cogent
education

EDUCATIONAL PSYCHOLOGY & COUNSELLING | RESEARCH ARTICLE

Evaluation of a two-phase experimental study of a small group ("MultiLit") reading intervention for older low-progress readers

Jennifer Buckingham[1,2*], Robyn Beaman-Wheldall[1] and Kevin Wheldall[1]

Abstract: The study reported here examined the efficacy of a small group (Tier 2 in a three-tier Response to Intervention model) literacy intervention for older low-progress readers (in Years 3–6). This article focuses on the second phase of a two-phase, crossover randomized control trial involving 26 students. In Phase 1, the experimental group (E1) received the 1 h literacy intervention daily for three school terms. The control group received regular classroom instruction. In Phase 2, the original control group received the intervention (E2). At the end of Phase 1, there was a statistically significant difference between groups and a large treatment effect on one of five measures—the Martin and Pratt Non-word Reading Test of phonological recoding. At the end of Phase 2, the large effect on phonological recoding was confirmed for the second experimental group, and there were also statistically significant differences with moderate or large effect sizes on four other measures—single word reading, fluency, passage reading accuracy, and comprehension.

1. Introduction

A large number of children in developed, English-speaking countries struggle to learn to read at even a functional level. The Progress in Reading Literacy Survey (PIRLS) is an international assessment of literacy of Year 4 students. In PIRLS 2011, the proportion of students who achieved at the minimum literacy benchmark or below ranged from 16% in the United States and Canada to 24% in Australia and

ABOUT THE AUTHORS

Jennifer Buckingham was a PhD candidate at Macquarie University and is a research fellow at The Centre for Independent Studies.

Kevin Wheldall AM, Emeritus Professor, of Macquarie University is the director of the MultiLit Research Unit and chairman of MultiLit Pty Ltd.

Robyn Wheldall is an honorary fellow of Macquarie University, deputy director of the MultiLit Research Unit, and a director of MultiLit Pty Ltd.

PUBLIC INTEREST STATEMENT

This study contributes to the evidence that effective reading instruction based on scientific research can accelerate the reading skills of low-achieving students from socially disadvantaged backgrounds. It also shows that older students (Years 3–6) respond well to comprehensive reading interventions designed for their age group. After students had three terms of a literacy intervention, their reading skills had increased significantly on five measures—decoding (phonics), word reading, fluency, accuracy, and comprehension.

25% in New Zealand (Thomson et al., 2012). Another international survey, Program for International Student Assessment (PISA), tests student literacy at age 15 years. In PISA 2012, the proportion of students achieving at the lowest literacy level or below ranged from 10.9% in Canada to 14.2% in Australia and 16.7% in the United Kingdom (Organisation for Economic Cooperation and Development, 2013).

Students from socioeconomically disadvantaged backgrounds are more likely to have low literacy achievement (Australian Curriculum, Assessment and Reporting Authority, 2013; Thomson, De Bortoli, & Buckley, 2013). The quality of reading instruction and intervention is a strong mediating factor in the literacy gap associated with socioeconomic status (Buckingham, Beaman, & Wheldall, 2013; Buckingham, Wheldall, & Beaman-Wheldall, 2013). Large-scale surveys of literacy research in the United States (National Institute of Child Health and Human Development, 2000), Australia (Department of Education, Science and Training [DEST], 2005), and England (Rose, 2006) have concluded that the best scientific evidence supports the finding that effective reading instruction has five "pillars": phonemic awareness, phonics, fluency, vocabulary, and comprehension. Each of these elements is necessary for the successful, early acquisition of reading skills and general literacy development. They are essential components of both effective classroom teaching and reading interventions for struggling readers.

The importance of early intervention for struggling readers cannot be overstated and is well recognized (Feinstein, 2007; Reynolds, Wheldall, & Madelaine, 2011; Stanovich, 1986; Torgesen, 2005). Many schools have at least one formal early reading intervention program, such as Reading Recovery, which targets Year 1 students (Clay, 1993; New South Wales Department of Education and Communities, 2013; Reading Recovery Council of North America, 2013; Tanner et al., 2011). Yet the statistics presented above indicate that at Year 4, substantial numbers of students are still in need of literacy support, whether because they missed out on early reading intervention, the intervention was ineffective, their reading difficulties were identified later, or they are students who require ongoing literacy support. There is therefore a need for literacy interventions aimed at older (Year 3 and above), low-progress readers. Low-progress readers are students whose literacy skills are well below those of their classmates'—around the lowest 25% of their age cohort (Pogorzelski & Wheldall, 2005).

