

EARLY LITERACY RESEARCH PROJECT

Time 1 Assessments of Children's Literacy Skills

A Report of Findings for the
Ministry of Education

James W. Chapman, Alison W. Arrow, and
William E. Tunmer
Massey University
September, 2015



EARLY LITERACY RESEARCH PROJECT

**Report for the Ministry of Education of Findings from the Time 1 (February/March 2015)
Assessments of Children's Literacy Skills**

James W. Chapman, Alison, W. Arrow and William E. Tunmer

Massey University

Palmerston North

September, 2015

Summary

This report presents data on the selection of schools and students for the Early Literacy Research project, together with demographic information about the students and baseline data for Time 1 assessments conducted during February and March, 2015. Of particular importance is whether the random allocation of schools to either the intervention or comparison groups resulted in adequately equivalent student characteristics.

Overall, findings from the school selection and randomisation process, together with data from the Time 1 assessments indicate children in the intervention and comparison groups are generally similar in terms of age, gender and decile rank of their school. Moreover, there were no statistically significant differences between the two groups on any of the Time 1 assessment measures. This finding suggests that the children in both groups are essentially equivalent in terms of these important baseline literacy variables.

An examination of the distribution characteristics of scores for almost all measures revealed strong to very strong positively skewed distributions. On many variables, large numbers of children in both groups scored at "floor" levels. This finding is normal and expected for young children who have just entered school.

As anticipated, differences as a function of school decile band were apparent in these Time 1 data. The differences were especially notable for receptive vocabulary, letter knowledge, and phonological processing, with children in low decile schools (deciles 1-3) performing at lower levels than children in high decile schools (deciles 8-10). If the project is successful in one of its major goals, children in intervention low decile schools should outperform their counterparts in the comparison schools, and reach or at least approximate the literacy development of children in higher decile schools.

Introduction

The main goal of this literacy research project is to improve the literacy learning outcomes of New Entrant children during and beyond their first year of schooling. A particular focus is on children from lower decile schools, and children from Maori and Pasifika backgrounds.

These high priority children have been the target of numerous government strategies, and they were identified as an area of concern in terms of inequitable schooling outcomes in the briefing papers prepared by the Ministry of Education following the 2011 general election.

Our approach in this research project is to work with teachers of New Entrant children in schools randomly selected from the southern half of the North Island. Our work with teachers is focussed on supplementary teaching strategies that are designed to enhance regular classroom literacy instruction, with a particular emphasis on the development of language-related and word-level decoding skills that are essential for successful literacy acquisition.

The goal of the research is to improve literacy learning outcomes for all children, especially those from backgrounds which are not adequately catered for in New Zealand's current approach to literacy instruction. To achieve this goal, we have developed and are implementing a professional learning and development (PLD) programme for teachers.

The content of the PLD comprises supplementary teaching strategies for enhancing literacy learning outcomes, along with the use of additional assessments that effectively identify specific literacy learning needs that should be addressed during children's first year of school. A further key aspect of the PLD programme is to encourage teachers to adopt *differentiated instruction* for New Entrant children.

Differentiated instruction takes into account children's development of skills on the literacy development continuum. Teachers will learn to provide targeted language-related literacy instruction (e.g., phonemic and phonological awareness) for some children whose assessment data reveal particular needs, whereas for other children who already have basic reading-related language skills teachers will be encouraged to provide instruction that recognises their different developmental progress.

An important part of the research is the randomised control design. This design provides the opportunity to evaluate the effectiveness of the teacher PLD programme for children in the "Intervention" schools compared to children in the "Comparison" schools, whose teachers are not presently receiving the PLD programme. A key part of the randomised control design is to ensure that the random allocation of schools to either the intervention or comparison "condition" has been effective in terms of the equivalence of children's baseline assessments.

This report presents results from these baseline assessments. If the sample characteristics of both groups are reasonably similar in terms of ages, gender, decile ranking of schools, and ethnic background, and if children in both groups perform similarly on key measures at school entry, we can assume that the randomisation process has been effective. Moreover, if children in the intervention group outperform children in the comparison group over the period of the project, we can reasonably assume that the teacher PLD variable is the key factor contributing to subsequent differences in literacy performance outcomes.

We present information on the sample characteristics, the "baseline" assessments that were administered at the start of the project, and the results of these initial assessments.

Sample

School Sampling Procedure

A random sample of schools was selected from regions of the lower North Island that include Wellington, Hutt Valley, Wairarapa, Kapiti, Horowhenua, Manawatu, Whanganui, Ruapehu, Tararua and Taranaki. A stratified frame was used in an attempt to maximise participation of lower decile schools in the project. The initial selection process was drawn from state and integrated primary schools listed on a Ministry of Education database. Included in the draw were 80 schools that were expected to enrol eight or more new Entrant students at the start of 2015.

