for this. One is that these are based on tribal Maori cultural knowledge, and the majority of the children, being urban Pasifika and Maori children may not have had wide access to or experience with these concepts or frameworks. Luke et. al. make the point that cultural groups tend to be homogenized by test developers, and there may be wide differences in practices and experiences between groups that relate to school tasks. A second is that the structure of retold legends might create a more difficult genre than a standard narrative or exposition in school texts (Graesser, McNamara & Louwerse, 2003).

The discussion above mirrors discussions that the cluster of schools and the research team are now having around these results. The report we have provided here illustrates features of the partnership between school professionals and researchers which is based on a shared problem solving approach using collected and analysed evidence. The achievement profiles and the classroom observations are contributing to a planned research and development programme. School-wide professional development across the cluster will be based on what the patterns have revealed or suggested. This means among other things, planning for and testing out the role of vocabulary in reading across the curriculum; testing the significance of children's awareness of their use of strategies, particularly checking strategies; and how to select and deploy literacy activities that build on the culturally and linguistically based expertise of the children. What we have learned is that it is critical not to make assumptions about what students can or cannot do, to look closely at instructional patterns as well as achievement patterns and to have a position that sees the students as full of promise; with cultural and linguistic resources that can be built on and from which high levels of comprehending can develop.

References

- Buly, M. R. & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. Educational Evaluation and Policy Analysis, 24, 3, 219-239.
- Dewitz, P. and Dewitz, P. K. (2003) They can read the words but they can't understand: Refining comprehension assessment. The Reading Teacher, 56, 5 422 – 435.
- Ministry of Education (2003). AsTTle. http://www.tki.org.nz/r/asttle/index_e.php
- Biemiller, A. (1999). Language and reading success. Cambridge, MA: Brookline Books.
- Dymock, S. and Nicholson, T. (1999). Reading comprehension: What is it? How do you teach it? Wellington: New Zealand Council for Educational Research.
- Elley, W. B. (2001). Supplementary Tests of Achievement in Reading. Wellington: New Zealand Council of Educational Research.
- Graesser, A. C. McNamara D.S. and Louwse, M M. (2003) What do readers need to learn in order to process coherence relations in narrative and expository text. A. P. Sweet, and C. E. Snow. (eds.). (2003). Rethinking reading comprehension. New York NY: Guilford Press.82-98
- Flockton, L. and Crooks, T. (2001). Reading and Speaking assessment results. National Education Monitoring Report 19. Wellington: NZ Ministry of Education.
- Hattie, J. (2002) A Research Agenda for New Zealand Schools. Vice Chancellor's Lecture Series. Unpublished manuscript. The University of Auckland, October 30, 2002
- Lee, C. D. (2000). Signifying in the zone of proximal development. In C. Lee and P. Smagorinsky, P. (eds.). Vygotskian perspectives on literacy research: constructing meaning through collaborative inquiry. Cambridge: Cambridge University Press. 191-225.
- Literacy Task Force (1999). Report of the Literacy Task Force. Wellington: New Zealand Ministry of Education.
- Literacy Experts Group (1999). Report of the Literacy Experts Group. Wellington, New Zealand Ministry of Education.
- Luke ,A. Woods, A., Land R, Bahr, M and McFarland, M. (2002). Accountability: Inclusive assessment, monitoring and reporting. Research Report prepared for the Indigenous Education Consultative Body: Brisbane. August 2002. School of Education, University of Queensland.

- McNaughton, S. (2002). Meeting of Minds. Wellington, New Zealand: Learning Media Limited.
- McNaughton, S., Phillips, G.E., & MacDonald, S. (2003). Profiling teaching and learning needs in beginning literacy instruction: The case of children in “low decile” schools in New Zealand. Journal of Literacy Research, 35(2), 703-730.
- Ministry of Education (2002). Curriculum Update. Issue 50, July.
- Ministry of Education. Annual Report 2003. Wellington: New Zealand Ministry of Education.
- Phillips, G., McNaughton, S., & MacDonald, S. (in press). Managing the Mismatch: Enhancing early literacy progress for children with diverse language and cultural identities in mainstream urban schools in New Zealand. Journal of Educational Psychology.
- Pressley, M. (2002). Comprehension strategies instruction: A turn of the century status report. In C. C. Block & M. Pressley (Eds.), Comprehension Instruction: Research-based best practice (pp11-27). New York: The Guilford Press
- Prochnow, J. E., Tunmer, W. E., Chapman, J. W. and Greaney, K. T. (2001). A longitudinal study of early literacy achievement asnd gender. New Zealand Journal of Educational Studies, 36, 2, 221-236.
- Reid, and Elley, W. B. (1991) Revised Progressive Achievement Tests: Reading Comprehension. Wellington: New Zealand Council for Educational Research.
- Snow, C. E., Burns, M. S. & Griffin, P. (1998). Preventing reading difficulties in young children Washington DC: National Academy Press.
- Stanovich, K. E., West, R. F., Cunningham, A. E., Cipelewski, J. and Siddiqui, S. (1996). The role of inadequate print exposure as a determinant of reading comprehension problems. In C. Cornoldi and J. Oakhill (Eds.). Reading Comprehension Difficulties: Processes and intervention (pp. 15 – 32) Mahwah, NJ: Lawrence Erlbaum. .
- Sweet, A.P. and Snow, C. E. (eds.). (2003). Rethinking reading comprehension. New York NY:Guilford Press.
- Tan, A. & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. Journal of Educational Psychology, 89, 276-288.
- Timperley, H. (2003). Shifting the Focus: Achievement information for professional learning. Wellington: NZ Ministry of Education.