The Educational and Developmental Psychologist

www.cambridge.org/edp

A small group intervention for older primary school-aged low-progress readers: Further evidence for efficacy

Kevin Wheldall[1,2], Nicola Bell[1], Robyn Wheldall[1,2], Alison Madelaine[1], Meree Reynolds[1] and Sarah Arakelian[1]

[1]MultiLit Research Unit, MultiLit Pty Ltd, Sydney, New South Wales, Australia and [2]Department of Educational Studies, Macquarie University, Sydney, New South Wales, Australia

Original Article

Cite this article: Wheldall K, Bell N, Wheldall R, Madelaine A, Reynolds M, and Arakelian S. (2019) A small group intervention for older primary school-aged low-progress readers: Further evidence for efficacy. *The Educational and Developmental Psychologist* **36**: 35–40. doi: 10.1017/edp.2019.15

Received: 12 July 2019
Revised: 17 September 2019
Accepted: 21 September 2019
First published online: 21 October 2019

Keywords:
reading; literacy; children; instruction; Tier 2; MacqLit

Author for correspondence:
Kevin Wheldall,
Email: kevin.wheldall@pecas.com.au

Abstract

The aim of the present study was to determine the efficacy of a small-group reading instruction program that was delivered over two school terms to Australian students in Years 3 through 6. A large cohort ($n = 239$) of primary school children was assessed on their literacy skills before and after receiving 'MacqLit', a phonics-based program designed for older struggling readers. Parametric and nonparametric difference tests were used to compare results at pre- and postintervention time points. Statistically significant improvements with large effect sizes were observed on all raw score measures of word reading, nonword reading, passage reading and spelling. Statistically significant improvements were also observed on standard score measures of nonword reading and passage reading, suggesting the gains were greater than what might be expected to have resulted from typical classroom instruction. The results indicate that older middle primary school-aged students may benefit from phonics-based, small-group reading instruction.

The negative consequences of poor literacy are well known. Studies have linked early reading difficulties with broader academic failure, reduced employability, heightened risk of mental illness, and antisocial behaviour (Snow, 2016; World Literacy Foundation, 2015). Given the predictive status of literacy in determining key life outcomes, it is imperative that children are exposed to high quality literacy instruction which, based on the best available evidence, is most likely to set them up for later success.

According to results from the most recent Progress in International Reading Literacy Study (PIRLS) in 2016, 20% of Australian children in Year 4 achieved Low or Below Low benchmark reading scores (Thomson, Hillman, Schmid, Rodrigues, & Fullarton, 2017). As such, the findings highlight the substantial proportion of Australian children who are underachieving, compared with global literacy standards. Foundational language and literacy skills in beginning readers are highly predictive of later reading success (Hjetland et al., 2019; Stanovich, 1986), which suggests early intervention will certainly play a role in positively influencing PIRLS results in future cohorts of at-risk or struggling students.

The provision of such instruction notwithstanding, educational resources should also be dedicated to supporting the large number of middle primary-aged students who demonstrate limited reading progress. By Year 3, Australian students are generally no longer exposed to whole-class instruction devoted to 'basic' word-level reading skills. In other words, students tend to transition at around middle primary school from 'learning to read' to 'reading to learn' (Chall, 1983). If unsuccessful, the manifestation of this transition is known as the 'fourth grade slump', which refers to the sudden and often unexpected emergence of a student's relative difficulty completing tasks with high reading, spelling and word meaning demands (Chall & Jacobs, 2003).

The educational context outlined above serves as a rationale for implementing a 'Response to Intervention' (RTI) model of instruction (Westwood, 2017). According to an RTI framework, students should receive literacy intervention at increasingly intensive tiers, based on the level of skill proficiency they exhibit in comparison with their peers (Balu et al., 2015). Tier 1 is whole-class instructional provision, while Tiers 2 and 3 are typically small-group and individual instruction, respectively. The premise of an RTI framework is, therefore, that instructional resources are allocated equitably, with the largest proportion going to those who need it the most.

