
Evaluation of a Two-Phase Implementation of a Tier-2 (Small Group) Reading Intervention for Young Low-Progress Readers

Jennifer Buckingham, Kevin Wheldall and Robyn Beaman-Wheldall

Macquarie University Special Education Centre, Macquarie University, Sydney, Australia

In a response to intervention (Rtl) model, reading is taught in increasingly intensive tiers of instruction. The aim of the study was to examine the efficacy of a Tier-2 (small group) literacy intervention for young struggling readers. This article focuses on the second phase of a randomised control trial involving 14 students in kindergarten as participants. In Phase 1 of the randomised control trial, the experimental group (E1) received the intervention for 1 hour, 4 days per week, for 3 school terms. The control group received regular classroom instruction. Large and statistically significant mean differences between groups were evident after 3 terms on 2 of 4 measures – the Martin and Pratt Nonword Reading Test and the Burt Reading Test, which measure phonological recoding and single word reading, respectively. Very large effect sizes were found. In Phase 2, the original control group received the intervention in the same way (E2). Testing at the end of Phase 2 confirmed the intervention's large effect on phonological recoding, but the results for the 3 other tests showed no acceleration in the Phase 2 experimental group (E2). This study evaluates the efficacy of the trialled intervention, adds to the research literature on Tier-2 interventions for young struggling readers, and yields practical implications for schools that offer literacy interventions without a strong Rtl framework.

Keywords: intervention, reading, phonics, low progress, instruction, Tier 2

Literacy is the bedrock of education and is the prime focus of the early years of schooling. In public schools of New South Wales (NSW), Australia, for example, the first 2 hours of each day are devoted to literacy. Even so, a substantial number of children cannot read at even a functional level after 4 years at school. In the 2012 National Assessment Program for Literacy and Numeracy (NAPLAN), 6.4% of Year 3 students failed to achieve the national minimum reading standards expected for their year of education. A further 10.4% achieved only the minimum standard (Australian Curriculum, Assessment and Reporting Authority, 2012). NAPLAN data contribute to a large body of literature showing a moderate and persistent relationship between literacy and socioeconomic status (SES; Marks, 2009; Rothman, 2002; Rothman & McMillan, 2003; Thomson,

Correspondence: Jennifer Buckingham, The Centre for Independent Studies, PO Box 92, St Leonards, NSW 1590, Australia. E-mail: jbuckingham@cis.org.au

de Bortoli, Nicholas, Hillman, & Buckley, 2010). In 2012, 33.4% of Year 3 children whose parents had not completed secondary school achieved at or below the minimum standard in the NAPLAN reading tests, as did 31.6% of children whose parents had not been in paid work in the previous year (Australian Curriculum, Assessment and Reporting Authority, 2012).

Within each socioeconomic group, a wide range of reading performance exists, showing that a student's SES does not determine reading performance. Although the research on SES and literacy achievement shows the significant impact of a number of out-of-school factors, including the early home literacy environment (Aikens & Barbarin, 2008; Dodd & Carr, 2003; Eamon, 2005; Hart & Risley, 2003), according to Fuchs and Fuchs (2006), the majority of reading difficulties arise from an instructional deficit and are therefore amenable to school-based strategies. Likewise, Carnine, Silbert, Kame'enui, Tarver, and Jungjohann (2006) argued that, irrespective of the reasons for children's reading difficulties, the most effective and immediate way to improve current reading ability of school-age children is through exemplary teaching, namely, explicit and systematic instruction in the fundamentals of reading: phonemic awareness, phonics, fluency, vocabulary and comprehension. Explicit and systematic instruction in these concepts and skills is effective for all children but is particularly important for children from disadvantaged backgrounds, who are less likely to have been exposed to these concepts in their family and home environments (Department of Education, Science and Training, 2005; National Institute of Child Health and Human Development [NICHD], 2000; Rose, 2006).

Some children will still struggle to learn to read even with effective initial instruction and will require extra instructional support at various levels of intensity and duration. Response to intervention (RtI) is a model for instruction and assessment that focuses principally on quality and quantity of instruction rather than the causes of reading difficulties. The RtI model has multiple levels or 'tiers' of instruction, which increase in intensity and duration (Gersten et al., 2009). In an RtI model with three tiers, Tier 1 is whole-class instruction and Tier 2 is small-group, supplementary, intensive, research-based instruction for students who are identified as 'struggling readers'— those who do not meet reading standards in the classroom setting, sometimes defined as the bottom 25% of their age cohort (Wheldall, 2009). Tier 3 is intensive individual instruction for the small number of children who do not respond to Tier-2 intervention and require specialised support. The RtI model uses rigorous assessment tools to identify struggling readers, to monitor their progress, and to determine which level of intervention they require (Fuchs & Fuchs, 2006; Gersten et al., 2009).

For children with the greatest reading difficulties, research indicates that one-to-one tutoring sessions are the most effective. For children with milder problems, however, a strong evidence base supports small-group interventions, especially in the early years of school (Gersten et al., 2009; Slavin, Lake, Davis, & Madden, 2011; Wanzek & Vaughn, 2007). Small-group instruction is also more cost-effective.

Ideally, intervention for children with reading difficulties should begin to be instituted no later than after one year of formal schooling (Gersten & Dimino, 2006; Torgesen, 2000) before problems become entrenched and the achievement gap widens (reviewed in Rose, 2006). Three major reports on reading in Australia (Department of Education, Science and Training, 2005), the United Kingdom (Rose, 2006), and United States (US; NICHD, 2000) conclude that reading skills are most likely to develop with phonics-based reading programs, in which children are taught explicitly about letter-sound correspondences, phonemic awareness and generative strategies in both initial instruction and remedial instruction settings. In a review of reading interventions for Year 1 students, Reynolds,

Wheldall, and Madelaine (2011) found that very few early interventions were comprehensive literacy programs that included instruction in phonemic awareness and phonics, and few programs had methodologically sound empirical evidence for their efficacy. The most widely used early reading intervention in Australia — Reading Recovery — does not include this content in a systematic way (New South Wales Department of Education and Communities [NSW DEC], 2012; Reading Recovery Council of North America, 2007).

