Efficacy of an evidence-based literacy preparation program for young children beginning school

Robyn Wheldall[1], Katharine Glenn[1], Sarah Arakelian[1], Alison Madelaine[1], Meree Reynolds[1], Kevin Wheldall[1]

[1] *MultiLit Research Unit, MultiLit Pty Ltd, Macquarie Park, Australia.*

Disclosure statement: The authors hereby declare a financial interest in the outcomes of this study. All authors work for MultiLit Pty Ltd, which publishes the PreLit program. This was made clear in the Ethics application for the project and in the consent forms signed by school principals and parents.

The following is the post-print version of an article that has been peer reviewed and accepted for publication. It was published by Taylor and Francis on 8 June 2016. Full reference details are below.

Wheldall, R., Glenn, K., Arakelian, S., Madelaine, A., Reynolds, M., & Wheldall, K. (2016). Efficacy of an evidence-based literacy preparation program for young children beginning school. *Australian Journal of Learning Difficulties, 21*(1), 21-39. https://doi.org/10.1080/19404158.2016.1189443

Abstract

This study aimed to provide evidence regarding the efficacy of an early literacy preparation program, *PreLit*, designed to improve the skills of young Australian children. Participants comprised 240 children in eight schools attending their first year of schooling. Children in the four experimental group schools received instruction in the program while children in the four comparison group schools continued with typical literacy activities in their classrooms. All children were assessed on measures of emergent literacy and language skills prior to and following intervention. It was found that 91% of students were in the bottom quartile for phonological awareness at pre-test. While neither the children nor their schools were randomly allocated to groups, the mean scores for the two groups were very similar at pre-test on all measures. Analyses showed that although the means for the two groups were not statistically different on any of the measures at post-test, significantly fewer students in experimental schools remained in the bottom quartile and more moved into the top quartile for phonological awareness skills, compared with students in the comparison schools. Fine grain analyses, taking into account additional qualitative data about the schools, helped to clarify these findings.

Efficacy of an evidence-based literacy preparation program for young children beginning school

Based on international standards, it is clear that too many Australian students are falling behind in reading. Recent international assessments have indicated significant differences between the reading skills of Australian students and those from other English-speaking countries (Mullis, Martin, Foy, & Drucker, 2012). Poor outcomes for Australian students indicate that the current educational system may not be adequately catering for the needs of a sizeable minority of students when first learning to read. The primary school years are when students should have mastered the skills of reading and should now be "reading to learn". If students are unable to make a timely transition from "learning to read" to "reading to learn" they will experience poor academic outcomes and may also develop negative social issues associated with failure. Early intervention is an important form of support that may prevent or reduce these difficulties (Connor, Morrison, & Slominski, 2006).

Over the past few decades, Australian governments have invested heavily in education with a particular focus on teaching reading. However, resources have not necessarily been directed towards effective methods of reading instruction (Buckingham, Wheldall, & Beaman-Wheldall, 2013). Furthermore, there is evidence to support the need for effective instruction to be provided early in a child's life (Campbell & Ramey, 1994; Hart & Risley, 1995). Hempenstall (2016) reports that 30% of first-graders do not appreciate the phonemic structure of words, with the proportion being even higher among disadvantaged children. Given the critical importance of understanding the sound structure of the English language in the process of learning to read, this is a serious cause for concern.

Early intervention in literacy and language is not a new concept and has been the focus of a large body of research (see discussion in Hart & Risley, 1995). Hart and Risley concluded that intervention may be required from as early as birth, noting that by school entry, the already apparent differences between the language experiences of children from families of different socio-economic status are difficult to resolve. However, they noted that such interventions would be costly, difficult to manage and difficult to resource.

Interventions implemented at the preschool level have provided children with initial strong benefits in the early school years. While Catts, Fey, Tomblin, and Zhang (2002) noted that children

who are successful in acquiring early reading skills are generally likely to maintain success, Hart and Risley (1995) learned from an early study on preschool intervention, that support may be required throughout the schooling years to maintain benefits for children from lower SES families.

It has been established that the emergent literacy skills children have acquired by the time they start school predict their ability to learn to read effectively. In a study in the United States, Chatterji (2006) estimated that 38% of the variability of children's reading ability by the end of their second year of schooling was accounted for by their early reading skills upon school entry. Many of these emergent literacy skills have been evaluated as predictors of the reading skills that comprise the Simple View of Reading (SVR), a framework now widely accepted. The SVR holds that reading comprehension, which is the purpose of reading, is the product of decoding ability and linguistic comprehension (Hoover & Gough, 1990).

In a series of meta-analyses involving more than 299 studies of children aged zero to five years, the National Early Literacy Panel in the United States (NELP, 2008) investigated which early literacy skills were predictive of the later conventional reading skills, decoding and reading comprehension. Alphabetic knowledge, phonological awareness (PA), rapid automatic naming and name writing ability were all found to be moderate to strong predictors of later decoding ability and reading comprehension, the most important of these being PA (Callaghan & Madelaine, 2012).

Evidence for the role of PA in later reading skills is consistently strong. PA is a metalinguistic construct referring to awareness that language is made up of a combination of sounds (Chapman, 2003). In a concurrent review on the language predictors of children's risk of reading failure, Serry, Rose, and Liamputtong (2008) concurred that phonological processing, a broader concept comprising PA, phonological memory and rapid automatic naming, had documented links to reading ability. The conclusions of the review supported the findings of NELP (2008) that PA and rapid automatic naming predict later reading ability (Serry et al., 2008) and are reliably related to reading ability from the first year of school (Kindergarten in New South Wales) to Grade 2 (NELP, 2008).

Other early literacy skills, including concepts about print, print knowledge, reading readiness, oral language skills and visual processing, have been found to be less reliably related (NELP, 2008).