It is with this framework in mind that the *Macquarie Literacy Program for Small Group Instruction* (or 'MacqLit') was developed as a Tier 2 program for older (i.e., Year 3 and above) primary school-aged low-progress readers (MultiLit, 2016a). Within the program, the reading skills of low-progress readers, defined by the program developers as the bottom 25% of students,

© Australian Psychological Society Ltd 2019.

CAMBRIDGE UNIVERSITY PRESS

are targeted on a small-group basis. The program's content is intended to address each of the five components identified by large-scale national reviews as necessary for reading instruction: phonemic awareness, phonics, fluency, vocabulary and comprehension (National Reading Panel, 2000; Rose, 2006; Rowe, 2005). According to the developers, the primary aim of the program is to improve readers' rapid, automatic and efficient word recognition skills through synthetic phonics instruction (MultiLit, 2016b). (Further information about the program content and structure is provided in the Methods section.) MacqLit has been implemented in almost 1,500 schools Australia-wide, and given its popularity, research to determine its efficacy is warranted.

To date, there have been several studies examining the efficacy of MacqLit instruction. The teacher's manual for the program provides a description of research results pertaining to 362 students who attended an early, draft version of the MacqLit program (MultiLit, 2016b). After two terms of instruction, the students obtained significantly better raw scores on measures of reading and spelling, with large (Cohen's $d \geq 0.8$) treatment effect sizes evident. Importantly, these students received an average of three hours of daily instruction, which differs from the one-hour structure stipulated in the standard published program. In addition, other literacy-based programs were implemented that were likely to have had some effect on observed improvements. More large-scale research into MacqLit would therefore be elucidative in better understanding its treatment effects when isolated from other literacy programs.

Efficacy of the MacqLit program was also investigated in a modest, though well-controlled randomised control trial (RCT) conducted by Buckingham, Beaman, and Wheldall (2012). In this study, 44 low-progress readers in Years 3 through 6 were randomly allocated to either an experimental or control group. Those in the experimental group received a pilot version of the MacqLit (or, as it was known then, 'MultiLit') program for a recommended one hour per day, while the remaining students received their usual classroom literacy instruction. After two terms, nonword reading accuracy was significantly higher in the experimental group, and this was associated with a very large effect size (partial $\eta^2 = 0.405$). A similar pattern of results was found for a subsample of the original cohort ($n = 30$) who received three terms of instruction: students in the experimental group had significantly better nonword reading accuracy, which was associated with a very large effect size (partial $\eta^2 = 0.520$). Group differences did not reach statistical significance with respect to word reading accuracy, reading fluency, reading comprehension, spelling accuracy, or passage reading accuracy. Nevertheless, at both testing time points, treatment effect sizes ranged from medium to large for many of the measures, suggesting that the study may have been underpowered to detect statistical group differences, due to the small sample size.

Twenty-six students from the RCT by Buckingham et al. (2012) also participated in a second phase of the study (Buckingham, Beaman-Wheldall, & Wheldall, 2014). Here, the original control group became the experimental group (hereafter E2), while the original experimental group (hereafter E1) returned to their usual literacy instruction. At the end of their three terms of MacqLit instruction, the E2 group made statistically significant gains, associated with large effect sizes (Cohen's $d = 1.27$–3.80), on measures of nonword reading accuracy, real word reading accuracy, passage reading accuracy, reading fluency, and reading comprehension. In contrast, the E1 group did not improve significantly over the same period of time. To summarise the results from Buckingham et al. (2012; 2014), low-progress readers appeared to improve as a result of receiving MacqLit instruction, although the modest sample size in each phase of the RCT limited the degree of statistical power available for analyses.

The present study was conducted with a cohort of low-progress readers, all of whom received school-based MacqLit instruction between 2016 and 2018. These readers formed a large convenience sample ($n = 239$). Since a 'gold-standard' RCT methodology was not employed here, our intention was not to attribute any observed treatment effects unequivocally to the MacqLit program; nevertheless, the large sample size meant that the results may be used to support prior RCT findings. Ultimately, the aim of the study was to determine the statistical and practical significance of reading progress, as measured over the course of two school terms. Accordingly, this research is expected to contribute to public knowledge about MacqLit program efficacy.

