InitiaLit–2

Extended program trial summary

During 2018, 2019 and 2020, the MultiLit Research Unit conducted trials to investigate whether students receiving InitiaLit–2 made literacy skill gains. InitiaLit–2 was officially published in late 2019; hence, students involved in the 2018 and 2019 trials received a version of the program that was technically still under development, although the instructional content was similar to that in the final version. Results from the 2018 trial are detailed in the InitiaLit–2 teaching manual. Results from the 2019 trial are detailed in a summary that was published on the MultiLit website. This extended program trial summary serves to outline the results from those years, as well as from 2020.

To determine whether students' literacy skills improved, the following areas were assessed at the start and end of the school year:

- Word reading accuracy (Burt Word Reading Test; Gilmore et al., 1981);
- Spelling accuracy (South Australian Spelling Test [SAST]; Westwood, 2005);
- Passage reading accuracy (Neale Analysis of Reading Ability – 3rd ed. [NARA-3]; Neale, 1999);
- Passage reading comprehension (NARA-3; Neale, 1999), and;
- Passage reading fluency (Wheldall Assessment of Reading Passages [WARP]; Wheldall & Madelaine, 2013).

In 2019 and 2020, nonword reading accuracy (Martin and Pratt Nonword Reading Test; Martin & Pratt, 2001) was also assessed.

There were approximately nine months between pre- and post-test time points. In 2018 and 2019, this duration was equivalent to approximately 33 weeks of actual instruction. In 2020, it was equivalent to approximately 27 weeks of actual, face-to-face instruction, due to COVID-19 pandemic-related school closures and interruptions.

Trial schools

Information about the schools involved in the trial each year is provided in Table 1. The level of socio-educational advantage in the general student populations of the schools ranged from average (ICSEA = 900-1100) to above-average (ICSEA >1100). A little over half the general student population (across schools) had a language background other than English. The mean age of all students at pre-test was 7 years, 3 months (SD = 4 months; range = 6;7-8;2). In 2018, two schools (90 students) participated in the trial, while in 2019, three schools (153 students) participated, and in 2020, three schools* (186 students) participated.

Table 1. Brief description of schools involved in trials.

Year	School	Location	*n*	% Females	ICSEA	% LBOTE**
2018	1	Perth, WA	56	54%	Above average	30%
2018	2	Sydney, NSW	34	38%	Average	70%
2019	3	Sydney, NSW	73	49%	Above average	50%
2019	4	Perth, WA	48	46%	Above average	30%
2019	5	Sydney, NSW	32	47%	Average	75%
2020	6	Sydney, NSW	36	61%	Average	60%
2020	7	Sydney, NSW	32	13%	Above average	5%
2020	8	Sydney, NSW	118	56%	Average	85%
	ALL		**Tot. = 429**	**Av. = 48%**	**Av. = 1103**	**Av. = 56%**

Note: *The 2020 trial involved two campuses of the same school. **Rounded to the nearest 5% to preserve schools' anonymity. ICSEA = Index of Community Socio-Educational Advantage; LBOTE = Language background other than English.

Did reading skills improve over the year?

The pre- and post-test raw score results were first analysed to determine whether there were any significant reading improvements between the start and end of the school year. As shown in Tables 2 through 4, students in all three trial years made statistically significant gains on the assessed areas of literacy. Based on the large effect sizes (Cohen's d > 0.8), these gains were also substantial.

Table 2. 2018 trial: Raw score means (and standard deviations) for pre- and post-test time points on all literacy skills measured.