What Works Clearinghouse (WWC) at the US Institute of Education Sciences has evaluated a number of Tier 2 reading programs. At least some statistically significant positive results in phonemic awareness and phonics were found among young students in these programs, but with varying degrees of supporting evidence. These programs included:

1. Lindamood Phoneme Sequencing[®], which focuses entirely on explicit instruction in phonemic awareness and phonics (WWC, 2008b, 2010).
2. Peer-Assisted Learning Strategies (PALS), in which highly skilled students coach and mentor their peers with limited reading skills within the classroom (WWC, 2012).
3. Success For All[®], which comprises both a whole-school, complete literacy curriculum and small-group supplementary instruction for struggling readers (WWC, 2009).
4. Early Intervention in Reading[®], which comprises both whole-class and small-group supplementary instruction (WWC, 2008a).
5. SpellRead[™], which is delivered in 60–90 minute lessons over 5–9 months, with explicit instruction in phonemic awareness and phonics (WWC, 2007).

Another Tier-2 program (not evaluated by WWC) with published research is called Early Literacy Support (ELS) in which students alternate daily between small-group and individual instruction (Hatcher et al., 2006). A randomised control trial demonstrated that students who had participated in a 10-week ELS program had significantly higher growth in letter knowledge, single word reading and phoneme awareness than the control group. Medium to large effect sizes were reported (Hatcher et al., 2006).

RtI models with three tiers of instruction are not common in Australia; instead, formal remedial reading programs are likely to be in the form of one-to-one tutor programs (possibly a Tier-3 intervention in an RtI model), with small-group instruction (Tier 2) being either informal or absent (Louden et al., 2000; van Kraayenoord, 2010). Only a few formal reading interventions for young struggling readers are identified in major reports on literacy interventions in Australia, the dominant program being Reading Recovery (Louden et al., 2000; van Kraayenoord, 2010; Wyatt-Smith, Elkins, Colbert, Gunn, & Muspratt, 2007). None meet the criteria of a Tier-2 program within an RtI model for children in the first few years of school, but the most well known of these are:

1. THRASS[®] (Teaching Handwriting, Reading and Spelling Skills) is a phonics teaching method that is used in some schools as a whole-class program and in others for remedial instruction for smaller groups (Louden et al., 2000). It is designed to teach student the ‘word level’ component of literacy; it is not a comprehensive program (The THRASS Institute, 2014). There is limited evidence that it is effective as an intervention for young students in improving word attack skills (Brooks, 2007; Symons & Greaves, 2006).
2. L3 (Language, Learning and Literacy) is a program that was introduced to NSW public schools in 2010. Students work in small groups on tasks differentiated for ability level while teachers move around to each group giving 10-minute targeted lessons. It is designed to work as an in-class intervention that precedes, and aims to reduce, progress

into Reading Recovery (NSW DEC, 2011c). No research has been published on trial implementations and no evaluations are publicly available.

The MiniLit (Meeting Initial Needs In Literacy) program is designed as a Tier-2 program for young students (in the first 3 years of school) who are performing below their peers in reading acquisition. It comprises all of the elements of effective early literacy instruction as determined by large-scale reviews of research (Reynolds, Wheldall, & Madelaine, 2007; Reynolds et al., 2011). Unlike other interventions, such as Reading Recovery, it is not focused on a single year of schooling (Year 1) and is inclusive of the very lowest performing students, moving them on to more intensive (Tier 3) instruction only if they do not make progress (Reynolds & Wheldall, 2007). More detail about the MiniLit program is provided in the Methods section.

Several pilot studies have guided the development and investigated the effectiveness of the MiniLit program. Three initial pilot studies of an early version of the intervention involved Year 1 and Year 2 students in MiniLit sessions for 1 hour each day, 4 days a week, over 15 weeks. The lessons took place during school time in a tutorial centre external to the students' schools. Participating students were tested on the same test battery prior to and after completing a 15-week MiniLit program. In each of the studies, participating students made statistically significant gains on all measures. Treatment causality cannot be claimed as these studies did not have control groups, but there were large effect sizes (Reynolds et al., 2007).

A fourth trial had an experimental design, with 16 Year 1 students randomly assigned into treatment and control groups. In the first phase, one group of eight students (Group 1) participated in a shortened version of the MiniLit program used in the previous trials while the other eight students (Group 2) were the control group. In the second phase, Group 1 became the control group and Group 2 participated in MiniLit. The program was delivered in a school by school staff who were trained to teach the program (Reynolds, Wheldall, & Madelaine, 2010). At the end of the 10-week MiniLit intervention no statistically significant differences between the experimental and the control group in either phase of the study were noted. Large effect sizes were evident, however. The large effect sizes suggest that a larger sample size with more statistical power would have achieved significance. Another possible contributing factor to the weaker results in this trial is that not all students in the study scored below the 25th percentile on all pretest measures, so could not strictly be described as the 'struggling readers' for whom MiniLit is designed (Reynolds et al., 2010). Finally, the intervention was given for only one term.

MiniLit has also been implemented at the Schoolwise Tutorial Centre at Ashfield in New South Wales since 2005. During the 3 years between 2009 and 2011, 161 students were tested before and after participation in a 15-week MiniLit program. There was no control group. After 15 weeks, students had made substantial and statistically significant gains on all of the measures of reading and related skills, with large effect sizes. Their average reading fluency was shown to have increased by 90% (Wheldall, Beaman, Madelaine, & McMurtry, 2012).

More recently, a further randomised control trial using a revised version of MiniLit was implemented with 22 students from kindergarten and Year 2 in a NSW public school. Students in the experimental group received MiniLit instruction for 1 hour a day, 4 days a week for three terms (27 weeks). At the end of the intervention, the experimental group had significantly higher scores than the control group on measures of phonological recoding and word reading, with very large effect sizes. No significant differences were noted between the groups in spelling (but a large effect size) and no discernible effect

on students' single word reading fluency scores were evident (Buckingham, Wheldall, & Beaman, 2012).

This paper focuses on a subset of 14 students from the randomised control trial reported in Buckingham et al. (2012). In this study, the kindergarten students formed a two-phase, crossover study extended over six terms. The study aimed to gather further experimental evidence of the efficacy of MiniLit in a school setting, and particularly the potential of MiniLit as a cost-effective, Tier-2 intervention, to improve literacy in schools with a high level of social disadvantage.

Method

Participants

Participants were a subset of students involved in a larger sample ($n = 22$) three-term randomised control trial, which has been reported in Buckingham et al. (2012). Fourteen students in kindergarten participated in the six-term trial. The Year 2 students were excluded from the six-term crossover study because they moved into Year 3 after the first three terms of intervention, and thus were outside the target group for MiniLit (K–2).

The site of the study was a regional NSW public school with a low socioeconomic profile. When the study began in 2010, the school had an Index of Community Socio-Educational Advantage (ICSEA; Barnes, n.d.), calculated for the federal government's My School website, of 897. (The national average ICSEA value is 1000, with a standard deviation of 100.)