Methods

Participants

Between the start of 2016 and the end of 2017, eight successive student intake cohorts ($n = 276$) attended MacqLit, a school-based small-group reading intervention program, for the first time. The majority of students from all schools had a language background other than English (65–98%). The schools had Index of Community Socio-Educational Advantage (ICSEA) scores ranging from 903 to 956 (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2019). For context, an average ICSEA score indicating average socio-educational status is 1000, whereas 500 indicates extreme disadvantage and 1300 indicates extreme advantage (ACARA, 2011). On the whole, the cohort therefore appeared to be from a slightly disadvantaged socio-educational context, relative to other Australian schoolchildren.

The process of recruitment originated with the developers of MacqLit (i.e., MultiLit) offering the program to schools, as funded by a charitable body. Tutors trained and employed by MultiLit implemented the program in the schools. Testers, also trained and employed by MultiLit, assessed students before and after intervention, in order to report this information back to the charitable trust funding the program. These data were accessed and collated for the present study.

Each student in the study was referred to the program after being identified by their school as a low-progress reader. Those students who were selected for the program also underwent a curriculum-based assessment screening protocol, to confirm their performance on word and passage reading measures was at or below the 25th percentile, in comparison to same-aged peers. Of the original 276 children, 37 started the program late, left the program part-way through, or were absent at post-testing. All data pertaining to these children were excluded from analyses. Consequently, complete pre- and posttest data were available for 239 students (112 females), none of whom had previously received MacqLit. These students attended schools in Sydney, New South Wales, and were in Year 3 ($n = 32$), Year 4 ($n = 45$), Year 5 ($n = 83$) or Year 6 ($n = 79$).

Measures

Students' data were gathered using a range of assessment measures, both before the program commenced, and then again after students had received approximately 18 weeks of instruction. All

https://doi.org/10.1017/edp.2019.15 Published online by Cambridge University Press

students were tested on the Burt Word Reading Test (Gilmore, Croft, & Reid, 1981), Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001), South Australian Spelling Test (SAST; Westwood, 2005), and the Wheldall Assessment of Reading Passages (Wheldall & Madelaine, 2000). Approximately a third of students ($n = 74$) were tested on the Neale Analysis of Reading Ability – 3rd edition (Neale, 1999), and the other two thirds of students ($n = 165$) were assessed on the York Assessment of Reading Comprehension – Passage Reading (Australian Edition; Snowling et al., 2012).

Burt Word Reading Test

The Burt Word Reading Test (Gilmore, Reid, & Croft, 1981) was used to assess word reading accuracy. Students are presented with 110 individual words on a card and are asked to read them aloud. Upon making 10 consecutive errors, examinees are given the opportunity to review and attempt any of the other stimuli on the sheet. The number of words read correctly is the raw score, which can be converted to a reading age using the standardised test norms. The test is suitable for children aged between 6.1 and 12.6 years. The Burt Word Reading Test has high test–retest reliability ($r > .95$) and internal consistency ($r > .96$; Gilmore et al., 1981). Results are significantly correlated ($r = .90$–$.98$) with both the Schonell Graded Word Reading Test (Shearer & Apps, 1975) and the Oral Word Reading Test (Fieldhouse, 1952), indicating it also has high criterion validity (Gilmore et al., 1981).

Martin and Pratt Nonword Reading Test

Nonword reading accuracy was assessed using the Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001). To read aloud nonwords, readers need to apply their knowledge of what phonemes are represented by specific letters or letter sequences. Performance is indicative of a student's ability to read novel or unfamiliar words in text. The examinee is asked to read aloud a series of nonwords, presented in groups of six per page. There are 54 items in total, and the test is discontinued when the student mispronounces eight consecutive words. The raw score (i.e., the total number of correct responses) can be converted to a reading age equivalent score, standard score or percentile rank. Standardised norms are available for examinees aged between 6.0 and 17.0 years. Form A of the Martin and Pratt Nonword Reading Test, which was the form used in the present study, has a high test–retest reliability ($r = .96$) and a high internal consistency reliability ($r = .96$; Martin & Pratt, 2001). The test is significantly correlated with the Woodcock Reading Mastery Test – Revised Word Attack subtest ($r = .89$; Woodcock, 1987), Coltheart and Leahy Nonword reading test ($r = .93$; Coltheart & Leahy, 1996), and Neale Analysis of Reading Ability accuracy score ($r = .78$–$.88$; Neale, 1999), indicating good criterion-related validity (Martin & Pratt, 2001).