Literacy variable	n	Pre-test Raw score (SD)	Post-test Raw score (SD)	Raw score (SD)	Gain		
					t	p	Cohen's d
Word reading accuracy	90	46.00 (15.96)	60.02 (17.02)	14.02 (7.28)	18.278	<.001	1.82 (L)
Spelling accuracy	90	28.19 (7.47)	33.93 (6.99)	5.74 (3.40)	16.011	<.001	1.55 (L)
Passage reading accuracy	89	35.43 (16.93)	53.27 (18.92)	17.84 (10.00)	16.829	<.001	1.71 (L)
Passage reading comprehension	89	7.58 (3.73)	12.74 (5.85)	5.16 (4.31)	11.285	<.001	1.17 (L)
Passage reading fluency 2020	88	72.49 (36.46)	118.05 (40.78)	45.56 (21.52)	19.863	<.001	2.12 (L)

Note: Where data were non-normally distributed, a Wilcoxon Signed-Rank Test was performed to confirm the statistical significance of parametric t-test results. When interpreting Cohen's d effect sizes, a small (S) effect is 0.2; a medium (M) effect is 0.5; and a large (L) effect is 0.8.

Table 3. 2019 trial: Raw score means (and standard deviations) for pre- and post-test time points on all literacy skills measured.

Literacy variable	n	Pre-test Raw score (SD)	Post-test Raw score (SD)	Raw score (SD)	Gain		
					t	p	Cohen's d
Word reading accuracy	153	44.69 (15.42)	60.82 (17.68)	16.14 (7.64)	26.121	<.001	2.11 (L)
Spelling accuracy	153	27.98 (7.80)	34.10 (6.83)	6.12 (3.78)	20.018	<.001	1.62 (L)
Nonword reading accuracy	153	27.40 (12.01)	34.08 (12.04)	6.63 (6.00)	13.744	<.001	1.11 (L)
Passage reading accuracy	153	33.63 (16.45)	52.24 (20.15)	18.58 (8.15)	28.198	<.001	2.28 (L)
Passage reading comprehension	153	8.36 (4.31)	15.03 (6.83)	6.67 (4.35)	18.984	<.001	1.53 (L)
Passage reading fluency	152	68.37 (37.38)	105.63 (39.45)	37.26 (17.62)	26.069	<.001	2.11 (L)

Note: Refer to Note under Table 2.

Table 4. 2020 trial: Raw score means (and standard deviations) for pre- and post-test time points on all literacy skills measured.

Literacy variable	n	Pre-test Raw score (SD)	Post-test Raw score (SD)	Raw score (SD)	Gain		
					t	p	Cohen's d
Word reading accuracy	186	52.22 (15.25)	65.44 (15.69)	13.22 (7.50)	24.029	<.001	1.76 (L)
Spelling accuracy	186	31.10 (5.87)	36.05 (5.47)	4.95 (3.90)	17.292	<.001	1.27 (L)
Nonword reading accuracy	186	30.43 (10.32)	37.01 (8.86)	6.58 (6.39)	14.022	<.001	1.03 (L)
Passage reading accuracy	186	39.94 (16.66)	55.72 (18.97)	15.78 (9.19)	23.428	<.001	1.72 (L)
Passage reading comprehension	186	9.51 (4.75)	13.56 (6.27)	4.05 (3.99)	13.865	<.001	1.02 (L)
Passage reading fluency	186	84.71 (36.76)	118.62 (35.90)	33.90 (16.66)	27.671	<.001	2.03 (L)

Note: Refer to Note under Table 2.

It was interesting to note that the gains observed in 2020 were reduced in comparison to those from 2018 or 2019, although they were still significant and associated with large effect sizes. There are a few possible reasons for this. The first is that, obviously, 2020 was an unusual year. Due to the COVID-19 pandemic, the students in the 2020 trial experienced at least six weeks of either complete school closure, or adapted delivery of learning content via an online platform. This content may or may not have included InitiaLit–2. Even beyond those six weeks, students may also have been withdrawn for additional periods of time by parents, and their learning while present at school may have been affected by pandemic-related stress. Relatedly, there was no opportunity to observe the program in action at any point during the year, which means instructional fidelity may have differed between (and within) schools.

Another factor worth considering is that the pre-test raw scores for the 2020 trial cohort were, across the board, higher than those from 2018 and 2019. As such, these students had less to gain from exposure to the InitiaLit–2 program.