The MultiLit reading intervention was designed specifically for older, low-progress readers. It exists in a number of formats. The MultiLit Reading Tutor Program is a 30–40 min a day, one-to-one format program, which is implemented in schools and at the MultiLit Literacy Centre. The MultiLit "Schoolwise" Program is conducted in tutorial centers which students attend for 3 h a day, five days a week. Students work in groups and individually with teachers. Evaluation of these programs has shown them to be highly effective (Wheldall, 2009; Wheldall & Beaman, 2000, 2010).

The growing body of research supporting a Response to Intervention (RtI) approach to teaching and assessment indicates that there is a missing step in reading intervention offered in schools. In an RtI model, students are provided with increasingly intensive "tiers" or levels of instruction, depending on their reading progress. In a three-tier RtI model, Tier 1 is whole class instruction, Tier 2 is small group instruction, and Tier 3 is individual instruction. Students who are not making good progress in reading in class are provided with supplementary instruction in a small group. Students who are still struggling to make reading progress in the small group are provided with specialist one-to-one instruction (Gersten et al., 2009). A review of reading interventions by Slavin, Lake, Davis, and Madden (2011) found that small group instruction with a strong phonics emphasis can be beneficial to students whose reading difficulties are not extreme. The RtI approach is therefore both effective and cost-effective. Small group interventions allow more students to be given extra reading support, reserving the most intensive (and expensive) one-to-one instruction for the few students with serious reading difficulties.

The MultiLit small group program was developed as a Tier 2 reading intervention for students in Year 3 and above. A randomised control trial of the small group MultiLit program over three terms is described in Buckingham, Beaman, and Wheldall (2012). Classroom teachers identified the lowest 20 readers in each year (a total of 80 students), who were then given screening tests by trained

testers. The 12 students with the lowest screening test scores from each year were selected for participation in the trial and randomly assigned into either the experimental or the control group. The control group had their usual classroom literacy instruction, while the experimental group attended MultiLit lessons for 1 h a day, four days a week, for three terms. All students in the control and experimental group were given a battery of tests pre-intervention, after two terms of intervention, and after three terms of intervention, and the results were compared.

At the end of three terms, the initial trial showed strong, statistically significant, positive results in phonological recoding only, with a very large treatment effect size (partial η^2 = .520). There were small treatment effects on single word reading (.057) and spelling (.037). Treatment fidelity had not reached an optimal level until the 14th week of the intervention and so on this basis, the school and researchers decided a second implementation would be worthwhile. In the second phase of the intervention, the original experimental group returned to their usual classroom literacy lessons and the control group replaced them in the small group MultiLit program, becoming a new experimental group. As there was some attrition of students from the school after the initial trial, 12 of whom were Year 6 students leaving for high school and another five of whom were students who moved away during the trial, the sample for the two-phase crossover study is smaller. This article focuses on the subset of students who participated in both phases of the trial. It compares the findings of the first and second implementations of the MultiLit intervention, evaluated as a two-phase, crossover study over six school terms.

2. Method

2.1. Participants
Participants were 26 students from Years 3 to 6 in a public primary school with a high proportion of socioeconomically disadvantaged students, located in a large New South Wales regional town. Participants in the two-phase, crossover study are a subset of the participants in the initial (three-term) randomised control trial (each school term is approximately 10 weeks). There were 30 participants in the initial three-term trial—15 in the first experimental group and 15 in the first control group. Several students left the school during the second three-term phase of the study—one from the first experimental group (E1) and two from the second experimental group (E2). In order to maintain comparability of the two groups, the data from their matched pairs have also been excluded. As a result of these departures and exclusions, a total of 26 students in two randomised, matched groups participated in the six-term, two phase crossover study.

2.2. Procedure
In Phase 1 of the study, students in the first experimental group (E1) were withdrawn from class to participate in the group MultiLit program for 1 h a day, four days a week, for three terms (27 weeks) during class literacy time. Students in the control group remained in their usual classrooms (detailed in Buckingham et al., 2012). In Phase 2 of the study, which took place over the next three terms, the first control group became the second experimental group (E2) and participated in the group MultiLit program. The first experimental group returned to their usual classroom literacy lessons and became the Phase 2 control group. All participants were given a battery of tests before commencement of the intervention, again at the end of Phase 1 when the first experimental group (E1) had completed the intervention, and a third time at the end of Phase 2, when the second experimental group (E2) had completed the intervention. At the end of the study, both groups had participated in the group MultiLit program for three terms.