South Australian Spelling Test

Single word spelling was assessed using the South Australian Spelling Test (SAST; Westwood, 2005). The student is required to write a list of real words dictated to them one word at a time by the test administrator. In total, the list contains 70 words, and the test is discontinued when the student makes 10 consecutive errors. Test–retest reliability for the SAST is high ($r = .96$) for most year groups (Westwood, 2005). Spelling age equivalent scores are available, ranging from 6.0 to 15.0 years.

Wheldall Assessment of Reading Passages

The Wheldall Assessment of Reading Passages (WARP; Wheldall & Madelaine, 2000) was used to measure passage reading fluency. The WARP is a curriculum-based measure, wherein students are given one minute to read a 200-word written passage. Each individual passage is scored based on the total number of words read accurately within the timeframe. In the present study, three Initial Assessment Passages (A, B and C) were administered. The raw score represented the average scores of all three passages. The WARP has high alternate forms reliability ($r = .95$–$.96$) and internal consistency ($r = .97$–$.99$; Wheldall & Madelaine, 2000). Furthermore, it correlates well with the accuracy score of the Neale Analysis of Reading Ability ($r = .78$–$.80$), indicating good criterion-related validity.

York Assessment of Reading Comprehension – Passage Reading (Australian Edition)

For the majority of the students ($n = 165$), passage reading accuracy, rate and comprehension were assessed using the York Assessment of Reading Comprehension – Passage Reading (YARC; Snowling et al., 2012). In the YARC, students read aloud two passages of text selected according to their age and reading ability and, following each passage, answered questions about the content. The test is discontinued when two passages are read with sufficient accuracy (i.e., ≤15 reading errors for passage Levels Beginner-2 and ≤20 reading errors for Levels 3–6) and comprehension (i.e., 2–6 correct responses to comprehension questions). Reading rate is measured as the number of seconds taken to read the passage. Using the test's standardised norms, raw scores for accuracy, comprehension and rate can be converted to 'ability scores', which are used to generate standard scores for students aged between 5 and 12 years old. Ability scores account for text difficulty, which is important, since this increases from passage Levels 1 through 6. For example, 10 errors (or 2 correct comprehension answers) indicate substantially different ability levels if observed on the Level 1 versus Level 6 passage.

Neale Analysis of Reading Ability – 3rd Edition

The remaining students ($n = 74$), were administered the Neale Analysis of Reading Ability – 3rd edition (NARA-3; Neale, 1999). In the NARA-3, students are asked to read aloud progressively longer and more difficult written passages, after which they answer literal and inferential comprehension questions about the passage content. The test is discontinued when the student reaches a certain threshold of reading errors (i.e., ≥16 for passage Levels 1–5 and ≥20 for Level 6). Raw scores are calculated for reading accuracy (i.e., the total number of reading errors) and reading comprehension (i.e., cumulated correct responses to comprehension questions). Using the test's standardised norms, raw scores can be converted to percentile ranks and reading age equivalent scores for students in Years 1 through 7 (Neale, 1999).

Intervention

The intervention employed in this study was 'MacqLit' (MultiLit, 2016a), an explicit and systematic reading program targeted towards small groups of struggling readers in Year 3 and above. The program is structured to comprise 122 scripted lessons. Each MacqLit lesson is an hour in duration, and includes synthetic phonics activities, sight word instruction, text reading, and shared

https://doi.org/10.1017/edp.2019.15 Published online by Cambridge University Press

Table 1. Means (and Standard Deviations) of All Raw Score Literacy Results Before and After Macqlit Instruction