Selection of students for participation in the study was completed through a screening and ranking process. The lowest 50% of students in each class in terms of reading ability was identified by kindergarten classroom teachers. Excluded from the study were students with a diagnosed (and documented) intellectual disability or severe language impairment, as their needs would have been addressed by alternative provision.

Identified students were screened by trained research assistants using two lists from the Wheldall Assessment of Reading Lists (WARL; Reynolds, Wheldall, & Madelaine, 2009) and then ranked according to their mean scores. The lowest ranked 16 students from kindergarten were selected for the study. Information and consent forms were sent home to the parents/carers of selected students. Passive consent for the students' participation in the study was required by the university research ethics committee and the State Education Research Approval Process (SERAP) office of the New South Wales Department of Education and Communities (NSW DEC).

Matched pairs were created using scores on the MiniLit Placement Test (described in the Appendix) and students were randomly allocated to two groups: the experimental (treatment) group and the control group. One student from kindergarten left the school in the ninth week of the study, reducing the total number of participants for the study from 16 to 14 students (the data for the matched student were also excluded from the study).

The participants included 10 boys and four girls, with a mean age of 67 months (5 years, 7 months) at the beginning of the intervention. The primary language for all participants was English. Standardised tests were administered to determine baseline measures (described in the Appendix).

After two terms, all but one student, who began the intervention in kindergarten, moved into Year 1. One student repeated kindergarten. All 14 students remained in the study for six terms.

Procedure

The study was implemented as a two-phase, crossover design. Each phase was three school terms (approximately 27 weeks of instruction). The group who received the intervention in Phase 1 of the study — Experimental Group 1 (E1) — became the comparison group in Phase 2. In Phase 2 of the study, the control group from Phase 1 received the intervention and became Experimental Group 2 (E2). Phase 1 was implemented as a randomised control trial, but in Phase 2 the group not receiving the intervention was not under control conditions and will be referred to as the ‘comparison’ group.

The MiniLit program was delivered to students in the experimental groups in each phase for 1 hour each day, 4 days a week, over three school terms (27 weeks). Students were withdrawn from class during classroom literacy time. Instructors were trained by MultiLit (Making Up Lost Time in Literacy) trainers. (MultiLit is the entity responsible for developing the intervention.) Two of the instructors were registered primary school teachers and one had no teaching qualifications. Students in the comparison group remained in class and received usual classroom literacy instruction for the duration of the study. In Phase 1, control conditions were specified for the comparison group, but in Phase 2, students in the comparison group may have received another formal reading intervention for part of the time. The reading activities of the comparison group were not within the remit of the study. Following the intervention, the data collected in the three testing phases were analysed to compare the experimental and control/comparison groups’ score growth on the test battery. Lesson observations for treatment integrity took place at 3-week intervals by highly experienced special educators familiar with MultiLit practices and procedures.

All experimental and control/comparison group participants undertook a battery of reading tests before the reading intervention started, at the end of three terms and again at the end of six terms. The tests were administered by trained research assistants and all tests were independently scored and double-scored. The test battery included the Burt Word Reading Test (Gilmore, Croft, & Reid, 1981), the South Australian Spelling Test (Westwood, 2005), the Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001), and the WARL (Reynolds et al., 2009; using different lists to the screening WARL). The preintervention test battery also included the MiniLit Placement Test (MultiLit, 2011). Details of the tests are in the Appendix.

The Intervention: MiniLit

The MiniLit early literacy intervention program is a small-group instruction program for struggling readers in the first few years of school. It includes instruction in all of the elements of effective reading instruction identified in research — phonemic awareness, phonics, fluency, vocabulary and comprehension — taught in a direct, explicit and sequential manner (Reynolds et al., 2007, 2010; MultiLit, 2011).

Students were grouped by instructional level and taught in three groups of three to four students. There was some movement initially between groups as skill acquisition varied, but the groups became stable after around 10 weeks of instruction. The average attendance rate for MiniLit lessons was 96% (with a range of 91–100%).

Each 1-hour MiniLit lesson had the following components: Sounds and Words Activities (30–40 minutes), Text Reading (5–10 minutes) and Story Book Reading (10–15 minutes). The Sounds and Words Activities component includes highly structured, carefully scripted and sequenced instruction of phonemic awareness and phonics. Students first learn and master letter-sound correspondences, quickly progressing to blending and

segmenting these sounds in words, both orally and in print. Sight words are initially taught through text reading as ‘tricky words’ and later more explicitly as short lists.

Text Reading is first introduced as part of Sounds and Words Activities and later becomes a separate component. In the Sounds and Words Activities, students read sentences or simple short stories that reinforce phonic word attack skills and/or sight words they have been learning. In Text Reading, students read aloud from a controlled vocabulary book at their instructional level. Their instructor uses the revised Pause, Prompt, Praise tutoring method, as used in Reinforced Reading (Ellis, Wheldall, & Beaman, 2007; MultiLit, 2011).

Story Book Reading (10–15 minutes) is the final part of the lesson. This activity is less structured; the teacher reads a children’s storybook to the group, engaging them by commenting on the story and asking questions. The teacher models fluent, expressive reading and the students develop listening comprehension and vocabulary skills (MultiLit, 2011).

Direct instruction teaching (such as model-lead-test procedures; Carnine, Silbert, Kame’enui, & Tarver, 2010) is intrinsic to the MiniLit program. Positive teaching behaviour management strategies (Merrett & Wheldall, 1990; MultiLit, 2011) are used to maximise time on task.

Analysis

To compare gains made by the experimental and control/comparison groups after each phase of the intervention, analyses of covariance were employed for each measure at posttest 1 (after three terms) and posttest 2 (after six terms), with pretest scores as the covariate in each analysis. Raw scores were used in the analyses because many of the participants were younger than 6 years old, the minimum age for which standard scores are available for the measures employed. The alpha level was set at 1% ($p < .01$) to allow for family-wise comparisons in lieu of the use of a Bonferroni correction (Howell, 2008). Treatment effects were also calculated for each measure in each phase of the study, using partial eta squared as the measure of effect size, as calculated by the SPSS Version 21.0 statistical analysis package (IBM Corp, 2012).

Results

In this six-term study, only Phase 1 (the first three terms) was a controlled trial, where students not receiving the intervention remained in their usual classrooms and did not participate in any other formal remedial reading programs. In Phase 2, students were no longer in controlled trial conditions. Students not receiving the intervention in Phase 2 (the original experimental group in Phase 1) may have participated in other remedial reading programs and are therefore more accurately described as a ‘comparison’ group than a ‘control’ group. The Phase 1 experimental group/Phase 2 comparison group will be called ‘E1’, and the Phase 1 control group/Phase 2 experimental group will be called ‘E2’.