Initial correlations of oral language with decoding and reading comprehension were found to be moderate, but oral language skills were among the weaker predictors of both reading skills. However, longitudinal investigation into the impact of oral language deficits on later reading ability have demonstrated a connection between the two (Catts et al., 2002).

Catts et al. (2002) conducted a study comparing children with language impairments to children with normally developing language, over their first four years of schooling. Broadly categorised, children who presented with language deficits in Years 2 and 4 had significantly poorer performance for various reading skills than their non-impaired peers, such that roughly half of this group qualified as having a reading disability. Of the oral language skills measured in Kindergarten and Year 2, however, only grammar was a significant predictor of Year 2 and Year 4 reading. Neither receptive nor expressive vocabulary were significant predictors of Year 2 or 4 reading once previous reading ability was controlled, in this study. As this indicates that some aspects of oral language may be more strongly related to later reading ability, a global measure of the skills may mask any significant relationships. Some skills are necessary but not sufficient for effectively learning to read. It is important for early education providers to consider this in planning for effective instruction.

While the predictive ability of emergent literacy skills for later reading has been well documented, the research is often correlational and therefore cannot demonstrate a causal relationship. Castles and Coltheart (2004) have cautioned about the difficulties involved in untangling the variables in terms of causality when it comes to early literacy skills including phonemic awareness and the subsequent acquisition of reading. Moreover, it is impossible to control the presence of deficits of phonological processing and oral language skills to determine a causal relationship with later reading ability. However, conducting experimental research into the efficacy of programs addressing gaps in such emergent literacy skills can go some way to providing evidence of such a relationship.

Research has shown that children commencing school with both PA and well developed general language skills are far more likely to learn to read easily and quickly (reviewed Callaghan & Madelaine, 2012). Although various reports have identified these predictors of reading ability, it does not necessarily follow that interventions involving instruction in these skills will result in change in

later reading ability. However, it is important to consider which elements are required for an intervention in emergent literacy skills to be effective.

Callaghan and Madelaine (2012) emphasised the need for pre-school literacy preparation to be systematic, play-based and delivered via short but frequent periods of instruction. Furthermore, the authors reviewed the emergent literacy skills that should be targeted and how they should best be taught. They concluded that effective interventions for emergent literacy should involve instruction in skills related to PA, some instruction in letter names and letter-sound correspondences and instruction in phoneme blending and segmenting skills. PA skills important for instruction in early literacy skills include the ability to detect phonemes (Chapman, 2003), and tasks involving the manipulation of sounds such as phoneme identification and deletion. However, Callaghan and Madelaine noted that these skills alone have not been shown to be sufficient to protect children from experiencing later reading problems. They argued that some teaching of letter names and letter-sound correspondences should also be provided, enabling children to bridge the gap between PA skills and later decoding skills. As Hempenstall (2016) points out, the reciprocal nature of the relationship between PA and the alphabetic principle is key. Sensitivity to phonology aids the understanding of the alphabetic principle with this operating as a "virtuous cycle", resulting in the acquisition of spelling and reading skills which in turn further enhance phonemic awareness (Hempenstall, 2016).

Turning to the development of oral language skills, storybook reading has been traditionally used as a method for providing a rich language environment for young children (Whitehurst et al., 1988). Callaghan and Madelaine (2012) noted that while most preschools employ traditional shared book reading in their programs, this type of instruction was insufficient in improving the oral language skills of children and recommended instead dialogic reading. A technique first used by Whitehurst et al. (1988), dialogic reading involves multiple readings (Doyle & Bramwell, 2006) and encourages children to actively participate by answering questions about the story (NELP, 2008). Teachers provide feedback by repeating, expanding on, and modelling answers that helps to foster richer language environments for children (Whitehurst et al., 1998).

Current study

This study constituted an attempt to provide empirical evidence for the efficacy of a whole class literacy preparation program, originally designed for preschool children in the year before starting formal schooling. This program, called *PreLit* (MultiLit, 2012a), described below under Method and Appendix 2, is appropriate for all children in the year before formal schooling but was developed particularly for children who do not have the necessary emergent literacy skills to learn to read successfully, particularly those from disadvantaged backgrounds.

While *PreLit* was primarily designed as a pre-school intervention, it was envisaged that the program might also prove to be beneficial for some young children commencing school who might be at risk of difficulties in learning to read. The current study sought to test this by comparing the progress made by students who undertook the 25 week program (the program would typically take up to one year in preschool settings) in their first three terms of Kindergarten with a comparison group of similar students who did not.

Data from pre- and post-tests of experimental and comparison groups on measures of language and early literacy skills were examined to determine whether there were differences in the children's gains over the time period. It was hypothesised that the group who received the *PreLit* program intervention would show greater growth in measures of emergent literacy skills than the comparison group.

While the original aim was to complete an experimental trial, this did not prove to be possible and neither children nor schools were randomly assigned to conditions/groups. It is recognised that this limits the conclusions that may be drawn from the data collected and that, hence, only tentative suggestions regarding potential efficacy may be made.

Method

Setting

Participants in the study were recruited from eight Catholic primary schools in a regional area of New South Wales, Australia. (The study was completed in close co-operation with the regional Catholic Schools Office.) Details of the eight schools involved in the study are provided as Appendix 1. It had originally been supposed that the children from the eight schools would largely be from disadvantaged backgrounds but the ICSEA values (Index of Community Socio-Educational

Advantage, Australian Curriculum, Assessment & Reporting Authority [ACARA], 2015) of the schools did not support this assumption. ICSEA values across Australian schools are scaled to have a median value of 1000 and a standard deviation of 100 (ACARA, 2015). All schools in this study had ICSEA values less than one standard deviation either side of the median. While there is some variability across schools, all schools may be said to fall into the average range in terms of socio-economic status.