Literacy skill (Assessment measure)	*n*	Pretest (*SD*)	Posttest (*SD*)	Gain (*SD*)	*t*	*p*	Partial η^2	Cohen's *d*
Real word reading accuracy (Burt Word Reading Test)	239	44.12 (14.93)	54.38 (15.89)	10.27 (6.50)	24.439	<.001	0.715	1.578
Nonword reading accuracy (Martin & Pratt)	238	17.82 (10.39)	27.39 (10.09)	9.57 (6.65)	22.180	<.001	0.675	1.437
Spelling (South Australian Spelling Test)	239	27.18 (7.96)	31.42 (6.35)	4.23 (3.47)	18.887	<.001	0.600	1.224
Reading fluency (WARP)	239	71.12 (32.19)	102.64 (37.38)	31.52 (14.26)	34.172	<.001	0.831	2.214
Passage reading accuracy (NARA-3)	74	40.59 (17.48)	52.57 (18.67)	11.97 (7.19)	14.33	<.001	0.738	1.667
Passage reading accuracy (YARC)	165	40.42 (5.48)	44.59 (5.00)	4.17 (3.15)	17.004	<.001	0.638	1.320
Reading comprehension (NARA-3)	74	9.18 (4.22)	11.99 (4.65)	2.81 (2.77)	8.735	<.001	0.511	1.017
Reading comprehension (YARC)	165	47.34 (8.90)	52.16 (7.63)	4.82 (7.31)	8.479	<.001	0.305	0.660
Passage reading rate (YARC)	151[a]	39.59 (12.36)	49.42 (10.24)	9.83 (5.92)	20.414	<.001	0.735	1.663

Note: [a]Reading rate scores could not be generated for students who read Beginner and Level 1 passages ($n = 14$). WARP = Wheldall Assessment of Reading Passages; NARA = Neale Analysis of Reading Ability; YARC = York Assessment of Reading Comprehension. Partial η^2 of 0.01 = small; 0.06 = medium; 0.14 = large. Cohen's *d* of 0.2 = small; 0.5 = medium; 0.8 = large.

Table 2. Means (and Standard Deviations) of All Standardised Literacy Results Before and After Macqlit Instruction

Literacy skill (Assessment measure)	*n*	Pretest (*SD*)	Posttest (*SD*)	Gain (*SD*)	*t*	*p*	Partial η^2	Cohen's *d*
Nonword reading accuracy (Martin & Pratt – standard score)	238	82.50 (13.30)	93.12 (11.90)	10.63 (9.00)	18.225	<.001	0.584	1.240
Passage reading accuracy (YARC – standard score)	85	80.31 (7.51)	84.54 (8.61)	4.24 (6.71)	5.823	<.001	0.288	0.631
Reading comprehension (YARC – standard score)	88	82.15 (7.99)	85.38 (10.00)	3.23 (10.46)	2.894	.005	0.088	0.309
Passage reading rate (YARC – standard score)	54	79.85 (7.94)	86.02 (9.27)	6.17 (5.67)	7.998	<.001	0.547	1.089

Note: YARC = York Assessment of Reading Comprehension. Partial η^2 of 0.01 = small; 0.06 = medium; 0.14 = large. Cohen's *d* of 0.2 = small; 0.5 = medium; 0.8 = large.

book reading. The lessons are sequenced so that simple single letter-sound correspondences are targeted first, followed by increasingly complex letter combinations and affixes. The lessons also include strategies for decoding multisyllabic words, and for recognising and decoding prefixes and suffixes. As stated, in the present study, MacqLit was delivered during class time by MultiLit tutors approximately four times per week over two school terms. Ordinarily (i.e., where school staff are delivering the program), two days of compulsory formal training are required prior to the program's implementation.

Results

Students' raw score results, collected both before and after receiving the reading intervention, are shown in Table 1. For the YARC, 'ability' scores served as raw scores. A series of paired *t* tests were conducted to calculate whether there was a significant difference between pre- and postintervention time points. Where the data for any measure were distributed nonnormally (i.e., Shapiro-Wilk $p < .05$), a Wilcoxon Signed-Rank Test was performed to confirm the statistical significance of results from parametric *t*-test analyses. To account for the multiple comparisons, the alpha level was set at 1% ($p < .01$) in lieu of a Bonferroni correction. Effect sizes were estimated using both partial η^2 and Cohen's *d* values, and were interpreted according to the following conventions: Partial η^2 of 0.01 = small; 0.06 = medium; 0.14 = large (Cohen, as cited in Pallant, 2005); Cohen's *d* of 0.2 = small; 0.5 = medium; 0.8 = large (Cohen, as cited in Piasta & Justice, 2010).