2.3. Measures
The test battery consisted of five measures—the Burt Word Reading Test (Gilmore, Croft, & Reid, 1981), Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001), Wheldall Assessment of Reading Passages (Wheldall & Madelaine, 2006), and Neale Analysis of Reading Ability (Accuracy and Comprehension) (Neale, 1999). Descriptions of these tests are provided in Buckingham et al. (2012) and brief descriptions are appended (see Appendix 1).

2.4. Intervention

The MultiLit program components are described in Buckingham et al. (2012). They were Word Attack (Accuracy), Word Attack (Fluency), Sight Words, and Reinforced Reading. The content and delivery of each component of the program were basically the same in Phases 1 and 2 of the study. However, some small changes in program implementation took place in Phase 2, including smaller group size and changes to the placement procedure. A brief description of the program is provided in Appendix 2.

2.4.1. Group size

In Phase 1, students in the MultiLit program were in instructional groups of six students for the first two terms, reduced to four students for the third term when the Year 6 students had left the school (Year 6 is the final year of primary school in NSW). In Phase 2, students in the MultiLit program were in instructional groups of four students for all three terms of the intervention.

2.4.2. Placement

In Phase 1, before beginning the MultiLit intervention, students were given the MultiLit Placement Test in order to determine the appropriate starting level of Word Attack (Accuracy) instruction. Students were allocated to instructional groups according to their starting level. The MultiLit Placement Test procedure in Phase 1 was to start instruction for each group at the lowest level required by any one group member, and then to proceed with instruction through each consecutive level. In Phase 1, almost all students were placed at the lowest level to start the program. It became apparent that this was too low for some of the older students in particular—their knowledge of phonics was uneven rather than consistently low.

In Phase 2, the placement procedure was changed to take this into account. Phase 2 MultiLit students were taught only the individual levels of the Word Attack program each group member had failed. Consecutive instruction of each level continued from the level failed by *all* group members. This change in procedure allowed some groups to quickly progress through the most basic Word Attack levels and move to their substantive instructional level. As a consequence, Phase 2 MultiLit students completed the Word Attack components more quickly and moved onto the additional program components developing fluency and comprehension.

2.5. Analysis

To compare the progress made by the experimental and their respective control or comparison groups, analyses of covariance (ANCOVA) were employed for each measure at post-test 1 (after three terms), with pre-test scores as the covariate in each analysis (with some exceptions detailed below). 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). In the second phase of the study, culminating in post-test 2, repeated measures *t*-tests were employed to evaluate the gains made by each group separately, again employing an alpha level of 1% ($p < .01$).

Treatment effects were also calculated for each measure in each phase of the study, using Cohen's *d*.

3. Results

The Phase 1 experimental group/Phase 2 control 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 Phase 2 experimental group (E2) at pre-test, post-test 1 (after three terms), and post-test 2 (after six terms) are shown in Table 1. Table 1 shows that the E1 group means were slightly lower than those for the E2 group's at pre-test on all measures but none of these differences was statistically significant. (The subsequent ANCOVA take these initial small group differences into account.)

Table 1. Means and standard deviations (raw scores) of Phase 1 experimental (E1) and Phase 2 experimental (E2) groups at pre-test, post-test 1 (after three terms) and post-test 2 (after six terms), results of ANCOVA and effect sizes (Cohen's d) for Phase 1 and repeated measures t-tests and effect sizes for both groups separately for Phase 2