One final factor worth highlighting is that the 2020 cohort was likely to comprise more students with a language background other than English (mean %LBOTE of general school community = 65%), when compared with the cohorts from 2018 (= 47%) and 2019 (= 51%). This factor was not controlled or accounted for on an individual level, and so we can only make tentative hypotheses as to the effect on learning.

With the above caveats in place, the gains in students' reading and spelling skills nevertheless increased significantly and substantially from pre- to post-test time points, for all three years of the trial.

Did reading skills improve compared to typical reading scores?

All the trials from 2018 to 2020 involved only participants who were receiving the InitiaLit–2 program (i.e., there were no comparison or control groups). As such, there is no *experimental* evidence to show that the observed gains were greater than what they would have been for a typically developing cohort of Year 2 students. However, age equivalent scores and percentile ranks were available for some of the assessment measures used in the trial. These give a good indication of how the students performed, relative to their same-aged peers.

As seen in Table 5, students made gains which were, in most cases, substantially greater than the actual nine months between pre- and post-test time points. This suggests that the observed reading and spelling improvements may be attributed to the program which targeted these skills: InitiaLit–2.

Table 5. Gains in age equivalent scores for each trial year.

Literacy variable	2018 Age equivalent gain (months)	2019 Age equivalent gain (months)	2020 Age equivalent gain (months)
Word reading accuracy	18	21	20
Spelling accuracy	17	17	14
Nonword reading accuracy	Not assessed	20	25
Passage reading accuracy	17	17	15
Passage reading comprehension	10	15	9

The area where students appeared to improve the least, relative to same-aged peers, was passage reading comprehension, as measured by the NARA-3. Comprehending text is a complex process, involving many different underlying skills. InitiaLit–2 targets many of these skills (e.g., word-level reading accuracy and automaticity, reading fluency, language comprehension, vocabulary). However, other factors contributing to reading comprehension are out of the program's scope (e.g., content knowledge). As such, we may expect to see the program positively affect students' reading comprehension, but not to the same degree as the other assessed skills, which were more closely related to what is specifically targeted in InitiaLit–2.

In 2020, students' average passage reading comprehension performance improved by nine months, which is equivalent to the actual elapsed time from pre- to post-test. This is the only measure and year in which

students' average increase in age equivalent score did not exceed their increase in chronological age. As previously hypothesised, this comparative lack of learning acceleration in 2020 could have been due to pandemic-related schooling interruptions, the higher starting point of reading comprehension scores in 2020, and/or the high proportion of students with a language background other than English. It may also have been due to an interaction between the first and third factors: for students who primarily spoke a different language at home, school closures would have reduced their exposure to spoken English. Given the importance of language comprehension to reading comprehension, it may be expected that this metric was more affected than others by reduced English exposure.

Percentile rank scores were available for the passage reading accuracy and comprehension measure (i.e., NARA-3) and for the nonword reading accuracy measure (i.e., Martin and Pratt Nonword Reading Test). For conciseness, percentile rank scores for all three years were combined and graphed (see Figures 1-3). Nonword reading accuracy percentile rank scores were based on the test's age-based standardised norms. Passage reading accuracy and comprehension percentile rank scores were based on the test's grade-based standardised norms. Since the students were in Year 2 at both time points (i.e., years of schooling = 3), these NARA-3 norms do not account for the fact that some reading development is expected to take place in response to nine months-worth of really any schooling. Hence, please note that Figures 1 and 2 likely overestimate pre- to post-test improvements, as captured by percentile rank scores. The standardised test norms for the NARA-3 were collected between September and November of 1997, which means that the post-test data from our trials are more accurate than the pre-test data, in terms of percentile ranks.