Schools from the Rangitikei district were excluded because of their small size and the small number of New Entrant children expected to be enrolled at the start of the 2015 school year. Schools in the Ruapehu district were also excluded because of their predominantly small enrolments and also because of the practicalities and costs associated with data collection from what would likely have been only a very few schools and New Entrant children. In addition, schools in the Porirua area and parts of Wellington and Hutt Valley that were participating in the *Shine Literacy Success for All* project were excluded from the sample to avoid confusion and "contamination" of the two projects.

Schools were randomly selected and randomly allocated to either the Intervention or Comparison conditions. This procedure was performed by means of a random number generator in the SPSS statistical package.

Following the sampling process, principals of the selected schools were contacted and invited to participate in the project and to attend meetings in Wellington, Palmerston North, and Whanganui to discuss the goals and activities of the project. Explanations included which of the two groups, intervention or comparison, to which the school had been randomly assigned.

Principals of the 80 schools drawn randomly and anonymously from a Ministry of Education database were contacted towards the end 2014 seeking agreement to participate in the project. By the start of the 2015 school year, 38 schools confirmed their willingness to take part in the research.

Decile rankings of schools showed some differences between the intervention and comparison schools. We grouped decile rankings as follows: low = deciles 1 to 3; medium = deciles 4-7; high = deciles 8-10. The spread across these three decile groups was more even for the intervention schools than the comparison schools. These data are presented in Table 1.

Table 1. Distribution of project participants by group and school decile band.

Decile Band	Group	
	Intervention	Comparison
	Percent (n)	Percent (n)
1-3	36.6 (72)	56.6 (90)
4-7	27.7 (56)	5.0 (8)
8-10	36.6 (74)	38.4 (61)

Sample Characteristics

Of the 38 schools that agreed to participate in the project, 24 had been randomly assigned to the intervention group, and 14 to the comparison group. A total of 62 teachers of New Entrant children were identified: 38 in the intervention schools and 24 in the comparison schools. These numbers fluctuate as teachers come and go for various personal or professional reasons. At the first of the scheduled intervention group teacher professional development workshops, 45 teachers attended from the 24 intervention schools.

Time 1 baseline assessment data were collected during February and early March 2015 from 361 New Entrant/Year 1 children. Of these, 202 (54%) were in intervention schools, and 159 (46%) were in comparison schools.

Age data reveal that the mean age of the sample at the time of first assessment was 60.58 months (SD = 1.85), which is 5 years, 1 month; the median age was 60 months, and the mode was 60 months. Clearly, the large majority of children are around 5 years of age. There is no statistically significant difference between the mean ages for the Intervention and Comparison groups: 60.71 (SD = 1.86) and 60.41 (SD = 1.84) respectively.

In terms of gender, 53% of the children were boys, and 47% were girls. However, there was a marginal imbalance for children in the intervention group: 54% (109) were boys compared to 46% (93) girls. The gender breakdown in the comparison group was generally more even: 51% (81) boys and 49% (78) girls.

Regarding ethnic background of the participating students, these were incomplete at the time of preparing this report. We anticipate providing these data in the report on Time 3 assessment data.

Assessments

The following assessments were undertaken during February and March, 2015.

Letter Identification

Research indicates that letter name knowledge and letter sound knowledge are important aspects of initial literacy acquisition. We have kept these two aspects of letter identification separate, because they differentially predict pathways to new word learning (Arrow, 2012).

Letter name and letter sound knowledge were assessed in terms of both upper case and lower case letters, using the Letter Identification task in the Diagnostic Survey (Clay, 1985). Children were asked to name each letter and to say the sound the letter represented for 26 upper case and 28 lowercase letters, two of which appeared in varying fonts. Scoring was based on the number of letters correctly identified by name, and by sound.

Vocabulary Knowledge

We used the British Picture Vocabulary Scale (BPVS: Dunn et al., 2009) to assess receptive language ability. This ability refers to understanding the meaning of words, which is necessary for the production of functional language. Raw scores are converted to standard scores, which are related to the age of each participant.

Word Recognition

Word recognition refers to the fluent, rapid reading of words as they appear. Such words are usually known as *sight words*. These words include the high-frequency words that children learn to read first, as well as any word that is read with automaticity. We used one of the *Ready to Read* test lists (Clay, 2002). These tests comprise 45 words of the most frequently occurring words in the 12 "little" books of the *Ready to Read* series. We administered the first 15 words in one of the lists. Scoring was based on the number of words read correctly by each child. In addition, attempts at word reading accuracy were assessed by scoring the number of correct letter-to-sound correspondences in each word.