Measure	Group	Pre-test	Post-1	F*	p	ES	Post-2	t	p	ES
	(n = 13)	Mean (SD)	Mean (SD)				Mean (SD)			
Martin and Pratt Nonword Test	E1	9.23 (8.47)	18.62 (10.60)	22.46	.000	.94	16.23 (10.02)	−1.87	NS	−.74
	E2	10.23 (4.44)	10.92 (4.54)				18.92 (3.93)	7.34	.000	2.54
Burt Word Reading Test	E1	27.62 (11.18)	36.31 (16.82)	.54	.468	.15	37.54 (14.65)	.81	NS	.35
	E2	30.92 (9.19)	38.46 (10.36)				45.92 (8.09)	6.42	.000	3.65
WARP (words correct per minute, wcpm)	E1	36.00 (24.90)	58.85 (35.87)	.34	.568	.22	62.15 (36.18)	2.35	NS	.92
	E2	43.77 (28.48)	66.62 (33.20)				97.92 (35.85)	9.47	.000	3.80
Neale accuracy	E1		25.85 (13.50)	.13	.725	.23	26.85 (14.14)	.816	NS	.32
	E2		28.54 (10.17)				37.54 (8.08)	4.11	.001	1.66
Neale comprehension	E1		10.46 (6.28)	1.75	.199	.07	9.92 (5.45)	−.87	NS	−.37
	E2		10.08 (3.59)				13.69 (5.09)	3.54	.01	1.27

*Pre-test scores are covariates for Phase 1 (Pre-test/Post-1) ANCOVA, except Neale measures which use Burt Pretest scores.

ANCOVA were conducted on the scores for each measure separately at post-test 1 with pre-test scores as the covariate for all measures except for the Neale Analysis of Reading Ability Accuracy and Comprehension components, as this test was not administered in the pre-test battery. Pre-test scores for the Burt (which correlates highly with the Neale) were used as the covariate for the two Neale measures.

For the second phase of the study, analyses of gains were evaluated using repeated measures *t*-tests, for the E1 and E2 groups separately.

Cohen's *d* was calculated for each measure at post-test 1 and post-test 2 to determine the size of the treatment effect. Results of all of these analyses are summarized in Table 1.

3.1. 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 one measure—the Martin and Pratt Nonword Reading Test. The treatment effect size for this measure was large (*d* = .94). No significant differences were found between the group means for the other measures and there were negligible treatment effects. These analyses confirmed for the reduced groups the results reported for the original group by Buckingham et al. (2012).

3.2 Results at the end of Phase 2—group means and treatment effects

For Phase 2, the two groups, E1 and E2, were analysed separately. The E1 group (the original group who received treatment in Phase 1 and who were now returned to regular classroom instruction) failed to make further gains that were statistically significant on any measure at the stated alpha level. Effect sizes for all but one measure, the Wheldall Assessment of Reading Passages (WARP), were small or, in two cases, indicated *a loss i.e. mean scores were lower at post-test 2 compared with post-test 1*. In the case of the WARP, the effect size was large (*d* = .92) but it should be emphasized that the gain was not statistically significant.

For the E2 group, however, who received treatment (MultiLit instruction) during Phase 2 of the study, the results were very different. Statistically significant gains at the stated alpha level were made on all five measures. Moreover, the effect sizes were all large, ranging from 1.27 to 3.80 (see Table 1).

In summary, at the end of Phase 2, the E2 group had made significant gains on all measures, with large effect sizes, whereas the E1 group made no significant gains.

4. Discussion

In this two-phase crossover study of a small group literacy intervention for older, low-progress readers (MultiLit), the second experimental group to receive the intervention appeared to display a much stronger response than the first experimental group. At the end of Phase 1, there was a statistically significant, positive effect of the intervention on only one of the five measures—the Martin and Pratt Nonword Reading Test. At the end of Phase 2, analyses of gains for the two groups separately showed statistically significant, positive gains for the second experimental group on all five measures (with large effect sizes) whereas the original experimental group (E1) (now experiencing control conditions) made no statistically significant gains.

It should be noted that a powerful effect of the program on phonological recoding was demonstrable in both phases of the intervention. At the commencement of the study, the two groups were very similar in terms of pre-test scores on the Martin and Pratt. At post-test 1, there was a statistically significant difference between the means and evidence for a large effect size (d = .94). In the second phase of the study, the initial control group, now experimental group 2, received the program and made large gains, as a result, reaching the gains made by the original experimental group. This provides strong evidence for the particular efficacy of the program on phonological recoding, arguably the most important skill to be mastered by older low-progress readers.

In Phase 1 of the study, there was a strong emphasis on phonics. All groups started at the lowest level of the Word Attack program. Approximately, half of each lesson was spent on phonics and this component of the program was the earliest to achieve treatment fidelity (Week 10). The skills learnt in the Word Attack component on the program most closely relate to those measured in the Martin and Pratt Nonword Reading Test; this is likely to explain the strong results on this measure of phonological recoding.