In spite of this, however, 91% of children fell into the bottom quartile for the measure of phonological awareness (SPAT-R) employed at pre-test. In terms of the Peabody Picture Vocabulary Test (PPVT) standardised score, however, commonly regarded as a proxy for overall verbal ability, this group of students had a mean of 106.37 (SD 12.57), indicating at least average ability. It is, therefore, difficult to assert with any confidence that these students were in any relative sense socially disadvantaged but they were clearly coming to school with less than adequate phonological skills. To this extent, they may be described as ill-prepared for learning to read.

Participants

A total of 264 children were originally pre-tested for participation in the study across the eight schools (131 boys and 133 girls). All students were commencing school for the first time in the same Kindergarten entry in 2015. Children were tested on a short battery of language and early literacy measures (see below) both prior to and following the experimental group receiving 25 weeks of *PreLit* intervention for half an hour per day. The average age of these children at pre-test was 65 months (range: 54–73 months). At post-test 240 of the original 264 were available for re-testing. There were 115 children in the experimental group and 125 in the comparison group. At pre-test the means for the two groups were very similar on all three measures and for age (See Table 1).

Measures

All participants were given a battery of tests measuring phonological awareness and oral language skills prior to and following the *PreLit* intervention. As children were in their first year of schooling and expected to have commenced formal reading instruction, an additional measure of early literacy skills was administered at post-test only. Tests were administered by trained research assistants and were independently double-scored. The children were tested "blind", i.e. the research

assistants did not know which children were in the experimental group and which children were in the comparison group.

Sentence Comprehension Test (SCT) – Revised Edition **(Wheldall, Mittler, & Hobsbaum, 1987).** The SCT measures a child's receptive language by assessing their ability to comprehend sentences of varying length and grammatical complexity. Children are presented with a spoken sentence and choose from four pictures which most closely represents the information contained in the sentence. The test has ten subtests of four items each, organised according to the different forms of language and grammatical structures that they present. The subtests measure the following sentence structures: simple intransitive, simple transitive, intransitive with adjective, plurals, past, future, negative, prepositions, embedded phrases and passive. The test manual presented evidence for adequate test-retest reliability (r = .79) and split half reliability (r = .77–.83). Adequate criterion related validity was established with the British Picture Vocabulary Scale (Dunn, Dunn, & Whetton, 1982; r = .78).

Sutherland Phonological Awareness Test – Revised (SPAT-R**; Neilson, 2003b).** The SPAT-R is designed to assess the phonological awareness skills required for literacy development of 5 to 8 year olds in the first few years of school. The test comprises 13 subtests that assess the following skills: syllabification, rhyming, phoneme identification, phoneme segmentation, sound blending, deletion of consonants, non-word reading and non-word spelling. The test is discontinued after the fifth subtest, measuring onset phoneme identification, if the child cannot segment or identify initial sounds. The test author reported high internal consistency (r = .95) and strong correlations with measures of word identification (the Word Identification subtest of the Woodcock Reading Mastery Test-Revised; Woodcock, 1987; r = .78) and another measure of phonological awareness (the Astronaut Invented Spelling Test; Neilson, 2003a; r = .86).

Peabody Picture Vocabulary Test – Fourth Edition (PPVT-IV; **Dunn & Dunn, 2007).** This test is a measure of receptive vocabulary and is commonly employed as a proxy measure for overall verbal ability. The test comprises pages of four pictures each. The test administrator provides a verbal stimulus word and students are required to respond by selecting the picture that they think best represents the spoken word. The test is suitable for people aged 2 to 90 years old. The test manual

provides evidence of high internal consistency (r = .89–.98) across age groups, grades and forms of the test, adequate alternate forms reliability (adjusted for range restriction, r = .87–.93) across age groups, and high test-retest reliability (adjusted for range restriction, r = .92–.96) across age groups. Correlation coefficients of the associations between the test and the Expressive Vocabulary Test (Williams, 2007; r = .80–.84), the Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999; r = .37–.77), and the Clinical Evaluation of Language Fundamentals (Semel, Wiig, & Secord, 2003; r = .67–.79) provide evidence of good criterion related validity of the test. Note that this is the only test to offer standardised scores.

***Letter-Sound Test (LeST;* Larsen, Kohnen, Nickels, & McArthur, 2015).** The LeST is a measure of grapheme-phoneme correspondences and includes a comprehensive list of single and multiple letters to assess these. Suitable for children in Kindergarten up to Grade 3, children are asked to provide the sounds that each letter or letter combination represent. The test authors reported high test-retest reliability (.88) and adequate criterion related validity with measures of phonological decoding ability using nonwords (r = .49–.70). This test was administered only at post-test.

Research design and procedure

The study design was originally conceptualised as (at least) quasi-experimental in nature. Schools were selected for participation via a regional Catholic Schools Office in New South Wales. Schools were initially excluded if they had composite Kindergarten classes or if they were geographically impractical. Schools were then paired on ICSEA value (ACARA, 2015), NAPLAN Year 3 Reading results (ACARA, 2015) and class size. The original idea was that schools from each pair would then be randomly allocated to experimental and control groups but this posed difficulties for the regional Schools Office, for administrative and pragmatic reasons, and who required that certain schools be placed in the experimental group. Therefore, true random allocation of matched schools to groups was *not* possible. For this reason, the students in the non-experimental group of schools are referred to as the comparison group. In spite of this, as may be seen from Table 1, the two groups were remarkably similar on all pre-test measures. It is acknowledged that the lack of randomization compromises the conclusions that may be drawn from this study.

Active consent was obtained from principals prior to pre-testing. Once school consent was obtained, passive consent was obtained from parents of children attending participating schools. Prior to the pre-testing, parents of nominated children were provided with information sheets and consent forms stating that consent for participation would be indicated if they did not return the consent form. (Passive consent had previously been agreed by the University Research Ethics Committee.)