Means and standard deviations for all measures (raw scores) for the Phase 1 experimental group (E1) and the Phase 2 experimental group (E2) at pretest, end of Phase 1 (posttest 1, after three terms), and end of Phase 2 (posttest 2, after six terms) are shown in Table 1.

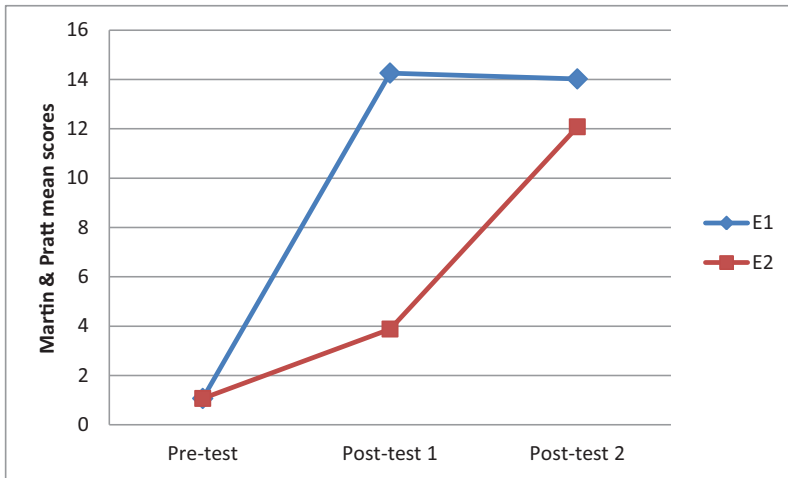
As may be seen in Table 1, the E1 group means were slightly lower than those for the E2 group at pretest for all measures, but none of these differences was statistically significant. (These small differences are taken into account in the analyses of covariance.) Analyses of covariance (ANCOVA) were conducted on the raw scores for each of these measures separately at posttest 1 and posttest 2 (with pretest scores as the covariate). Treatment

TABLE 1

Means and Standard Deviations (Raw Scores) of Phase 1 Experimental (E1) and Phase 2 Experimental (E2) Groups at Pretest, Posttest 1 (After Three Terms) and Posttest 2 (After Six Terms), Results of Analyses of Covariance and Effect Sizes (Partial Eta Squared)

Measure	Group (<i>N</i> = 14)	Pretest	Posttest 1	<i>F</i>	<i>p</i>	ES*	Posttest 2	<i>F</i>	<i>p</i>	ES
		Mean (<i>SD</i>)	Mean (<i>SD</i>)				Mean (<i>SD</i>)			
Martin & Pratt Nonword Test	E1	1.00 (1.16)	14.14 (5.58)	18.13	.001	.622	14.00 (7.19)	.397	.541	.035
	E2	1.14 (1.35)	4.00 (3.92)				12.29 (6.80)			
Burt Word Reading Test	E1	2.29 (1.11)	20.57 (5.91)	12.21	.005	.526	29.86 (8.82)	6.67	.025	.377
	E2	4.00 (3.96)	12.71 (9.53)				23.00 (11.96)			
South Australian Spelling Test	E1	1.29 (1.60)	14.29 (4.31)	4.13	.067	.273	20.00 (5.39)	1.38	.265	.111
	E2	2.86 (3.58)	9.43 (7.04)				16.57 (9.33)			
WARL (words correct per minute)	E1	3.43 (1.40)	21.86 (7.60)	3.55	.086	.244	38.57 (10.37)	2.93	.115	.210
	E2	4.71 (3.86)	16.43 (14.16)				32.00 (18.93)			

Note. ES = partial eta squared. A large effect size is evident when partial eta squared is $\geq .138$.

**FIGURE 1**

(Colour online) Martin and Pratt Nonword Reading Test Mean Scores at Pretest, Posttest 1 and Posttest 2.

effect size using partial eta squared was calculated for each measure at posttest 1 and posttest 2. Table 1 presents the results of these analyses.

Results at the end of Phase 1: Group means and treatment effects

Statistically significant, positive treatment effects at the stated alpha level ($p < .01$) were found for two measures: the Martin and Pratt Nonword Reading Test and the Burt Reading Test. The treatment effect sizes for these measures were very large (Martin & Pratt, 2001; partial eta squared = .622; Burt = .526). (Effect sizes calculated using partial eta squared that are larger than .138 are considered to be large; Howell, 2008.)

No significant differences were found between the group means for the other two measures, but treatment effect sizes were substantial. Effect sizes were large for both the South Australian Spelling test (partial eta squared = .273) and large for the WARL (partial eta squared = .244). These findings confirm the findings of the larger randomised control trial (Buckingham et al., 2012).

Results at the end of Phase 2: Group means and treatment effects

No significant differences between group means were found at the stated alpha level ($p < .01$) for any measure at the end of Phase 2; that is, after both groups had each had three terms of the intervention. Mean scores and treatment effects are shown in Table 1.

Figures 1 to 4 show important differences in progress between the two study phases on each of the measures. (Note that these graphs show corrected mean scores generated by the covariance analyses; i.e., allowing for small initial differences between groups at pretest.) Figure 1 shows that test scores on the Martin and Pratt increased strongly for the experimental group in both phases, but there was virtually no growth in scores for the control/comparison groups, with the end result being similar mean scores for E1 and E2 at the end of Phase 2 and an overall treatment effect that is very small (partial eta squared = .035). In other words, the second treatment group (E2), who were formerly the

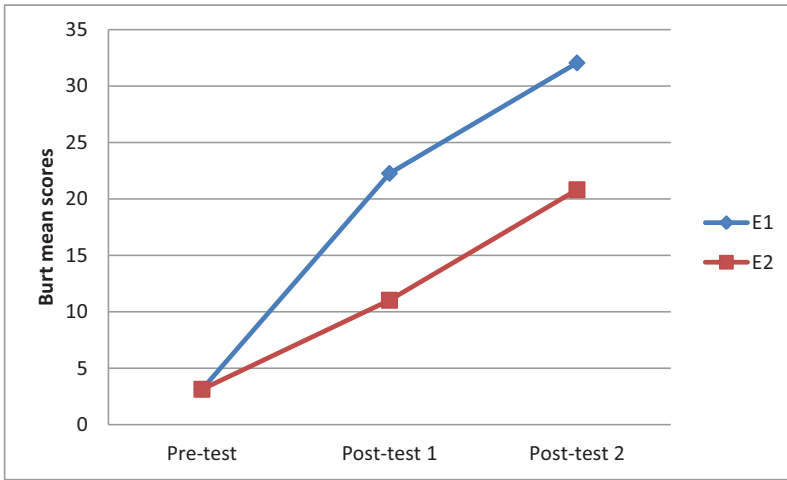


FIGURE 2

(Colour online) Burt Word Reading Test Mean Scores at Pretest, Posttest 1 and Posttest 2.