The results of Phase 2 are similarly strong for phonological recoding, but were also strong and significant for single word reading (Burt), fluency (WARP), passage reading accuracy (Neale), and comprehension (Neale). The powerful results in Phase 2 might be attributed to a number of factors.

Group size was lower in Phase 2 than in most of Phase 1. For two terms of the three-term Phase 1 intervention, the MultiLit students were in groups of six, decreasing to groups of four when the Year 6 students left the school. In Phase 2, MultiLit students were in groups of four for the entire intervention. In the smaller groups, testing time was shorter (allowing increased teaching time), there was a narrower range of ability levels in each group and, perhaps most importantly, the amount of time for each student to do reinforced reading was greater. All of these had a potentially positive impact on the program's efficacy.

Changes to MultiLit program implementation in Phase 2 may also have influenced the results. As noted in the method section, the placement of groups on the Word Attack component of the program was revised in Phase 2 so that all students would reach their substantive instructional level more quickly. This change in the placement procedure allowed students to progress through and complete the Word Attack component earlier in the intervention. Students were then able to expend more time on activities designed to develop the higher order skills of fluency and comprehension. The Phase 2 results provide evidence of the positive effect of this change in instructional focus.

There was also a noticeable difference in the behaviour of the Phase 1 and Phase 2 experimental groups. It is not clear whether this was a cohort effect, a function of the changes in group size and program implementation, or perhaps a third factor—instructional quality. Although treatment fidelity data was not collected in the Phase 2 implementation, it is likely that the MultiLit instructors were more proficient in teaching the program in Phase 2, and this positively influenced both behaviour and learning.

cogent •• education

The lack of significant differences between the two groups at the end of Phase 1 in all measures except the Martin and Pratt was not because the experimental group had not made progress over the period of the intervention, rather it was because both groups made similar progress in this period. This suggests that all students had benefitted from another, unmeasured factor. During Phase 1, the school was undergoing substantial reforms to its teaching processes in all classrooms, including adopting explicit teaching methods in literacy (but without an increase in phonics instruction). This may have contributed to the control group's improved performance. Furthermore, during the trial all classrooms had fewer low-progress readers (half of whom were in MultiLit for 1 h each day during literacy time), which may have positively affected the instruction received by control students.

This two-phase, crossover study of the small group MultiLit program had some limitations. The final sample size was not large (26 students) and confined to one school, and the measures used do not cover the full range of literacy skills. Nonetheless, the study has provided evidence of the efficacy of the intervention, showing statistically significant and educationally important increases in both first- and second-order reading skills, especially in the second implementation. These results contribute to the research evidence on reading interventions for older students.

The study also reinforces the necessity of good experimental trials. The randomised control trial methodology is an important feature of this study. Without a control group for comparison, the Phase 1 results would have appeared to be stronger than they really were. It also demonstrates the benefits of trialling new programs over a realistic period of time. Even though the MultiLit program was based on the best available research and had good evidence of efficacy in other formats and in other settings, the initial results of the small-group school program trialled in this study were strong only in phonological recoding. It was not until the second phase that highly positive results were yielded, indicating that abandoning programs too early can be imprudent. Ethical judgements are required—if the Phase 1 implementation had shown an adverse effect on students' reading skills it would not have been repeated—but it may be too much to expect immediate strong benefits of even the most well-designed program.

Further research on this program to support the results would be ideal, but this study adds to the evidence that a comprehensive literacy intervention which explicitly develops the five essential skills of reading can markedly improve literacy skills among older low-progress readers.

Author details
Jennifer Buckingham[1,2]
E-mail: jbuckingham@cis.org.au
Robyn Beaman-Wheldall[1]
E-mail: robyn.beaman@pecas.com.au
Kevin Wheldall[1]
E-mail: kevin.wheldall@pecas.com.au
[1] Macquarie University Special Education Centre, Macquarie University, Sydney, Australia.
[2] Centre for Independent Studies, P.O. Box 92, St. Leonards, Sydney, New South Wales 1590 since December 1998.

Citation information
Cite this article as: Evaluation of a two-phase experimental study of a small group ("MultiLit") reading intervention for older low-progress readers, J. Buckingham, R. Beaman-Wheldall & K. Wheldall, *Cogent Education* (2014), 1: 962786.