Children attending the schools allocated to the experimental intervention (*PreLit*) received lessons in the *PreLit* program (MultiLit, 2012a) while those in the comparison group continued with their usual daily lessons. Instruction for schools in the experimental group was delivered by kindergarten teachers at the schools who had received thorough training in the intervention program. By providing training to existing teachers, it was intended that the schools could continue to provide the program to children following the completion of the study. Kindergarten teachers at schools assigned to the comparison group received training in the intervention program upon completion of the study, following post-testing of all participating children. Thus, children attending comparison group schools were provided with the opportunity to receive instruction in the program at a later date.

Two and a half terms (approximately 25 weeks) of instruction was delivered to allow teachers time to administer the entire program. This decision was based on previous field trials of the program. Formal observations of lessons were conducted with each teacher during program delivery in order to measure how reliably it was being delivered. Data collected at pre- and post-testing were analysed to compare the growth in scores of the experimental and comparison groups on the measures used.

Intervention

The *PreLit* Early Literacy Preparation Program (MultiLit, 2012a) is a play-based, systematic and explicit program of instruction that provides young children with the pre-requisite skills needed for learning to read. It was developed for whole class, small group or one-to-one instruction, primarily for preschool children in the year prior to Kindergarten. The program was intended for delivery to children throughout the year, at least three times per week. As the program was being trialled in Kindergarten classes at schools in this study, the program was delivered daily and therefore the content was delivered in a shorter period of time. The program also adheres to the principles of the

Early Years Learning Framework (Department of Education, Employment & Workplace Relations [DEEWR], 2009).

Each 30-min session comprises two main components; PA instruction and storybook reading to develop oral language skills. Each PA lesson is paired with a shared storybook reading session according to the hierarchical progression of skills. Each component was delivered to children in the experimental groups for 15 min, five days a week for 25 weeks. Classes at schools in the experimental group were split up into groups during PA lessons to accommodate a limit of 15 students per teacher. Typically, classes had more than one group to allow for this limit. Teachers conducted PA lessons with each group consecutively after which the groups would combine for the Storybook Reading session. Classes at one school remained in separate groups for the Storybook Reading sessions as the children were usually split into groups during the time in which the intervention was being delivered. Further details of the *PreLit* Program are provided in Appendix 2.

Analysis

Analyses of covariance were conducted to compare gains made by the experimental and comparison groups. Analyses compared post-test scores of the groups on each measure covarying pre-test scores, except for the LeST which was administered at post-test only. (A proxy pre-test measure, SPAT-R, was used in the case of LeST.) An alpha level of $p < .01$ was set to control for family-wise comparisons, as an alternative to Bonferroni corrections. Both partial eta squared and Cohen's *d* were used as measures of effect size.

Treatment integrity

Treatment integrity checks were conducted on two occasions throughout the study, one at the end of the first school term (after about 7 weeks of *PreLit* instruction) and again towards the end of the second school term (after about 17 weeks of *PreLit* instruction). One observer conducted the treatment integrity checks and these were conducted over the period of a week on each occasion. Following the observations, feedback was provided to each teacher individually. There were 13 teachers in 11 classes/groups in four schools involved in delivering *PreLit* instruction over the course of this study. All but two of the teachers had attended the *PreLit* training conducted at the beginning of the study.

The treatment integrity checklist consisted of two components, one for the PA lesson (with 15 dimensions) and one for the Storybook Reading component (with 11 dimensions). Teachers were observed and were given a score for each dimension on a three-point scale, signifying Always (1), Sometimes (.5) and Never (0). An example of the treatment integrity dimensions for the PA component was "Pacing – appropriate pacing of lesson to ensure students remain engaged and to complete lesson within allocated time". An example for the Storybook Reading component was "Provides opportunity for oral language development by expanding on limited student responses, modelling more detailed responses".

On the first occasion, 10 teachers were observed for the PA lesson and achieved a mean score (expressed as a percentage) of 94% (range 87–97%). On the second occasion 11 teachers were observed (3 of whom were not observed on the first occasion) and achieved a mean score of 97% (range 93–100%). On this second occasion, four of the 11 teachers scored 100%. Clearly, there was high treatment fidelity in this study for the PA component.

In terms of the Storybook Reading observations, 10 teachers were observed on the first occasion and achieved a mean score (expressed as a percentage) of 85% (range 73–95%). There were three teachers who scored below 80% at this first observation point. On the second occasion, 11 teachers were observed (two of whom had not been observed on the first occasion), achieving a mean score of 90% (range 82–100%). It was clear that by the second observation there was a high degree of treatment integrity, given that all the teachers observed scored over 80%. Given the nature of the (unscripted) task in the Storybook Reading component, it is not surprising that there was slightly lower treatment fidelity initially and that it took time to adhere more closely to the program. Overall, it can be said that the teachers delivering the *PreLit* instruction in this study did so with a high degree of fidelity.

Results

As may be seen from Table 1, and as previously noted, the experimental and comparison groups were highly comparable on all measures employed at pre-test. Statistical analyses of group differences (*t*-tests) confirmed that the group means were not significantly different on any of the three measures.

Highly statistically significant gains were made by the sample as a whole on all three measures ($p < .0005$), as would be expected following 25 weeks attendance at school between testing, but the means for the two groups remained very similar at post-test on all three measures and also on the additional post-test measure (LeST) – see Table 2. None of the group mean differences on the measures was significantly different.

This was confirmed by analyses of covariance of post-test scores between the experimental and comparison groups, covarying pre-test scores. (In the case of the LeST measure, which was not given at pre-test, the pre-test scores for SPAT-R were employed as a proxy covariate because the SPAT-R correlated most closely with LeST at post-test; $r = .731$.) The mean scores and standard deviations of the experimental and comparison groups for each measure are presented in Table 2. There were no statistically significant differences between the post-test mean scores of the two groups, covarying pre-test scores (or equivalent), on any measure.