According to the results, performance improved significantly from pre- to postintervention time points for all raw score measures. For reading comprehension, the effect size was large (partial $\eta^2 = 0.305$) or medium-to-large (Cohen's $d = 0.660$). In all other cases, the effect sizes were large.

The difference between raw scores obtained at pre-and posttest were associated with crude reading age gains of 11 months for real word reading (8;1 to 9;0 years), 18 months for nonword reading (7;5 to 8;11 years), 12 months for spelling (7;4 to 8;4 years), 11 months for passage reading accuracy on the NARA-3 (8;1 to 9;0 years), 11 months for passage reading accuracy on the YARC (6;11 to 7;10 years), 6 months for reading comprehension on the NARA-3 (6;11 to 7;5 years), 12 months for reading comprehension on the YARC (6;8 to 7;8 years), and 10 months for reading rate on the YARC (6;6 to 7;4 years).

Standard scores were available for two of the assessment measures: the Martin and Pratt Nonword Reading Test and the YARC. Standard scores on the YARC could not be calculated if they fell under 70; hence, students who scored at floor levels were removed from analyses. A single extremely high difference score outlier was also removed from the Martin and Pratt raw and standard score results, because the score substantially affected normality and was more than three times greater than the interquartile range. Again, to account for the multiple comparisons, the alpha level was set at 1% ($p < .01$). As seen in Table 2, performance as measured by standard scores improved significantly from pre- to postintervention time points. The effect sizes were large for nonword reading accuracy and passage reading rate. For passage reading accuracy, the effect size was medium-to-large

https://doi.org/10.1017/edp.2019.15 Published online by Cambridge University Press

(Cohen's $d = 0.631$) or large (partial $\eta^2 = 0.288$). For passage reading comprehension, the effect size was small-to-medium (Cohen's $d = 0.309$) or medium-to-large (partial $\eta^2 = 0.088$).

Discussion

Based on the results of this large-cohort investigation into MacqLit efficacy, students receiving the program appeared to make substantial gains on a wide range of literacy measures over time. The findings therefore support previous studies in which MacqLit has elicited reading and spelling improvements in older school-age children (e.g., Buckingham et al., 2012, 2014).

A significant limitation of the study was that no control group was included with which to compare students' progress. Hence, it is not possible to unequivocally attribute observed improvements to provision of MacqLit specifically, and some may argue that similar gains would have been made in the absence of the program. Indeed, since MacqLit was delivered over the course of two school terms (i.e., an average of 18 weeks), some degree of improvement resulting from usual classroom instruction is to be expected. That said, standard scores available for measures of nonword reading accuracy, passage reading accuracy and reading comprehension all indicated that the gains were significantly greater than what might be expected in a same-aged cohort. This, in addition to the large effect sizes across raw score measures, and in the context of students' suboptimal literacy development prior to MacqLit, suggests the program produced significant improvements. Evidence for the reliability of reported outcomes is also derived from the fact that broadly similar results were found for each of the eight student intakes separately, as reported to students' schools and school sponsors.

It is interesting to note that the improvements extended to reading comprehension, although not to the same extent as reading (or spelling) accuracy. While connected text reading does feature as an instructional component of the MacqLit program, more time is devoted to targeting students' reading skills at the single word level. Hence, proximal word-level literacy tasks were arguably more susceptible to the effects of MacqLit instruction than distal reading comprehension skills.

Another explanation for the comparatively reduced growth observed in reading comprehension is that it is a more stable cognitive construct. Further longitudinal research may elucidate the results in this regard. In the present study, progress was tracked longitudinally only insofar as students' literacy skills were assessed before and after receiving the MacqLit program. In future, it would be useful to commit to research examining longer-term longitudinal outcomes – particularly for reading comprehension, which is more distally related to the word-level instructional focus of MacqLit. The early gains observed here are suggestive of this cohort of older low-progress readers progressing at a rate similar to – or greater than – similarly aged peers. To determine whether such gains are maintained, or to determine whether the students even 'catch up' with the standard of their peers, additional longitudinal research is warranted.