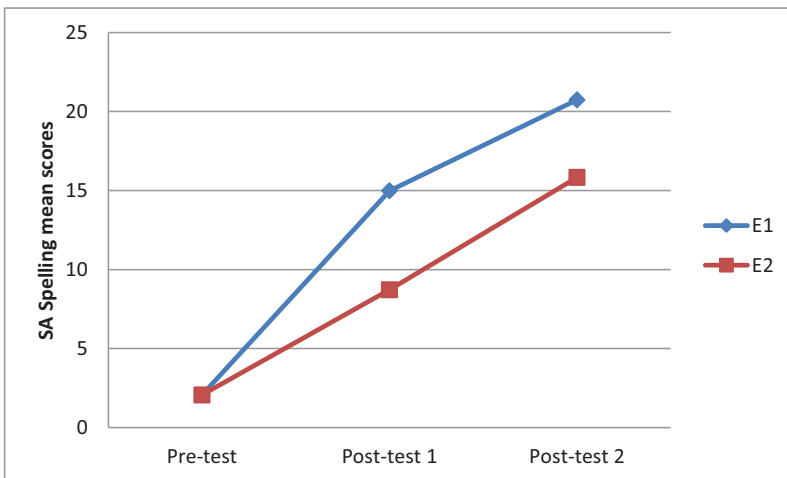
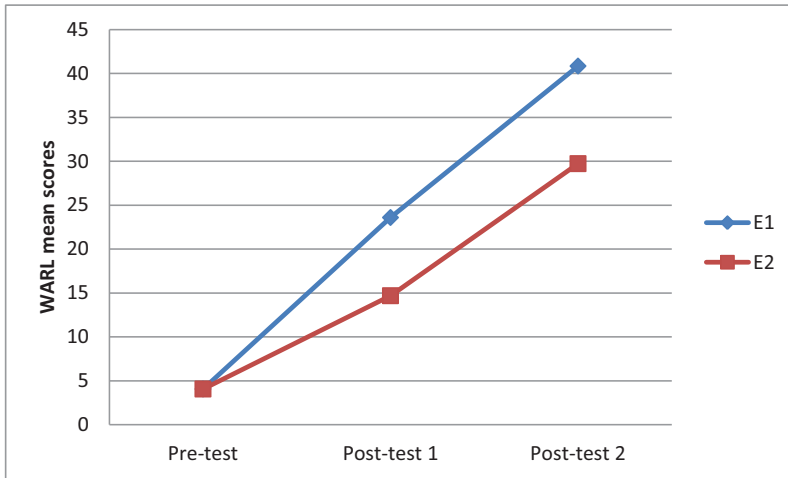


FIGURE 3

(Colour online) South Australian Spelling Test Mean Scores at Pretest, Posttest 1 and Posttest 2.

control group, had almost caught up with the original experimental group (E1) following intervention.

Figures 2, 3 and 4 show different patterns of score growth to Figure 1. At the end of Phase 1 of the study, the experimental group (E1) had made larger gains than the control group in the Burt (Figure 2), the South Australian Spelling Test (Figure 3) and the WARL (Figure 4), as shown by the divergent posttest 1 means, although the differences for the WARL and South Australian Spelling were not statistically significant. At the end of Phase 2, there had been little divergence or narrowing of the scores, as can be seen in the almost

**FIGURE 4**

(Colour online) Wheldall Assessment of Reading Lists (WARL) Mean Scores at Pretest, Posttest 1 and Posttest 2.

parallel slopes between the posttest 1 and posttest 2 means on these three measures. This indicates that the E1 students continued to grow on these measures after they had finished MiniLit.

The effect sizes (reported in Table 1) calculated at the end of Phase 2 for each measure were smaller than at the end of Phase 1, indicating that some of the gap evident at the end of Phase 1, when only one group had had the intervention, was reduced when the second group received the intervention, as we would expect. The effect size for the Martin and Pratt test at the end of Phase 2 was negligible (partial eta squared = .035), in accordance with the nonsignificant difference between the means of the two groups at posttest 2. In other words, at the end of Phase 2, the Phase 2 experimental group (E2) had almost completely closed the gap with the Phase 1 experimental group (E1), because E1 had shown no growth in Martin and Pratt scores in the period after they completed the intervention. For the Burt Word Reading Test and the South Australian Spelling Test, the effect size was reduced at the end of Phase 2, but not substantially (Burt partial eta squared = .377; Spelling = .111), indicating that although the E2 students had made progress it was not sufficient to 'catch up' to E1, because the E1 continued to grow after they completed the intervention. For the WARL, there was little change in the effect size at the end of Phase 2 (partial eta squared = .210). Therefore, with the exception of phonological recoding (as measured by the Martin and Pratt), the earlier Phase 1 intervention had a stronger effect.

Discussion

The efficacy of a small-group (Tier 2) intervention was evaluated in a two-phase, crossover study, implemented over six terms, involving 14 kindergarten students randomly allocated in two groups from matched pairs. The seven students in the Phase 1 control group became the experimental group in Phase 2 of the study. At the end of the six-term study, both groups of students had completed three terms of intervention.

Phase 1 was the first three terms; the experimental group for this phase will be called E1. Phase 2 was the last three terms of the study; the experimental group for this phase (the original, Phase 1 control group) will be called E2.

At the end of Phase 1, the E1 group mean was significantly higher on the Martin and Pratt Nonword Reading Test and the Burt Word Reading Test, and the effect sizes were very large, confirming the powerful effect of the intervention on phonological recoding and word reading. No significant differences were found between groups for the other two measures, but large effect sizes for both the South Australian Spelling Test and the WARL were found.

These effect sizes are in keeping with the observable differences in the growth in raw scores of the two groups in Phase 1, under experimental conditions (see [Table 1](#)). At the end of Phase 1, the E1 group could spell twice as many more words correctly as the control group (13 vs. 6.5) on the South Australian Spelling Test, and the mean WARL scores indicate that the E1 group's fluency grew by 50% more than the control group (18.4 words correct per minute vs. 11.7). These results are informative, and suggest that statistical nonsignificance for these measures might be attributable to the lack of power in the small sample size (Bell, 2011; Slavin & Smith, 2009).