Figure 1. Percentage of students scoring in the lower quartile, middle range and upper quartile for passage reading accuracy, at pre-test and at post-test.
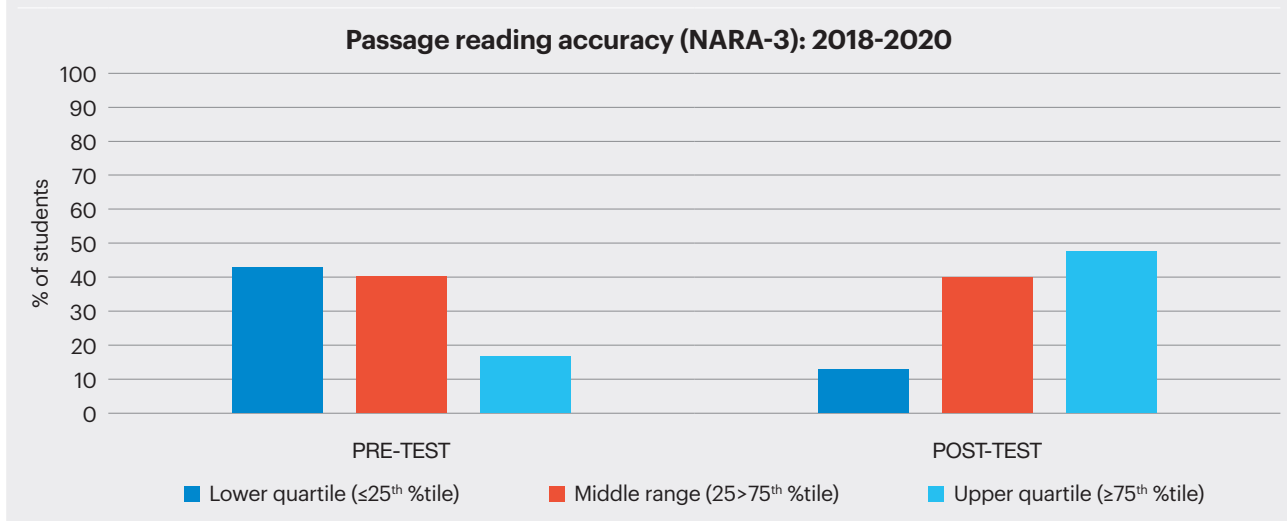