The *variability* of the two groups on the measures of phonological awareness (SPAT-R) and early reading (LeST) was, however, significantly different (Levene's test for equality of variances, $p < .0005$). While the means were similar, the variability in scores for the two groups was statistically significantly different.

At pre-test, the percentage of students scoring in the bottom quartile on the SPAT-R, was 90% in the comparison group and 92% in the experimental group, the remainder falling in the average range (middle 50%). The distributions were not statistically significantly different. At post-test, however, the situation was very different – see Table 3. Less than 3% of students were still in the bottom quartile in the experimental group while 16% of students remained in the bottom quartile in the comparison group. Moreover, 55% of experimental group students were in the top quartile at post-test compared to 45% of comparison group students. The distributions of the two groups were statistically significantly different (Likelihood ratio 14.068, $p < .001$). The distributions for the LeST measure (given at post-test only) were not significantly different.

Closer inspection of the data revealed that there appeared to be greater differences among school means in the comparison group than in the experimental group. For SPAT-R, all experimental group schools had means of between 32 and 33; for comparison schools the means varied from 20 to

41. For LeST, all experimental group school means were between 28 and 34; for comparison group schools, the means varied from 21 to 43. As may be seen from Figure 1, not only is the variability of means greater among the comparison schools but the mean for C3 school is the highest of all schools for both measures.

Analyses of co-variance comparing the means across schools for SPAT-R at post-test for the comparison group confirmed that the means were significantly different ($p < .0005$) with a large effect size (partial eta squared .321). For the experimental group schools, the means were not significantly different. For the LeST, the means across the four comparison schools were also significantly different ($p < .0005$), again with a large effect size (partial eta squared .575) but the differences between experimental schools on this measure were also significantly different ($p < .0005$), albeit with a much smaller effect size (partial eta squared .181).

Given the exceptional status of school C3, the only school that appeared to have implemented a thoroughgoing phonics instruction program (Jolly Phonics; Lloyd, 1992), and aspects of MiniLit (MultiLit, 2011) plus other programs (see Appendix 1), the data was re-analysed, omitting the data for this one comparison school. Analyses of covariance on this reduced sample were conducted as before. The group means were again not statistically significantly different for the two language measures, PPVT and SCT, but for the two phonological processing measures, SPAT-R and LeST, the mean differences were now statistically significant, favouring the experimental group, with medium effect sizes evident for both ($p < .0005$; partial eta squared values of .126 for SPAT-R and .072 for LEST; medium effect size confirmed by Cohen's *d* values of .59 and .50 respectively).

Discussion

The aim of this study was to determine the effectiveness of a literacy preparation program (*PreLit*) designed to improve the emergent literacy skills of children in the year before Kindergarten, but deployed with young children entering Kindergarten in this case. As noted earlier, the original plan to randomly allocate schools to conditions proved to be unacceptable to the relevant Schools Office and, hence, the performance of the students in the four PreLit schools was simply compared with four schools who did not receive the program. This methodological problem compromises the extent to which sound, unequivocal conclusions can be drawn from the study and we acknowledge

this. However, some tentative suggestions may be drawn regarding the relative performance of the two groups.

While it had initially been suggested that the students could be classified as socially disadvantaged, the ICSEA (the Index of Community Socio-Educational Advantage) values of their schools showed that none of the schools were outside the average range for ICSEA. Moreover, and as previously noted, the verbal ability for the total sample as indicated by the mean PPVT standardised score was, in fact, average. To this extent, the sample could not be regarded as socially disadvantaged. That they may be said, however, to be ill prepared for learning to read is indicated by the fact that over 90% of the sample scored in the bottom quartile on the measure of phonological awareness. They may, therefore, be regarded as ill-prepared for learning to read compared with the national average. This background information on the children at program commencement, moreover, suggests that while they may not have been in need of additional preparation in terms of oral language skills, they were likely to benefit from instruction in phonological awareness skills. This proved to be the case.

The results initially failed to show any difference between the mean gains on measures of emergent literacy skills of children in the experimental compared to the comparison group. This was surprising, considering the success of other intervention programs addressing emergent literacy skills in preschools (Hart & Risley, 1995; Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013).

Research has shown that short daily periods of instruction in PA resulted in significantly better performance on measures of PA, letter sound blending and elision, with the effects on PA and blending being the most consistent (Lonigan et al., 2013). Similarly, instruction in oral language skills using a dialogic reading method has been shown to have a significant positive effect on children's vocabulary. Furthermore, children who received instruction in letter knowledge improved their ability to identify letter sounds over children who did not receive the instruction. These findings indicate that the type of instruction used in the present study should have been effective for improving children's emergent literacy skills. The results of the present study were also initially surprising given that the *PreLit* program used has been based on years of research into necessary components of successful early literacy programs (Callaghan & Madelaine, 2012).

The outcome of the study may have resulted from limitations in the methodology but program delivery issues are unlikely to have contributed to the outcomes of the study. Treatment integrity was generally good throughout the period of delivery. The teachers delivering *PreLit* instruction did so with a high degree of fidelity following a one-day professional development workshop. They delivered the program with enthusiasm and had positive experiences using the program, indicating that they could see the benefits of the instruction for the children. This enthusiasm grew over the course of the intervention.

Over the course of three terms (25 weeks of instruction), appreciable gains in performance were to be expected and while all students did indeed make substantial gains on the measures of vocabulary (PPVT-IV) and sentence comprehension (SCT), there was no evidence from the study to show differential improvement by the experimental group students over those in the comparison group, in any of our analyses. The students were not in particular need of further language intervention and no gains attributable to the program were shown. We shall not discuss the results for the language measures further but shall focus on the measure of phonological awareness (SPAT-R) and the measure of letter-sound skills (LeST).