At the end of Phase 2, when both groups had had three terms of intervention, there were no significant differences between groups on any measure (see [Figures 1 to 4](#) and [Table 1](#)). For the Martin and Pratt Nonword Reading Test, the convergence of the two groups was preceded by strong growth of both E1 and E2 while receiving the intervention, and virtually zero growth of these groups in their control/comparison periods. For the other measures, however, both groups grew during their experimental and control/comparison phases, but at different rates. E1 had higher mean scores than E2 for the Burt, South Australian Spelling Test and WARL at the beginning of Phase 2, and E1 continued to make progress on these measures while it was the comparison group. With a lower starting point at the beginning of Phase 2 on these measures, E2 made progress while it received the intervention but did not accelerate enough to catch up to the still-growing E1 group. Treatment effect sizes at the end of Phase 2 confirm this. After both groups had received the intervention, no difference was apparent on the Martin and Pratt — the original control group (E2) had completely closed the gap. For the other measures, however, a treatment effect was still evident, showing that the E1 group was still achieving superior results, even three terms after they completed the intervention.

The small sample size and implementation on a single school site limit generalisation, but the results have several implications specifically regarding literacy practices within the participating school and more generally for the ability of the program to be successful in a regular school setting.

The Martin and Pratt results imply that students were not receiving good phonics instruction in the classroom, as no growth in phonological recoding occurred for either group during their control/comparison period. Classroom literacy instruction for the Phase 1 control group was delivered through a program called L3. This was a new kindergarten program developed by the NSW DEC, and which was being implemented by the school for the first time at the same time as the intervention. Little information about L3 is publicly available so the program can only be described in rudimentary terms. No scope and sequence is available for the content of each lesson; the information for parents states only that 'L3 focuses on providing rich literacy experiences' and that children will 'listen to the reading of stories, poems and songs' (NSW DEC, 2011b, 2011c). There is no mention of phonemic awareness, phonics or any other instruction related to alphabetic knowledge, such as letter-sounds, and if such instruction occurs there is no

publicly available information about the form such instruction takes. Data collection is in the form of continuous text reading levels, writing vocabulary, and a listening test (NSW DEC, 2011a). Phase 2 of the study took place when students were in Year 1. Classroom literacy teaching for the comparison group during this phase was not through a formal program.

The stronger and enduring results for the E1 group on the Burt, the South Australian Spelling Test and the WARL provide some evidence for the need for early intervention. Even though E2 were only in Year 1 when they received the intervention, they did not make as much progress over the course of the intervention as the E1 group, which began the intervention in Term 3 of kindergarten and continued to make progress on all measures except phonological recoding once they returned to the classroom. That the E2 did not catch up to the E1 group in three of the four measures demonstrates that it is very difficult to close reading gaps, even with a highly structured and explicit intervention.

In terms of general implementation, the study offers some useful lessons. First, a good Tier-2 program can be an effective way to provide supplementary reading instruction, but its impact is diminished if high-quality core instruction is not evident in the general education classroom. Given the students' low levels of letter-sound knowledge at pretest and their failure to make any progress in phonological recoding while not in the intervention, it is reasonable to assume that the intervention was, in some cases, providing initial instruction rather than supplementing and reinforcing what was being taught in class. It therefore is to be expected that it would take some time for these students to develop secondary skills such as spelling and fluency.

The objective of this trial was to evaluate the efficacy of the small-group 'MiniLit' program and assess its potential as a Tier-2 intervention in a three-tier RtI model. Although the use of a randomised control trial was necessary to obtain the most scientifically valid evidence, in this case it required a relatively simple two-group comparison. This meant that the other important component of RtI — monitoring and data-based decision making about instruction — could not be included.

Consequently, the results may have been affected by the constraints of the trial methodology, as students could not be moved out of the program. Under standard three-tier RtI conditions, students who are clearly not making progress with Tier-2 instruction would be diverted to Tier-3 intervention. Furthermore, students who made rapid progress would ideally return to regular class instruction when they reached an agreed criterion (say, the 40th percentile in the Martin and Pratt test), allowing them to more quickly generalise to the higher order skills being developed in the classroom. In sum, Tier-2 programs work best when part of a dynamic RtI model, involving careful and timely data-based decisions for each child.

This study supports the efficacy of the small-group program implemented in the trial, particularly in developing the necessary and fundamental skills required to decode words. It has also provided important cautionary information about the limitations of Tier-2 interventions in a regular school setting, when not embedded in an exemplary RtI model. Further study of the intervention when implemented alongside strong and consistent Tier-1 and Tier-3 instruction would be instructive.

Author note

Professor Kevin Wheldall and Dr Robyn Beaman-Wheldall are directors of MultiLit Pty Ltd, the developers of the MiniLit program.

References

- Aikens, N.L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*, 235–251. doi:10.1037/0022-0663.100.2.235
- Australian Curriculum, Assessment and Reporting Authority. (2012). *NAPLAN Achievement in Reading, Writing, Language Conventions and Numeracy: National Report for 2012*. Sydney: ACARA.
- Barnes, G. (n.d.). *Report on the generation of the 2010 Index of Community Socio-Educational Advantage (ICSEA)*. Sydney: Australian Curriculum, Assessment and Reporting Authority. Retrieved from http://www.acara.edu.au/verve/_resources/2010_Index_of_Community_Socio-Educational_Advantage_Generation_Report.pdf
- Bell, J.F. (2011). The small-study effect in educational trials. *Effective Education, 3*, 35–48. doi:10.1080/19415532.2011.610642
- Brooks, G. (2007). *What works for pupils with literacy difficulties? The effectiveness of intervention schemes* (3rd ed.). London, UK: Department for Children, Schools and Families. Retrieved from https://www.education.gov.uk/publications/eOrderingDownload/pri_lit_what_works0068807.pdf
- Buckingham, J., Wheldall, K., & Beaman, R. (2012). A randomised control trial of a Tier-2 small-group intervention ('MiniLit') for young struggling readers. *Australian Journal of Learning Difficulties, 17*, 79–99. doi:10.1080/19404158.2012.717537
- Carnine, D.W., Silbert, J., Kame'enui, E.J., & Tarver, S.G. (2010). *Direct instruction reading* (5th ed.). Upper Saddle River, NJ: Pearson.
- Carnine, D.W., Silbert, J., Kame'enui, E.J., Tarver, S.G., & Jungjohann, K. (2006). *Teaching struggling and at-risk readers: A direct instruction approach*. Upper Saddle River, NJ: Pearson.
- Department of Education, Science and Training (DEST). (2005). *National Inquiry into the Teaching of Literacy (NITL). Teaching reading: Report and recommendations*. Barton, ACT: Commonwealth of Australia.
- Dodd, B., & Carr, A. (2003). Young children's letter-sound knowledge. *Language, Speech, and Hearing Services in Schools, 34*, 128–137. doi:10.1044/0161-1461(2003/011)
- Eamon, M.K. (2005). Social-demographic, school, neighborhood, and parenting influences on the academic achievement of Latino young adolescents. *Journal of Youth and Adolescence, 34*, 163–174. doi:10.1007/s10964-005-3214-x
- Ellis, L.A., Wheldall, K., & Beaman, R. (2007). The research locus and conceptual basis for MULTILIT: Why we do what we do. *Australian Journal of Learning Disabilities, 12*, 61–65. doi:10.1080/19404150709546831
- Fuchs, D., & Fuchs, L.S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 93–99. doi:10.1598/RRQ.41.1.4
- Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W.D. (2009). *Assisting students struggling with reading: Response to intervention (RTI) and multi-tier interventions in the primary grades. A practice guide* (NCEE 2009–4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wvc/publications/practiceguides/>
- Gersten, R., & Dimino, J.A. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41*, 99–108. doi:10.1598/RRQ.41.1.5
- Gilmore, A., Croft, C., & Reid, N. (1981). *Burt Word Reading Test – New Zealand revision*. Wellington, NZ: New Zealand Council for Educational Research.
- Hart, B., & Risley, T.R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*(1). Retrieved from <http://www.aft.org/pdfs/americaneducator/spring2003/TheEarlyCatastrophe.pdf>
- Hatcher, P.J., Hulme, C., Miles, J.N.V., Carroll, J.M., Hatcher, J., Gibbs, S., ... Snowling, M.J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised control trial. *Journal of Child Psychology and Psychiatry, 47*, 820–827. doi:10.1111/j.1469-7610.2005.01559.x