Figure 2. Percentage of students scoring in the lower quartile, middle range and upper quartile for passage reading comprehension, at pre-test and at post-test.
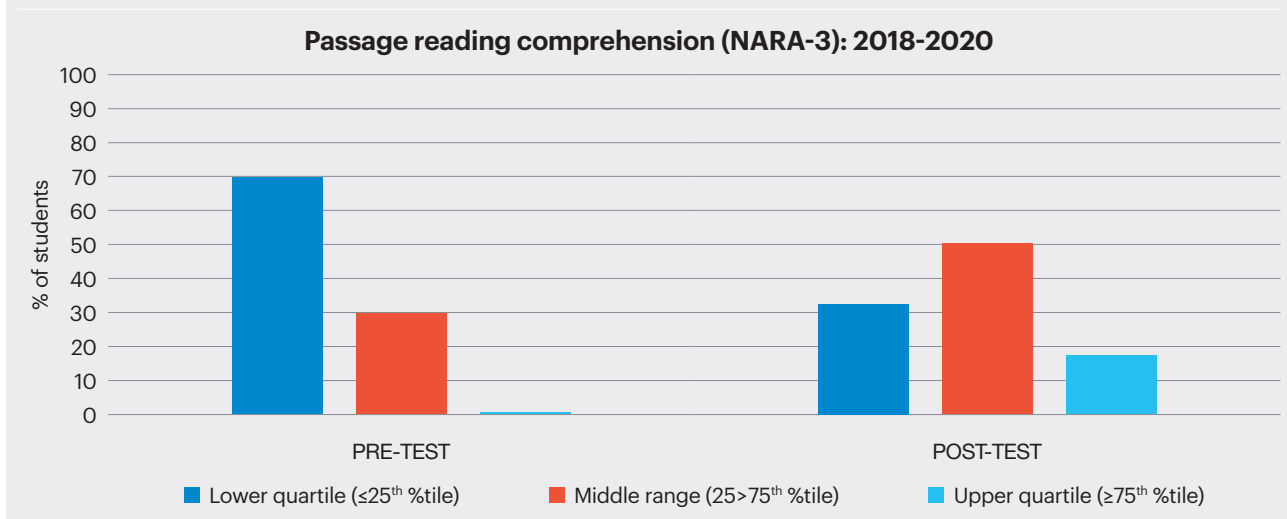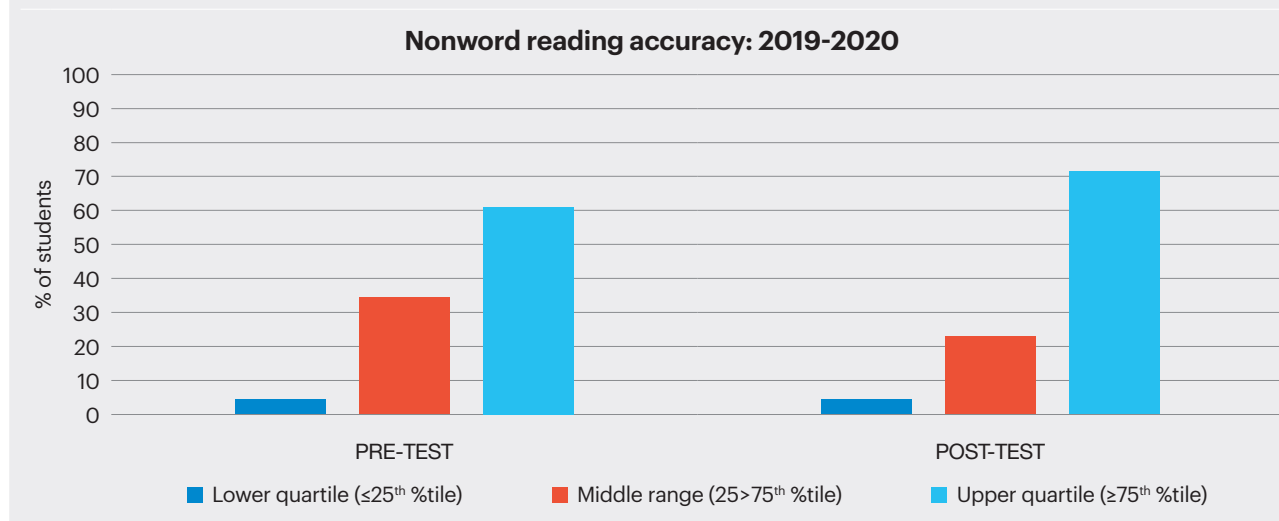
Figure 3. Percentage of students scoring in the lower quartile, middle range and upper quartile for nonword reading accuracy, at pre-test and at post-test.

Nonword reading accuracy: 2019-2020

Legend: ■ Lower quartile (≤25th %tile)　■ Middle range (25>75th %tile)　■ Upper quartile (≥75th %tile)

As may be seen in Figures 1-3, the proportion of students performing in the lower quartile (i.e., percentile rank scores equal to or below 25) appear to shift into the middle range (i.e., percentile rank scores between 25 and 75) and upper quartile (i.e., percentile rank scores equal to or above 75), from pre- to post-test time points. This pattern suggests that InitiaLit–2 students were, on the whole, attaining these reading skills at a greater rate than their peers.

In particular, more than 70% of students performed in the upper quartile on the nonword reading measure at post-test (Figure 3). This score represents a reader's ability to decode unfamiliar words accurately, using knowledge of common letter-sound correspondences. This may have been a strength of the schools in this trial, because most of those schools implemented InitiaLit–F and InitiaLit–1 in previous years, and these programs systematically target word-level spelling patterns. InitiaLit–2 targets word-level skills too, though to a lesser extent, as the focus is more so on text reading comprehension. In this context, it is pleasing but not too surprising to note that the students' decoding skills are mostly above average, both at the start of the year and even more so at the end of the year.