As was the case with the two language measures, the two groups were very similar in terms of mean pre-test scores on SPAT-R, the measure of phonological awareness. None of the mean differences were statistically significant at pre-test. Similarly, the means for the two groups were very similar and not significantly different at post-test on any of the three measures; nor was the difference between the two groups significantly different on the LeST (letters and sounds) measure, given at post-test only. Unlike the two language measures, however, the variability in scores at post-test was considerably and significantly greater for the comparison group than for the experimental group on both the SPAT-R and the LeST ($p < .0005$), as previously noted.

This was confirmed by a comparison of the distributions of scores for the two groups on the SPAT-R. While at pre-test, the percentage of students scoring in the bottom quartile was very similar, about 90% in both groups (the remainder in the middle 50%), at post-test, the distributions had changed (see Table 3). At post-test, as already noted, less than 3% of students were still in the bottom quartile in the experimental group compared with 16% of students in the comparison group; 55% of

experimental group students were in the top quartile compared to 45% of comparison group students. The two distributions were statistically significantly different (Likelihood ratio 14.068, $p < .001$).

These findings warranted further exploration of the data including the descriptive data regarding the schools available to us (see Appendix 1). As Figure 1 clearly illustrates, the variability of post-test mean scores across the four schools in the comparison group is greater than for the four schools in the experimental group, with school C3 clearly ahead of the other seven schools, for both SPAT-R and LeST. The descriptive data regarding the eight schools suggested a possible explanation for these findings. School C3 was the only one of the four comparison schools where a systematic program of phonics instruction (Jolly Phonics, aspects of MiniLit, and other programs) was faithfully and fully implemented in both of the classes included in the study. For this reason an exploratory secondary analysis was carried out in which School C3 was excluded from the analyses. The results this time, for both measures, revealed statistically significant differences between the two groups ($p < .0005$) with medium effect sizes ($d = .59$ for SPAT-R and .50 for the LeST) in favour of the experimental group.

It could be tentatively suggested that first, the issue of differential variability between the two original groups and greater similarity across schools in the experimental group suggests the possibility that *PreLit* is more uniform in its effects on pre-literacy skills, in effect keeping the students closer together. This possibility is supported by the finding that less than 3% of students were still in the bottom quartile in the experimental group at post-test on the SPAT-R measure compared with 16% in the original comparison group. Considering that Hempenstall (2016) reports that 30% of first graders do not appreciate the phonemic structure of words, this result on the SPAT-R is notable in that it provides some evidence that *PreLit* has provided a solid foundation in phonological awareness for almost every student in the experimental group. It should also be emphasised that this was not at the expense of the higher performing students; 55% of experimental group students at post-test were in the top quartile compared to only 45% of students in the original comparison group. (As already noted, the distributions for LeST, given at post-test only, were not significantly different, however.)

Second, there is tentative evidence to support the efficacy of the *PreLit* program with beginning Kindergarten Year students, *in the absence of a systematic program of phonics instruction*

in schools, such as an elaborated Jolly Phonics approach (as received by students in the excluded school, C3). The performance of the experimental group was clearly and substantially superior to that of the students in the attenuated comparison group of schools who received a generic mix of instructional approaches in their classes.

At this point, it is worth emphasising that these suggestions are speculative, not least because the two groups were not initially randomly allocated to treatments. They may hold true for the groups included in this study but they may not be generalisable to the wider population. It should also be noted that the comparison group schools were effectively a "business as usual" condition; they did not receive an alternative, novel treatment.

There is some evidence, then, that *PreLit*, while having no appreciable effect on language skills, in this study, may be effective in terms of improving the emergent literacy skills of beginning Kindergarten students. On the other hand, it may be more effective to include a thoroughgoing phonics program instead with these children. *PreLit* may be more sensibly reserved for the pre-school children for whom it was designed and/or, possibly, for students commencing school who come to school with low levels of oral language facility and phonological awareness from very socially disadvantaged backgrounds.

Disclosure statement

The first and last authors are directors of MultiLit Pty Ltd, the publisher of the PreLit program which is the subject of this research.

Acknowledgments

We sincerely and thankfully acknowledge the cooperation of the Catholic Schools Office, Armidale and the staff and students of the schools attended.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2015). *Guide to understanding ICSEA (Index of Community Socio-educational Advantage) values: From 2013 onwards.* Australia: Australian Curriculum, Assessment and Reporting Authority. Retrieved from http://www.acara.edu.au/verve/_resources/Guide_to_understanding_ICSEA_values.pdf

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why Jaydon can't read: The triumph of ideology over evidence in teaching reading. *Policy, 29*, 21–32. Retrieved from http://www.cis.org.au/publications/policy-magazine

Callaghan, G., & Madelaine, A. (2012). levelling the playing field for kindergarten entry: Research implications for preschool early literacy instruction. *Australasian Journal of Early Childhood, 37*, 13–23. Retrieved from http://www.earlychildhoodaustralia.org.au/our-publications/australasian-journal-early-childhood/

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*, 684– 698. doi:http://dx.doi.org/10.1111/j.1467-8624.1994.tb00777.x

Carrow-Woolfolk, E. (1999). *Comprehensive assessment of spoken language*. Circle Pines, MN: American Guidance Service.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77–111. doi:http://dx.doi.org/10.1016/S0010-0277(03)00164-1

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*, 1142– 1157. doi:http://dx.doi.org/10.1044/1092-4388(2002/093)

Chapman, M. L. (2003). Phonemic awareness: Clarifying what we know. *Literacy Teaching and Learning, 7*, 91–114. Retrieved from http://readingrecovery.org/rrcna/journals/ltl-archive

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology, 98*, 489–507. doi:http://dx.doi.org/10.1037/0022- 0663.98.3.489

Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology, 98*, 665–689. doi:http://dx.doi.org/10.1037/0022- 0663.98.4.665

Department of Education, Employment and Workplace Relations (DEEWR). (2009). *Belonging, being and becoming: The early years learning framework for Australia.* Australia: Commonwealth of Australia. Retrieved from http://docs.education.gov.au/system/files/doc/other/belonging_being_and_becoming_the_early_years_learning_framework_for_australia.pdf

Doyle, B. G., & Bramwell, W. (2006). Promoting emergent literacy and social-emotional learning through dialogic reading. *The Reading Teacher, 58*, 554–564. doi:http://dx.doi.org/10.1598/RT.59.6.5

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: NCS Pearson Inc.