References
Australian Curriculum, Assessment and Reporting Authority. (2013). *National assessment program for literacy and numeracy—Achievement in reading, persuasive writing, language conventions and numeracy* (National Report for 2013). Sydney: Author.

Buckingham, J., Beaman, R., & Wheldall, K. (2012). A randomised control trial of a MultiLit small group intervention for older low-progress readers. *Effective Education, 4,* 1–26. doi:10.1080/19415532.2012.722038

Buckingham, J., Beaman, R., & Wheldall, K. (2013). Why poor children become poor readers: The early years. *Educational Review.* doi:10.1080/00131911.2013.795129

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education, 57,* 190–213. http://dx.doi.org/10.1177/0004944113495500

Clay, M. M. (1993). *Reading recovery: A guidebook for teachers in training.* Portsmouth, NH: Heinemann.

Department of Education, Science and Training. (2005). *National inquiry into the teaching of literacy (NITL). Teaching reading: Report and recommendations.* Retrieved December 21, 2007, from www.dest.gov.au/nitl/documents/report_recommendations.pdf

Feinstein, L. (2007). Mobility in students' cognitive attainment during school life. *Oxford Review of Economic Policy, 20,* 213–229.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier interventions in the primary grades. A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved August 30, 2012, from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/rti_reading_pg_021809.pdf

Gilmore, A., Croft, C., & Reid, N. (1981). *Burt word reading test-New Zealand revision*. Wellington: New Zealand Council for Educational Research.

Howell, D. (2008). *Fundamental statistics for the behavioural sciences*. Belmont, CA: Thomson Wadsworth.

Martin, F., & Pratt, C. (2001). *The Martin and Pratt nonword reading test*. Melbourne: ACER.

MultiLit. (2007a). *MultiLit reading tutor program (revised)*. Sydney: Author.

MultiLit. (2007b). *MultiLit word attack skills (revised)*. Sydney: Author.

MultiLit. (2007c). *MultiLit sight words (revised)*. Sydney: Author.

MultiLit. (2007d). *MultiLit reinforced reading (revised)*. Sydney: Author.

National Institute of Child Health and Human Development. (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: US Government Printing Office. Retrieved August 30, 2012, from http://www.nationalreadingpanel.org/publications/publications.htm

Neale, M. (1999). *Neale analysis of reading ability* (3rd ed.). Melbourne: Australian Council for Educational Research.

New South Wales Department of Education and Communities. (2013). *Reading recovery—A research-based early intervention program*. Welcome to Reading Recovery. Retrieved August 1, 2013, from http://www.curriculumsupport.education.nsw.gov.au/earlyyears/reading_recovery/

Organisation for Economic Cooperation and Development. (2013). *PISA 2012 results: What students know and can do—Student performance in mathematics, reading and science* (Vol. 1). Paris: Author.

Pearce, S., Wheldall, K., & Madelaine, A. (2006). MULTILIT book levels: Towards a new system for levelling texts. *Special Education Perspectives, 15*, 38–56.

Pogorzelski, S., & Wheldall, K. (2005). The importance of phonological processing skills for older low-progress readers. *Educational Psychology in Practice, 21*, 1–22. http://dx.doi.org/10.1080/02667360500035074

Reading Recovery Council of North America. (2013). Retrieved August 1, 2013, from http://readingrecovery.org/

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. http://dx.doi.org/10.1080/1034912X.2011.598406

Rose, J. (2006). Independent review of the teaching of early reading (Final report). Retrieved August 31, 2012, from www.education.gov.uk/publications/eOrderingDownload/0201-2006PDF-EN-01.pdf

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

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407. http://dx.doi.org/10.1598/RRQ.21.4.1

Stuart, M., Dixon, M., Masterson, J., & Gray, B. (2003). Children's early reading vocabulary: Description and word frequency lists. *British Journal of Educational Psychology, 73*, 585–598.

Tanner, E., Brown, A., Day, N., Kotecha, M., Low, N., Morrell, G., … Purdon, S. (2011). *Evaluation of every child a reader (ECaR)* (Research Report DFE-RR114). Cheshire: Department for Education.

Thomson, S., De Bortoli, L., & Buckley, S. (2013). *PISA 2012: How Australia measures up—The PISA 2012 assessment of students' mathematical, scientific and reading literacy*. Camberwell: Australian Council for Educational Research.