In the context of MacqLit's word-level instructional focus, it may be considered surprising that reading comprehension improvements were observed at all in the present study. Given the focus on targeting word-level skills, such findings are in broad agreement with the considerable body of research showing that word decoding and recognition skills – in combination with language comprehension – contribute to reading comprehension (Catts, Hogan, & Adlof, 2005; Hoover & Gough, 1990; Lonigan, Burgess, & Schatschneider, 2018). For students in the present study, word-level limitations may have formed a 'bottleneck' for higher level reading development. As such skills were progressively improved through instruction, additional cognitive resources could then be devoted to understanding the text meaning. This shift in skills underlying reading comprehension, resulting from increasingly automatic word reading, is captured by the assumptions of the Simple View of Reading model and has consistently been found in previous studies of typically developing children (Catts et al., 2005; Catts, Herrera, Nielson, & Bridges, 2015; Garcia & Cain, 2014; Language & Reading Research Consortium, 2015; Lonigan et al., 2018).

It is worth considering what aspects of the MacqLit program itself may have resulted in the observed changes to word-level reading and spelling skills. First, phonics instruction within the program was delivered according to a set scope and sequence. This systematicity is supported by prior research-based evidence (Castles, Rastle, & Nation, 2018). Specifically, the program aligns with a *synthetic* (i.e., part-to-whole) phonics approach, for which there is some evidence of comparative success over an *analytic* (i.e., whole-to-part) phonics approach (Rose, 2006). In addition, the synthetic phonics content was delivered alongside instruction that targeted other components identified as important to include in reading instruction (i.e., phonemic awareness, vocabulary, fluency and comprehension; National Reading Panel, 2000; Rose, 2006; Rowe, 2005). These factors are likely to have contributed to the improvements.

In spite of its lack of a control group and long-term follow-up, the present study contributes to a better understanding of MacqLit efficacy. Previous RCT studies have found that students receiving MacqLit progressed significantly in nonword reading ability, compared with students receiving usual literacy instruction (Buckingham et al., 2012, 2014). With the benefit of a large sample size, the present study had much more power to identify significant changes in performance over time. Hence, the results support – and are supported by – those from better controlled but smaller studies involving children on the MacqLit program.

More broadly, the results also provide insight into the efficacy of small-group literacy intervention. MacqLit was designed as a Tier 2 program, according to an RTI framework of instructional delivery. As such, the lowest-performing 25% of students in the population could receive more targeted, better paced instruction than what is delivered in a whole-class environment. Conversely, the program is not so intensive as to require individualised teacher resources. Hence, particularly in the context of the reported results, this small-group intervention program may be considered more cost-effective than other one-on-one intervention options.

In conclusion, the present study builds on existing evidence in support of MacqLit. The results suggest that middle primary school-aged low-progress readers who received the small-group literacy program developed improved literacy outcomes over a two-term duration of instruction.

Financial Support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of Interest. All authors are members of the MultiLit Research Unit. MultiLit Pty Ltd is the publisher of the MacqLit program.

Ethical Standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

https://doi.org/10.1017/edp.2019.15 Published online by Cambridge University Press

Data Sharing Statement. The data that support the findings of this study are available from the corresponding author, subject to research ethics restrictions.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2011). *Guide to understanding ICSEA*. Sydney, Australia: Author.

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2019). *My School*. Retrieved from https://www.myschool.edu.au/.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of Response to Intervention practices for elementary school reading*. Washington, DC: U.S. Department of Education.

Buckingham, J., Beaman, R., & Wheldall, K. (2012). A randomised control trial of a MultiLit small group intervention for older low-progress readers. *Australian Journal of Learning Difficulties, 17*, 79–99.

Buckingham, J., Beaman-Wheldall, R., & Wheldall, K. (2014). Evaluation of a two-phase experimental study of a small group ('MultiLit') reading intervention for older low-progress readers. *Cogent Education, 1*, 1–11.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*, 5–51.

Catts, H.W., Hogan, T.P., & Adlof, S.M. (2005). Developmental changes in reading and reading disabilities. In H.W. Catts & A.G. Kahmi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Mahwah, NJ: Erlbaum.