- Hempenstall, K. (2009). Research-driven reading assessment: Drilling to the core. *Australian Journal of Learning Difficulties*, 14, 17–52. doi:10.1080/19404150902783419
- Howell, D. (2008). *Fundamental statistics for the behavioral sciences* (6th ed.). Belmont, CA: Thomson Wadsworth.
- IBM Corp. (2012). IBM SPSS Statistics for Windows, Version 21.0 [Computer software]. Armonk, NY: Author.
- Louden, B., Chan, L.K.S., Elkins, J., Greaves, D., House, H., Milton, M., . . . van Kraayenoord, C. (2000). *Mapping the territory: Primary students with learning difficulties: Literacy and numeracy*. Canberra, Australia: Commonwealth Department of Education, Training and Youth Affairs.
- Marks, G. (2009). Accounting for school-sector differences in university entrance performance. *Australian Journal of Education*, 53, 19–38. Retrieved from doi:10.1177/000494410905300103
- Martin, F., & Pratt, C. (2001). *The Martin and Pratt Nonword Reading Test*. Melbourne, Australia: ACER.
- Merrett, F., & Wheldall, K. (1990). *Positive teaching in the primary school*. London, UK: Chapman.
- MultiLit. (2011). *MiniLit early literacy intervention program*. Sydney, Australia: MultiLit Pty Ltd.
- National Institute of Child Health and Human Development (NICHD). (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office. Retrieved from <http://www.nationalreadingpanel.org/publications/publications.htm>
- New South Wales Department of Education and Communities (NSW DEC). (2011a). *Language, learning and literacy: Guidelines*. Sydney, Australia: NSW Curriculum and Innovation Centre.
- New South Wales Department of Education and Communities (NSW DEC). (2011b). *Language, learning and literacy: Information for parents*. Sydney, Australia: NSW Curriculum and Innovation Centre. Retrieved February 12, 2012, from <http://www.curriculumsupport.education.nsw.gov.au/beststart/lll/general/index.htm>
- New South Wales Department of Education and Communities (NSW DEC). (2011c). *Language, learning and literacy: Information for the kindergarten team*. Sydney, Australia: NSW Curriculum and Innovation Centre. Retrieved February 12, 2012, from <http://www.curriculumsupport.education.nsw.gov.au/beststart/lll/general/index.htm>
- New South Wales Department of Education and Communities (NSW DEC). (2012). *Reading Recovery: A research-based early intervention program. What happens in a Reading Recovery lesson?* Retrieved June 26, 2012, from http://www.curriculumsupport.education.nsw.gov.au/earlyyears/reading_recovery/lesson.htm
- Reading Recovery Council of North America. (2007). *Phonemic awareness*. Retrieved from http://www.readingrecovery.org/reading_recovery/federal/Essential/phonemic.asp
- Reynolds, M., & Wheldall, K. (2007). Reading Recovery 20 years down the track: Looking forward, looking back. *International Journal of Disability, Development and Education*, 54, 199–223. doi:10.1080/10349120701330503
- Reynolds, M., Wheldall, K., & Madelaine, A. (2007). Meeting Initial Needs In Literacy (MINILIT): Why we need it, how it works, and the results of pilot studies. *Australasian Journal of Special Education*, 31, 147–158. doi:10.1080/10300110701716188
- Reynolds, M., Wheldall, K., & Madelaine, A. (2009). Building the WARL: The development of the Wheldall Assessment of Reading Lists, a curriculum-based measure designed to identify young struggling readers and monitor their progress. *Australian Journal of Learning Difficulties*, 14, 89–111. doi:10.1080/19404150902783443
- Reynolds, M., Wheldall, K., & Madelaine, A. (2010). An experimental evaluation of an intervention for young struggling readers in Year One. *Special Education Perspectives*, 19(2), 35–57.
- Reynolds, M., Wheldall, K., & Madelaine, A. (2011). What recent reviews tell us about the efficacy of reading interventions for struggling readers in the early years of schooling. *International Journal of Disability, Development and Education*, 58, 257–286. doi:10.1080/1034912X.2011.598406
- Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. Retrieved from <https://www.education.gov.uk/publications/standard/publicationdetail/page1/DFES-0201-2006>