In addition to targeting word-level skills, InitiaLit–2 provides instruction and opportunities for practice in sentence- and text-level reading. All of these elements of the program may have resulted in students' improved passage reading accuracy, as seen in Figure 1.

Similarly, reading accuracy and fluency are known to contribute to reading comprehension, and so targeting these areas, as well as vocabulary and language comprehension, may have resulted in the improvements observed in Figure 2. It would be interesting in future to examine how language background factors into students' development of reading comprehension skills, in response to InitiaLit–2 (and InitiaLit–F and –1) instruction. The student populations for the majority of schools involved in the trials comprised a high proportion of English language learners, and it is likely that this had some effect on the results.

Conclusion

This document provides an extended summary of the results from trials of InitiaLit–2. The trials were conducted by the MultiLit Research Unit over a three-year period (2018-2020). The findings indicated that students who received InitiaLit–2 made excellent reading and spelling gains between the start and end of the school year. Based on percentile distribution shifts and reading age equivalent gains, the observed progress exceeded what might be expected based on increasing chronological age and duration of schooling.

References

Neale, M.D. (1999). *Neale analysis of reading ability – 3rd edition.* Australian Council for Educational Research.

Martin, F., & Pratt, C. (2001). *Martin and Pratt Nonword Reading Test.* Australian Council for Educational Research.

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

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

Wheldall, K., & Madelaine, A. (2013). 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*(4), 371–382.

Appendix A. Description of assessment measures

Burt Word Reading Test

The Burt Word Reading Test has a long history and is a measure of single word recognition. The version we employ is based on a standardisation carried out in the early 1980s in New Zealand by the New Zealand Council for Educational Research. While doubts may be expressed over the utility of reading single words in isolation, it remains a robust test especially when used as part of a battery of reading tests. Our experience with this test suggests that it now frequently overestimates reading age. This needs to be remembered when interpreting the results obtained with this test, but it still provides a good reliable measure of relative reading gain over time.

South Australian Spelling Test

Spelling performance is assessed using the South Australian Spelling Test. The revised norms for this simple spelling test provide estimates of spelling age based on a sample of South Australian students. This test assesses spelling performance from age six to over 15 years.

Martin and Pratt Nonword Reading Test

The Martin and Pratt Nonword Reading Test is a measure of phonological decoding skills. Phonological decoding is the ability to match a sequence of letters to its corresponding sounds and is indicative of a student's ability to read novel or unfamiliar words in text. Students who are poor decoders are more likely to rely on contextual information when they read. The use of nonwords in a test such as this allows for the detection of those students who are largely relying on compensatory strategies rather than generative decoding strategies when attempting to read. The Martin and Pratt Nonword Reading Test consists of pseudowords, which range from simple three letter nonwords to more difficult multisyllabic nonwords. This test assesses performance between the ages of six to 16 years.

Neale Analysis of Reading Ability – 3rd ed.

The widely used Neale Analysis of Reading Ability (3rd edition) provides global indicators of performance in two of the main skills involved in reading: reading accuracy and reading comprehension. Reading accuracy refers to how well a child can identify and accurately pronounce words when presented in written form, either as isolated words or in the context of text, in this case a short story. Reading comprehension refers to how well a child has understood what he or she has read, and this is assessed by presenting the child with questions about what has happened in the story. The Neale Analysis measures, and provides reading-age estimates and age-group comparisons for, both reading accuracy and reading comprehension.

Wheldall Assessment of Reading Passages

The WARP is a curriculum-based measure of reading and is used to measure oral reading fluency. Reading fluency refers to how rapidly a child can read words correctly – to be a competent reader, one needs to be able to read at a reasonable speed. The student is asked to read three 200-word stories, each with a one-minute time limit. Errors are subtracted from the total number of words read to give the total number of words read correctly per minute for each story, and the average is then calculated. The WARP has repeatedly been shown to be both highly reliable and valid.