Catts, H.W., Herrera, S., Nielsen, D.C., & Bridges, M.S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing, 28*, 1407–1425.

Chall, J.S. (1983). *Stages of reading development*. New York: McGraw-Hill.

Chall, J.S., & Jacobs, V.A. (2003). Poor children's fourth-grade slump. *American Educator, 27*, 14–15.

Coltheart, M., & Leahy, J. (1996). Assessment of lexical and nonlexical reading abilities in children: Some normative data. *Australian Journal of Psychology, 48*, 136–140.

Department of Education, Science & Training. (2005).

Fieldhouse, A. E. (1952). *Oral word reading test*. Wellington, New Zealand: New Zealand Council for Educational Research.

Garcia, J.R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*, 74–111.

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

Hjetland, H.N., Lervåg, A., Lyster, S.H., Hagtvet, B.E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology, 111*, 751–763.

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

Language and Reading Research Consortium. (2015). Learning to read: should we keep things simple? *Reading Research Quarterly, 50*, 151–169.

Lonigan, C.J., Burgess, S.R., & Schatschneider, C. (2018). Examining the Simple View of Reading with elementary school children: Still simple after all these years. *Remedial and Special Education, 39*, 260–273.

Martin, F., & Pratt, C. (2001). *Martin and Pratt nonword reading test*. Melbourne, Australia: Australian Council for Educational Research.

MultiLit. (2016a). *MacqLit: the Macquarie program for small group instruction*. Sydney, Australia: MultiLit Pty Ltd.

MultiLit. (2016b). *MacqLit Manual*. Sydney, Australia: MultiLit Pty Ltd.

National Reading Panel. (2000). *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). National Institute of Child Health & Development. Retrieved from https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

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

Pallant, J. (2005). *SPSS survival manual* (2nd ed.). Sydney, Australia: Allen & Unwin.

Piasta, S.B., & Justice, L.M. (2010). Cohen's d statistic. In N.J. Salkind (Ed.) *Encyclopedia of research design* (pp. 181–185). Thousand Oaks, CA: Sage Publications.

Rose, J. (2006). *Independent review of the teaching of early reading*. Nottingham, UK: Department for Education and Skills.

Rowe, K. (2005). *Teaching reading: National inquiry into the teaching of literacy*. Melbourne, Australia: Australian Council for Educational Research. Retrieved from https://research.acer.edu.au/tll_misc/5/.

Shearer, E., & Apps, R.A. (1975). Restandardization of the Burt-Vernon and Schonell graded word reading tests. *Educational Research, 18*, 67–73.

Snow, P.C. (2016). Elizabeth Usher Memorial Lecture: Language is literacy is language – positioning speech-language pathology in education policy, practice, paradigms and polemics. *International Journal of Speech-Language Pathology, 18*, 216–228.

Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Nation, K., Truelove, E., & Hulme, C. (2012). *York assessment of reading for comprehension – passage reading (Australian ed.)*. London, UK: GL Assessment.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Thomson, S., Hillman, K., Schmid, M., Rodrigues, S., & Fullarton, J. (2017). *Reporting Australia's results: PIRLS 2016*. Melbourne, Australia: Australian Council for Educational Research.

Westwood, P. (2005). *Spelling: Approaches to teaching and assessment* (2nd ed.). Melbourne, Australia: Australian Council for Educational Research.

Westwood, P.S. (2017). *Learning disorders: A response-to-intervention perspective*. New York, NY: Routledge.

Wheldall, K., & Madelaine, A. (2000). A curriculum-based passage reading test for monitoring the performance of low-progress readers: The development of the WARP. *International Journal of Disability, Development and Education, 47*, 371–382.

Woodcock, R.N. (1987). *Woodcock reading mastery tests-revised examiner's manual*. Circle Pines, MN: American Guidance Service.

World Literacy Foundation. (2015). *WLF final economic report*. Retrieved from https://worldliteracyfoundation.org/wp-content/uploads/2015/02/WLF-FINAL-ECONOMIC-REPORT.pdf

https://doi.org/10.1017/edp.2019.15 Published online by Cambridge University Press