Invented Spelling

Invented spelling was assessed by having children write 18 words that were read aloud by the research assistant. The 26 (lower case) letters of the alphabet were displayed across the top of the children's response sheets. Each word that children wrote down received a score from 0 to 4. Maximum points were awarded if the sounds in the word were represented with letters, although unconventionally (e.g., *kik* for *kick*, *fil* for *fill*, *sid* for *side*). Two points were awarded if more than one phoneme (but not all) was represented with phonetically related or conventional letters (e.g., *sd* for *side*, *lup* for *lump*). One point was awarded where the initial phoneme was represented with the correct letter (e.g., *f* for *fat*). Children were also asked to identify the sounds in the words that were read aloud. The total number of possible points for letters and sounds was 72 each.

Phonological Processing

Phonological processing was assessed using the Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2: Wagner, Torgesen, Rashotte, & Pearson, 2013). This test is normed in the United States for use with people from 4 years to 25 years. The CTOPP-2 is used to help evaluate phonological processing abilities as a prerequisite to reading fluency. We administered three of the subtests: elision, blending and matching. Elision measures the ability to remove phonological segments from spoken words to form other words. There are 34 items in this test, with discontinuation occurring when each child missed three consecutive items. Blending Words measures the ability to synthesize sounds to form words. There were 33 items in this section; again, discontinuation occurred following three consecutive missed items. Sound Matching measures the ability to select words with the same initial and final sounds. This section comprised 26 items; testing was discontinued following three missed items.

Results

Score Distributions

We examined the distribution of scores on the Time 1 variables for the total sample. Not surprisingly, most variables had positively skewed distributions, with nine of the 12 variables showing modal scores of 0: the four Letter ID measures, CTOPP elision, Clay word test and phonemes, and the two measures of invented spelling. Scores for CTOPP blending and matching showed a greater range in the distribution of scores, however they were still positively skewed. For the BPVS, the distribution of scores was relatively normal, with a mean of 98.75 (SD = 11.88), a median of 101, and a mode of 109. The skewness was -.29 and the kurtosis was -.78.

Group Comparisons

We performed simple *t*-tests to test the hypothesis that there would be no significant differences in mean scores between the intervention and comparison groups on the Time 1 assessment variables. The hypothesis was supported for each of the variables. Means and standard deviations are presented in Table 2.

Table 2. Means and standard deviations for intervention and comparison students at Time 1.

Assessments	Intervention Group			Comparison Group		
	Mean	SD	n	Mean	SD	n
BPVS	98.10	13.64	198	98.97	12.10	148
Letter ID UC Name	11.09	8.90	201	11.73	9.07	158
Letter ID LC Name	9.78	8.70	201	10.59	8.79	158
Letter ID UC Sound	5.20	7.34	201	6.76	7.79	158
Letter ID LC Sound	5.07	7.46	201	6.25	7.74	158
Clay Word Test	0.52	1.56	201	0.56	1.86	158
Clay Word Phonemes	1.50	4.33	201	1.97	5.53	158
Invented Spelling	0.11	0.62	201	0.24	1.41	158
Invented Spelling Sounds	3.82	9.84	201	5.37	11.02	158
CTOPP Elision	4.31	4.26	201	4.54	4.27	158
CTOPP Blending	6.91	4.19	201	6.29	4.24	158
CTOPP Matching	7.48	5.80	201	7.00	5.58	158

This finding is important and indicates that receptive vocabulary, alphabet knowledge, word knowledge, phonological processing, and phonemic awareness scores are similar for the two groups. Having similar baseline means for the Time 1 variables suggests that the randomisation process has worked well insofar as the intervention and comparison students were performing at similar levels at the start of the project. This similarity will strengthen the ability to make inferences about the effectiveness of the intervention if, as

we anticipate, students in the intervention group progress at an accelerated rate on subsequent assessments.

Decile Band Comparisons

We examined scores for the 12 variables in terms of decile bands for the total sample. These comparisons were conducted by means of one-way analyses of variance (ANOVAs). Seven variables showed statistically significant effects, all with probabilities less than .01: the four Letter ID measures, CTOPP elision and blending, and the BPVS. The other measures did not result in statistically significant effects, with all probabilities greater than .17.

The means and standard deviations for the 12 variables are presented in Table 3. These data show that mean scores for children in the low decile band were generally lower than for children in the high decile band, and in one case (BPVS) lower than those in the middle decile band. No differences between with middle and high decile groups were statistically significant.

We also examined the distributions of scores in terms of school decile band for all variables except the BPVS. This examination was undertaken because of the large numbers of students who scored 0 on many of the measures. We calculated the percentages of students scoring 0 or 1 on each of the 11 variables. These percentages are presented in Table 4.

Table 3. Means and standard deviations for Time 1 variables as a function of school decile band.