Dunn, L. M., Dunn, L. M., & Whetton, C. (1982). *The British picture vocabulary scale*. Windsor: NFER- NELSON.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes.

Hempenstall, K. (2016). *Read about it: Scientific evidence for effective teaching of reading* (CIS Research Report 11). Sydney: The Centre for Independent Studies. Retrieved from https://www.cis.org.au/app/uploads/woocommerce_uploads/2016/03/rr11.pdf

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160. doi:http://dx.doi.org/10.1007/BF00401799

Larsen, L., Kohnen, S., Nickels, L., & McArthur, G. (2015). The Letter-Sound Test (leST): A reliable and valid comprehensive measure of grapheme-phoneme knowledge. *Australian Journal of Learning Difficulties, 20*, 129–142. doi:http://dx.doi.org/10.1080/19404158.2015.1037323

Lloyd, S. (1992). *The jolly phonics handbook*. Essex: Jolly learning.

Lonigan, C. J., Purpura, D. J., Wilson, S. B., Walker, P. M., & Clancy-Menchetti, J. (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk

for reading difficulties. *Journal of Experimental Child Psychology, 114*, 111–130. doi:http://dx.doi.org/10.1016/j. jecp.2012.08.010

Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading.* Retrieved from http://timssandpirls.bc.edu/pirls2011/downloads/P11_IR_FullBook.pdf

Multilit. (2011). *MiniLit early literacy intervention program*. Sydney: Multilit Pty Ltd.

Multilit. (2012a). *PreLit early literacy preparation program*. Sydney: Multilit Pty Ltd.

Multilit. (2012b). *PreLit manual*. Sydney: Multilit Pty Ltd.

National Early Literacy Panel (NELP). (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for literacy. Retrieved from http://www.pathstoliteracy.org/sites/default/files/uploaded-files/NElPReport09.pdf

Neilson, R. (2003a). *The astronaut invented spelling test*. Jamberoo: language, Speech and literacy Services.

Neilson, R. (2003b). *Sutherland phonological awareness test* (revised ed.). Jamberoo: Roslyn Neilson.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: Psychological Corporation.

Serry, T., Rose, M., & Liamputtong, P. (2008). Oral language predictors for the at-risk reader: A review. *International Journal of Speech-Language Pathology, 10*, 392–403. doi:http://dx.doi.org/10.1080/17549500802056128

Wheldall, K., Mittler, P., & Hobsbaum, A. (1987). *Sentence comprehension test* (revised ed.). Berkshire: NFER-NELSON.

Whitehurst, G. J., Falco, F. l., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24*, 552–559. doi:http://dx.doi.org/10.1037/0012-1649.24.4.552

Williams, K. T. (2007). *Expressive vocabulary test* (2nd ed.). Circle Pines, MN: AGS Publishing.

Woodcock, R. W. (1987). *Woodcock reading mastery test* (revised ed.). Circle Pines, MN: American Guidance Service.

Table 1

Means and standard deviations for the experimental and comparison groups on the three measures at pre-test and chronological age.

		N	Mean	Std. Deviation
Pre SPAT-R				
	Experimental	115	8.878	5.3511
	Comparison	125	8.840	5.8797
Pre SCT				
	Experimental	115	31.809	4.3405
	Comparison	125	31.312	5.1485
Pre PPVT				
	Experimental	115	97.183	15.7350
	Comparison	125	96.816	20.5550
Chronological Age (months)				
	Experimental	115	64.696	4.1383
	Comparison	125	65.624	4.3564

Table 2

Means and standard deviations for the experimental and comparison groups on the three measures at post-test and on the LeST.

	Comparison		Experimental			
Measure	Mean	SD	Mean	SD	F	p
SPAT-R	31.26	13.13	32.93	8.69	1.74	0.188
SCT	35.97	3.40	36.32	2.76	0.27	0.606
PPVT	109.93	17.54	110.15	15.78	0.00	0.971
LeST	31.52	10.38	30.09	6.37	1.63	0.203

Table 3

Proportions of students falling into the top and bottom quartiles and middle 50% on post-test SPAT-R measure for the experimental and comparison groups.

		Post-test SPAT-R Quartile			
		Bottom Quartile	Average	Top Quartile	Total
Comparison group					
	Count	20	49	56	125
	Expected Count	12.0	51.0	62.0	125.0
	% within Group	16.0%	39.2%	44.8%	100.0%
Experimental group					
	Count	3	49	63	115
	Expected Count	11.0	47.0	57.0	115.0
	% within Group	2.6%	42.6%	54.8%	100.0%

Appendix 1. Details of schools

Experimental schools

E1

This school had an ICSEA value of 978.