Thomson, S., Hillman, K., Wernert, N., Schmid, M., Buckley, S., & Munene, A. (2012). *Monitoring Australian Year 4 student achievement internationally: TIMSS and PIRLS 2011*. Melbourne: Australian Council for Educational Research.

Torgesen, J. K. (2005). Recent discoveries on remedial interventions for children with dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook*, Retrieved April 3, 2009, from http://www.blackwellreference.com/subscriber/tocnode?id=g9781405114882_chunk_g978140511488236

Wheldall, K. (2009). Mona Tobias award winner, 2008 Effective instruction for socially disadvantaged low-progress readers: The schoolwise program. *Australian Journal of Learning Difficulties, 14*, 151–170. http://dx.doi.org/10.1080/19404150903264294

Wheldall, K., & Beaman, R. (2000). *An evaluation of MultiLit: "Making up lost time in literacy"*. Canberra: Commonwealth Department of Education, Training and Youth Affairs.

Wheldall, K., & Beaman, R. (2010). Effective instruction for older low-progress readers: Meeting the needs of indigenous students. In C. Wyatt-Smith, J. Elkins, & S. Gunn (Eds.), *Multiple perspectives on difficulties in learning literacy and numeracy* (pp. 255–274). New York, NY: Springer.

Wheldall, K., & Madelaine, A. (2006). The development of a passage reading test for the frequent monitoring of performance of low-progress readers. *Australasian Journal of Special Education, 30*, 72–85. http://dx.doi.org/10.1080/10300110609409366

Appendix 1

Measures

Burt Word Reading Test (Gilmore et al., 1981). A test measuring single word reading using a list of 110 words that increase in difficulty. The reading age achievable on the Burt is about 13 years maximum. The Burt has high internal consistency (>.96), high test–retest reliability (>.95), and high criterion validity (correlations of .90–.98 between the Burt Word Reading Test and the Oral Word Reading Test and the Schonell Graded Word Reading Test (Gilmore et al., 1981).

Martin and Pratt Nonword Reading Test, Form A (Martin & Pratt, 2001). A test measuring phonological recoding ability in students aged from 6 to 16 years, using pseudowords of increasing difficulty (Martin & Pratt, 2001). It has a high test–retest reliability coefficient of .96, high alternative-forms reliability coefficients of .92–.96, and a high internal consistency reliability coefficient of .96 (Martin & Pratt, 2001). Positive correlations between the Martin and Pratt and the WRMT-R Word Attack (Woodcock, 1987) (.89), Coltheart and Leahy Nonword reading lists (.93), and the Neale Analysis of Reading Ability (.78–.88) indicate good criterion-related validity (Martin & Pratt, 2001).

Wheldall Assessment of Reading Passages (Wheldall & Madelaine, 2006). A curriculum-based measure of passage reading fluency for older low-progress readers (Year 3 and above), the test consists of standardized passages in the form of short stories of 200 words in length. Students are presented with the passage on a page and instructed to read the words aloud quickly and carefully. They are asked to stop after 1 min and their score is the number of words read correctly per minute. In the test battery, the score was averaged over three passages. For the purposes of screening in this study, two additional WARP passages, not included in either the assessment passages, were employed. The WARP has been found to have high parallel forms and repeated measures reliability, with coefficients typically exceeding .95. The validity of the WARP has been shown with high correlations between WARP scores and the accuracy measure of the Neale Analysis of Reading Ability (.87) and the Burt Word Reading Test (.85) (Wheldall & Madelaine, 2006).

Neale Analysis of Reading Ability (Neale, 1999). A test assessing students' reading accuracy and reading comprehension. The student is presented with six text passages of increasing difficulty and they are asked to read them aloud. Errors are recorded and used to calculate an accuracy score. After each passage has been read, the examiner asks the students a set of comprehension questions and the number of correct responses provides a comprehension score. The Neale Analysis has been shown to have high levels of internal consistency for accuracy and comprehension with correlations ranging from .71 to .96 (Neale, 1999). Good criterion-related validity has been demonstrated with Pearson product moment correlations ranging between .88 and .96 found between the Neale Analysis for accuracy and comprehension and the Schonell Graded Word Reading Tests (Neale, 1999).