Variables	Low Decile Band		Middle Decile Band		High Decile Band	
	Mean	SD	Mean	SD	Mean	SD
BPVS*	94.43 ^{ab}	11.28	101.06	11.68	102.63	11.07
Letter UC Name*	8.99 ^a	8.64	11.84	8.58	14.01	8.83
Letter LC Name*	7.69 ^a	8.05	10.67	8.37	12.82	8.93
Letter UC Sound*	4.47 ^a	7.04	5.34	7.15	7.84	8.00
Letter LC Sound*	4.05 ^a	6.71	5.52	7.57	7.48	8.22
CTOPP elision*	3.43 ^a	4.13	4.59	4.17	5.51	4.20
CTOPP blending	6.35	3.86	6.23	4.50	7.17	4.45
CTOPP matching*	6.14 ^a	5.48	8.06	5.84	8.24	5.68
Clay word	0.46	1.64	0.31	0.94	0.75	2.00

Clay phonemes	1.54	4.97	0.97	2.56	2.27	5.56
Invented spelling	0.18	1.33	0.09	0.75	0.19	0.73
Invented spelling sounds	3.81	9.71	3.63	8.53	5.74	11.84

* Statistically Significant 1-way ANOVA, $p < .01$

^a Low decile group significantly lower than high decile group

^b Low decile group significantly lower than middle decile group

BPVS = British Picture Vocabulary Scale

UC = upper case; LC = lower case

CTOPP = Comprehensive Test of Phonological Processing

Table 4. Percentages of students obtaining scores of 0 or 1 on each of the listed variables, as a function of school decile band.

Variables	Low	Middle	High
Letter UC Name	21.1	7.8	9.7
Letter LC Name	29.8	20.3	12.7
Letter UC Sound	59.0	56.3	34.3
Letter LC Sound	65.8	57.8	34.3
CTOPP elision	47.8	32.8	25.4
CTOPP blending	15.0	18.7	16.5
CTOPP matching	22.4	14.1	9.7
Clay word	91.3	92.2	85.8
Clay word phonemes	85.1	82.8	73.1
Invented spelling	96.3	98.4	94.8
Invented spelling sounds	73.9	75.0	68.6

The most notable decile band differences between low and high decile students are for receptive vocabulary (BPVS), letter knowledge (name and sound; upper and lower case), and phonological processing (elision and blending). These differences are consistent with our view that students in low decile schools tend to have lower levels of literate cultural capital which impact on some of the key language-related factors associated with literacy acquisition (Tunmer, Chapman, Greaney, Prochnow & Arrow, 2013). We anticipate that the effects of the intervention programme will overcome any disadvantage associated with these initial differences.

Discussion

Overall, findings from the school selection and randomisation process, together with data from the Time 1 assessments, indicate children in the intervention and comparison groups are generally similar in terms of age, gender and decile rank of their school. Slightly more children are in the intervention group (54%) compared to the comparison group (45%), however, this difference is of little consequence in terms of statistical processes for treating the assessment data.

Of particular significance is the finding that the standardised scores for the British Picture Vocabulary Scale (BPVS) are very similar for both groups of children. This measure of receptive vocabulary is an important indicator of verbal ability, and as such, shows that this key variable in literacy development is consistent across both groups.

An examination of the distribution characteristics of scores for almost all measures (the exception is the PBVS) revealed strong to very strong positively skewed spreads of scores. On many variables, large numbers of children in both groups scored at "floor" levels. This finding is normal and expected for young children who have just entered school.

Finally, as expected, differences as a function of school decile band were apparent in these Time 1 data. The differences were especially notable for receptive vocabulary, letter knowledge, and phonological processing. If the project is successful in one of its major goals, children in intervention low decile schools should outperform their counterparts in the comparison schools, and reach or at least approximate the development of children in higher decile schools.

References

- Arrow, A.W. (2012, July). *Alphabet knowledge in the acquisition of early orthographic representations*. Paper presented at the annual meeting of the Society for the Scientific Studies in Reading, Montreal, Canada.
- Clay, M.M. (1985). *The early detection of reading difficulties*. Auckland: Heinemann.
- Clay, M.M. (2002). *An observation survey of early literacy achievement* (2nd. Ed.). Auckland, Heinemann.
- Dunn, L.M., Dunn, D.M., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., et al. (2009). *British picture vocabulary test* (3rd. ed.). London: GL Assessment.
- Tunmer, W.E., Chapman, J.W., Greaney, K.T., Prochnow, J.E., & Arrow, A.W. (2013). Why the New Zealand literacy strategy has failed and what can be done about it: Evidence from the Progress in International Reading Literacy Study (PIRLS) 2011 and Reading Recovery monitoring reports. *Australian Journal of Learning Difficulties*, 18, 139-180.
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., & Pearson, N.A. (2013). *Comprehensive test of phonological processes* (2nd. Ed.). Austin, TX: Pro-Ed.