The children participating in the study from this school comprised students from three classes. The three class teachers were supported by three student support teachers who taught half the class (the "higher" performing students) for literacy sessions of two hours (including *PreLit*) for three sessions per week. There were six literacy groups taught by eight teachers. The approaches to literacy teaching across the three classes appeared to vary (from the data supplied by the teachers and support teachers), although all seem to reflect a phonics approach. One class teacher made reference to explicit instruction. Letter-sound correspondences were introduced in clusters over several weeks. A variety of commercial programs were used by the teaching staff including Ants in the Apple, Reading Eggs, Jolly Phonics songs, PM Oral Literacy Kit, in addition to the *PreLit* instruction. Total literacy instructional time (including *PreLit*) took place for 120–150 min per day, depending on the day.

E2

This school had an ICSEA value of 1013.

The children participating in the study from this school comprised students from two classes, each with one teacher. Aspects of the commercial program Sound Waves was used in both these classes, in addition to the *PreLit* instruction. Total literacy instructional time (including *PreLit*) took place for 90–150 min per day.

E3

This school had an ICSEA value of 1021. The children participating in the study from this school comprised students from one class, with two teachers. No reference was made to phonics instruction in the listed literacy activities that took place in addition to the *PreLit*instruction. Total literacy instructional time (including *PreLit*) took place for 100 min a day on average.

E4

This school had an ICSEA value of 987. The children participating in the study from this school comprised students from one class, with one teacher. This teacher used Jolly Phonics from Term 2, in addition to the *PreLit* instruction, as well as other literacy activities. Total literacy instructional time (including *PreLit*) took place for between 105 and 120 min a day.

Comparison schools

C1

This school had an ICSEA value of 1065.

The children participating in the study from this school comprised students from one class, with one teacher. There was strong phonics and phonological emphasis in the program content, the program being supported by aspects of the MiniLit program (sounds introduced in line with the MiniLit program sequence, for example) and the Sound Waves program. Total literacy instructional time was 120 min per day, each morning.

C2

This school had an ICSEA value of 990.

The children participating in the study from this school comprised students from one class, with one teacher. The teacher incorporated phonemic and phonological awareness activities daily, as well as a "daily" phonics-based lesson, as well as other literacy activities. (No specific commercial programs were specified.) The students participated in literacy lessons four days a week (Mon–Thurs) for 120 min each day.

C3

This school had an ICSEA value of 1009.

The children participating in the study from this school comprised students from two classes, each with one teacher. One teacher was very experienced at teaching Kindergarten students and the other was a first year out teacher. Both teachers appear to have followed a similar program that took a "phonics first and fast" approach, using the Jolly Phonics program. All 42 sounds were introduced in the first term of the year, with further work being done using Jolly Phonics materials (Book 2) in the second term. The experienced teacher also used MiniLit program techniques, as well as other strategies "to supplement and enrich the Jolly Phonics program". This teacher commented, "This program was implemented by myself last year with relatively good success – children learn to sound out, read and write with quite good competence by the end of the year". The students participated in literacy lessons every day for 90 min a day in the classroom of the experienced teacher, and from 60 to 120 min per day in the class of the novice teacher. It is interesting to note that the experienced teacher taught literacy for the least amount of time daily of all the teachers in the study (90 min per day).

C4

This school had an ICSEA value of 1037.

The children participating in the study from this school comprised students from two classes, each with one teacher. Students were taught on average about 100 min per day, in both classrooms. Both teachers used a phonics approach to teaching reading, with the use of commercial programs such as Jolly Phonics (in one class) and ABC Reading Eggs, Sunshine Online being used in both classrooms, as well as other materials. The introduction of letter-sound correspondences was done more slowly

than in the Jolly Phonics program (see above C3), with letters and sounds being introduced in clusters of five sounds over five weeks, explicitly taught. Consonant blends were taught earlier to students in the “top” group.

Appendix 2. Details of the PreLit early literacy intervention program

The *PreLit* program comprises two components. The first component is aimed at improving children's phonological and phonemic awareness. This instruction comprised scripted lessons that targeted these specific skills through play-based activities. There are 108 scripted lessons ordered hierarchically from word level to phoneme level (MultiLit, 2012b) and according to task difficulty, linguistic complexity and the position of phonemes within words. The lessons develop PA skills using activities involving identifying a particular sound in a word, blending spoken syllables or phonemes into whole words and segmenting spoken words into their component sounds.

The second component focused on the development of oral language skills. This involves instruction in definitional vocabulary, listening comprehension, grammar and syntax (MultiLit, 2012b) through structured storybook reading where children are encouraged to engage with the story being read, to answer questions about the story and to relate the events to their own lives. Questions asked by the teacher are open-ended and expansion of answers is encouraged via modelling. This method is known as dialogic reading, described previously. Question are literal, inferential, evaluative and definitional. Teachers read selected storybooks to the children three times on separate days, using storybook cards comprising instructions for three sessions.

The first session provides an introduction to the storybook and involves exposure to concepts about print, providing some background for the book, drawing on children's experience. The teacher then reads the story to the children for the first time and definitions of target words are provided. At the end of the session, the teacher links the story to the content of the PA lesson with which the session was paired. In the second session children are encouraged to discuss the book via questions which require them to remember the story, facilitating listening comprehension, expressive language and event sequencing (MultiLit, 2012b). The teacher then reads the story again incorporating questions that develop oral language and comprehension skills. Words that were defined in the first session are revisited and understanding is determined. Questions are then asked which summarise the story. In the final session, previously defined target words are revised and used in language activities; understanding of the words is determined again.

Furthermore, both the PA lessons and shared Storybook Reading incorporate instruction in print awareness. In the scripted PA lessons, this is done through teaching children to recognise some of the written letters, linking them to their most common sounds (MultiLit, 2012b). (An extract from the *PreLit* manual showing the scope and sequence for the PA component is shown below.) Children also engage in activities such as blending onset and rime (in this case, first and remaining sounds in a word respectively) and matching printed words to pictures. In the shared Storybook Reading sessions, children are taught that print has meaning and are exposed to the different elements of print in the context of a picture