MultiLit Word Attack Skills Placement Test (MultiLit, 2007b). A test assessing students' knowledge of letter-sound correspondences and their ability to read phonetically regular words and non-words. Letters and words are presented in a specific sequence of increasing complexity, beginning with single sounds and simple words and progressing to digraphs, blends, and words requiring rule-based strategies. The test is used to determine the appropriate starting level of the MultiLit Word Attack Skills component of the MultiLit program and, in this trial, to create instructional groups of students beginning at a similar level.

cogent •• education

Appendix 2

The intervention

(From Buckingham, Beaman, & Wheldall, 2012)

The MultiLit program implemented in this trial was an adapted version of the MultiLit Reading Tutor Program (MultiLit, 2007a), a one-to-one intervention for struggling readers in Year 2 and above. The adapted version was developed for small group instruction with the objective of providing an effective and less costly intervention for older low-progress readers within a school setting.

The MultiLit program incorporates the key elements of effective reading instruction identified in large-scale reviews, including the (Australian) National Inquiry into the Teaching of Literacy (DEST, 2005). These elements are phonemic awareness, phonics, fluency, vocabulary, and comprehension.

MultiLit has three main components: Word Attack Skills (Accuracy and Fluency), Sight Words, and Reinforced Reading. Each 1 h MultiLit session is divided into four discrete lessons of a set length.

Word Attack Accuracy (15 min). This lesson provides explicit instruction in phonemic awareness and phonic word decoding skills. The level at which students begin instruction is determined by their performance on the MultiLit placement test, which is directly aligned with the teaching program. Students progress through specifically sequenced sub-levels and levels as they master the content. Their rate of progress is dependent on the speed of mastery. The aim of this lesson is to provide students with skills to read familiar and unfamiliar words using phonetic rules and decoding strategies (MultiLit, 2007b).

Word Attack Fluency (10 min). In this lesson, students work on the same set of materials as in Word Attack Accuracy, but in this case, working on the development of fluency—speed and accuracy. Students work at least one level below their Word Attack Accuracy level, but are usually several levels below. As in Word Attack Accuracy, rate of progress is determined by mastery. Students must achieve a fluency target within each sub-level to progress to the next sub-level. The aim of this lesson is to develop fluency and automaticity in reading decodable text (MultiLit, 2007b).

Sight Words (15 min). In this lesson, students are taught 20 lists of 10 sight words, moving onto a new list each time they demonstrate mastery. The 200 sight words are high frequency words, both phonically regular and irregular, derived from a content analysis of children's storybooks (Stuart, Dixon, Masterson, & Gray, 2003). The aim of this lesson is to develop students' rapid recognition of commonly occurring words, and thereby allow them more immediate access to text (MultiLit, 2007c).

Reinforced Reading (20 min). Using a book of an appropriate instructional level, the teacher first introduces any unknown vocabulary or sight words. Comprehension questions about the previous reading are asked if it is a book being continued, or the teacher introduces a new story. Students take turns to read aloud from the book, with feedback and guidance provided by their teacher using the revised Pause, Prompt, Praise (PPP) tutoring method (MultiLit, 2007d), in which students are given several seconds to attempt an unknown word without help, followed by a sequence of prompts, and finally given specific praise for their efforts. Comprehension questions are again asked at the end of the session. The student's instructional level is determined using a 100-word sample of text from a book. An instructional-level book is one which students read with 90–95% accuracy. MultiLit has its own system for leveling texts consisting of 10 "M-levels" which have been graded against other leveling systems to allow the selection of appropriate books (Pearce, Wheldall, & Madelaine, 2006).

The above four lesson types are the main components of the MultiLit program. As students move through different parts of the program at different rates, however, components can also be completed at different rates. Several additional MultiLit program components were introduced to the MultiLit lessons as groups completed the main components.

PPP-C (Comprehension). The procedure is similar to regular PPP, but students read aloud from a book that is two levels below their instructional level to allow students to read fluently and focus on meaning. After a few minutes, or a few paragraphs, students are asked to turn their books face down and the teacher asks a number of "who, what, when, where, why, how" comprehension questions (MultiLit, 2007d).

PPP-F (Fluency). This comprises guided text reading using the PPP method that focuses on the development of fluency and prosody. Teachers model fluent and expressive reading and provide guidance to students who take turns to read aloud from books two levels below their instructional level (MultiLit, 2007d).