- Rothman, S. (2002). *Achievement in literacy and numeracy by Australian 14-year-olds, 1975–1998* (Longitudinal Surveys of Australian Youth Research Report no. 29). Melbourne: Australian Council for Educational Research. Retrieved from http://research.acer.edu.au/lsay_research/33/
- Rothman, S., & McMillan, J. (2003). *Influences on achievement in literacy and numeracy* (Longitudinal Surveys of Australian Youth Research Report no. 36). Melbourne: Australian Council for Educational Research. Retrieved from http://research.acer.edu.au/lsay_research/40/
- Slavin, R.E., Lake, C., Davis, S., & Madden, N.A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, 6, 1–26. doi:10.1016/j.edurev.2010.07.002
- Slavin, R.E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31, 500–506. doi:10.3102/0162373709352369
- Symons, A., & Greaves, D. (2006). The use of THRASS program with younger children with literacy difficulties. *Australian Journal of Dyslexia and Specific Learning Difficulties*, 1 (July), 31–36.
- Thomson, S., de Bortoli, L., Nicholas, M., Hillman, K., & Buckley, S. (2010). *Challenges for Australian education: Results from PISA 2009: The PISA 2009 assessment of students' reading, mathematical and scientific literacy*. Melbourne: Australian Council for Educational Research. Retrieved from <http://research.acer.edu.au/ozpisa/9/>
- The THRASS[®] Institute. (2014). For parents THRASS[®]. Retrieved from <http://www.thrass.com.au/for-parents/>
- Torgesen, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, 15, 55–64. doi:10.1207/SLDRP1501_6
- van Kraayenoord, C.E. (2010). Response to intervention: New ways and wariness. *Reading Research Quarterly*, 45, 363–376. doi:10.1598/RRQ.45.3.5
- Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36, 541–561. Retrieved from <http://www.nasponline.org/publications/spr/index.aspx?vol=36&issue=4>
- Westwood, P. (2005). *Spelling: Approaches to teaching and assessment* (2nd ed.). Camberwell: Australian Council for Educational Research Press.
- What Works Clearinghouse (WWC). (2007). *SpellRead[™]* (WWC Intervention Report, July 9, 2007). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Kaplan_Spellread_070907.pdf
- What Works Clearinghouse (WWC). (2008a). *Early Intervention in Reading[®]* (WWC Intervention Report, November 2008). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_eir_112508.pdf
- What Works Clearinghouse (WWC). (2008b). *Lindamood Phonemic Sequencing[®]* (WWC Intervention Report, December 2008). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_lindamood_121608.pdf
- What Works Clearinghouse (WWC). (2009). *Success For All[®]* (WWC Intervention Report, August 2009). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_sfa_081109.pdf
- What Works Clearinghouse (WWC). (2010). *Lindamood Phoneme Sequencing[®]* (WWC Intervention Report, March 2010). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_lindamood_031610.pdf
- What Works Clearinghouse (WWC). (2012). *Peer-Assisted Learning Strategies* (WWC Intervention Report, May 2012). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_pals_050112.pdf
- Wheldall, K. (2009). Effective instruction for socially disadvantaged low-progress readers: The Schoolwise Program. *Australian Journal of Learning Difficulties*, 14, 151–170. doi:10.1080/19404150903264294
- Wheldall, K., Beaman, R., Madelaine, A., & McMurtry, S. (2012). *Evaluations of the efficacy of MultiLit and MiniLit programs provided by the Exodus Foundation, 2009–2011*. Unpublished research report submitted to the Exodus Foundation. Sydney, Australia: MultiLit Research Unit, Macquarie University.

Wyatt-Smith, C., Elkins, J., Colbert, P., Gunn, S., & Muspratt, S. (2007). *Changing the nature of support provision: Students with learning difficulties: Interventions in literacy and numeracy project (InLaN). Volume 1: Design, methodology and findings*. Canberra, Australia: Commonwealth Department of Education, Science and Training.

Appendix

Source. Buckingham, J., Wheldall, K., & Beaman, R. (2012). A randomised control trial of a Tier-2 small-group intervention ('MiniLit') for young struggling readers. *Australian Journal of Learning Difficulties*, 17, 79–99. doi:10.1080/19404158.2012.717537

Measures

Burt Word Reading Test (Gilmore et al., 1981). This test measures single word recognition using a list of 110 words that increase in difficulty. The maximum reading age achievable on the Burt is about 13 years. The Burt has high test–retest reliability (> 0.95), high internal consistency (> 0.96) and high criterion validity (correlations of 0.90–0.98 between the Burt Word Reading Test and the Schonell Graded Word Reading Test (Schonell, 1955) and the Oral Word Reading Test (Fieldhouse, 1952, as cited in Gilmore et al., 1981).

South Australian Spelling Test (Westwood, 2005). This test provides a spelling age for children in the age range 6 years to over 15 years. It can be administered individually or in groups. The test manual reports good internal reliability with a test–retest reliability coefficient of 0.96 for most year groups. Alternative forms reliability ranges from 0.89 to 0.94 depending on age level (Westwood, 2005).

Martin and Pratt Nonword Reading Test, Form A (Martin & Pratt, 2001). This test measures phonological recoding ability in students aged from 6 to 16 years, using pseudowords of increasing difficulty (Martin & Pratt, 2001). The test has a high test–retest reliability coefficient of 0.96, high alternative-forms reliability coefficients of 0.92–0.96 and a high internal consistency reliability coefficient of 0.96 (Martin & Pratt, 2001). Good criterion-related validity is indicated through positive correlations between the Martin and Pratt and the WRMT-R Word Attack (Woodcock, 1987, as cited in Martin & Pratt, 2001; 0.89), Coltheart and Leahy Nonword reading lists (Coltheart & Leahy, 1996; 0.93) and the Neale Analysis of Reading Ability (Neale, 1988, as cited in Martin & Pratt, 2001; 0.78–0.88). Nonword tests are an important measure of early reading progress as they avoid the possibility of students reading words from memory (Hempenstall, 2009).

Wheldall Assessment of Reading Lists (WARL; Reynolds, Wheldall, & Madelaine, 2009). The WARL is a curriculum-based measure of word identification fluency for young students. The test consists of parallel lists of 100 high-frequency words from children's texts and storybooks. The student is presented with the list on a page and instructed to read the words aloud quickly and carefully. They are asked to stop after one minute. The score is the number of words read correctly per minute, averaged over three parallel lists. The WARL has been found to be highly reliable, with reliability coefficients for parallel forms between 0.85 and 0.94 (Reynolds et al., 2009). The validity of the measure has been demonstrated through high correlations between the WARL and the Burt Word Reading Test ($r = 0.79$) and the TOWRE Sight Words Test (Torgesen, Wagner, & Rashotte, 1999; $r = 0.95$; Reynolds et al., 2009).

MiniLit Placement Test (MultiLit, 2011). The MiniLit Placement Test assesses students' phonic word attack knowledge and abilities, namely, letter-sound correspondences and reading of words containing specific letter-sound correspondences. The test includes only real words, but many are not frequently used, and therefore are very unlikely to be recognised as sight words by young students.