LanguageLift Extended program trial summary

In 2021, the MRU conducted program trials to determine the effects of the program on students' language development. The students involved in these trials received a version of the program that was still under development, although the instructional content was similar to that in the final version. Results from trials are detailed in the LanguageLift teaching manual. This extended program trial summary serves to outline these results.

Program trial outcomes

- **Oral language gains:** Over the course of the intervention trial, participating students made statistically significant and substantial gains on all of the assessed areas of oral language (see the results section below for more details).

- **Oral language gains compared to standard scores:** Over the course of approximately six months, participating students experienced accelerated growth in average standardised scores, making improvements beyond what would be expected in a six-month timeframe. However, there was one exception, the CELF-5 Sentence Comprehension subtest, on which the Year 1 and 2 cohort's average improvement was equivalent to the average intervention duration.

- **More students moved out of the lowest quartile:** The goal of improving the language skills of students with difficulties was achieved, as demonstrated by a visible shift in the proportion of students performing at or below the 25th percentile from pre- to post-test on all language measures.

- **Social, emotional and behavioural skills:** A survey conducted before and after the completion of LanguageLift instruction revealed statistically significant and positive changes in classroom communication behaviours. These improvements were observed in social, emotional and behavioural sub-areas.

The program trial

Trial schools and participating students

Information about the schools involved in the trials 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). Almost half of the general student population (across schools) had a language background other than English. The mean age of all students at pre-test was 5 years 10 months. The trial sample consisted of 45 students from five schools in Western Australia. This total comprised 22 students in Foundation, 15 students in Year 1, and eight students in Year 2.

Due to the constantly evolving threat of pandemic-related school closures, as well as school staff capacity shortages impacting delivery, post-intervention testing was conducted earlier at some schools than others. Two schools were only able to deliver 25 and 35 lessons respectively. The remaining schools delivered an average of 51 lessons. In most cases, the duration between pre- and post-test time points was around two school terms (see Table 1 for details).

MultiLit

Table 1. Descriptive statistics about research sample

School	Pre/post Duration	Lessons delivered	*n* in each grade	Tot. *n*	Av. age	Gender	ICSEA	% LBOTE*
1	15–16 weeks	35	Y1 = 7 Y2 = 4	11	6;5y	F = 4 M =7	Average	55%
2	19 weeks	49	Y1 = 4	4	6;2y	F = 2 M = 2	Average	75%
3	18 weeks	25	YF = 12	12	5;2y	F = 4 M = 8	Above average	15%
4	YF: 22 weeks Y1 (Group 1): 25 weeks Y1 (Group 2): 27 weeks	58	YF = 6 Y1 = 4	10	5;8y	F = 1 M = 9	Average	75%
5	19 weeks	46	YF = 4 Y2 = 4	8	6;1y	F = 4 M = 4	Above average	25%
			Tot. YF = 22 Tot. Y1 = 15 Tot. Y2 = 8	45	Av. = 5;10y	Tot. F = 15 Tot. M = 30	1074	46%

Note: *Rounded to the nearest 5% to preserve schools' anonymity. ICSEA = Index of Community Socio-educational Advantage. LBOTE = Language Background Other Than English. Av. = average. Tot. = total.

Participant selection and evaluation measures

Students were screened into LanguageLift on the basis of their results on:

- Foundation students – Wheldall Sentence Comprehension Screener (WSCS; Wheldall et al., 2022)
- Year 1 and 2 students – TILLS Student Language Scale (TILLS SLS; Nelson et al., 2015)

For the purposes of the trial, to reduce possible influence from confounding factors, students who had been exposed to English for less than 12 months, who had a diagnosis of a childhood developmental disorder affecting language development (e.g., Autism Spectrum Disorder) or who were already accessing language-focussed speech pathology intervention were not included in the program.

To determine whether students made gains over the course of program instruction, the following language literacy skills (depending on students' age) were assessed after students had received a sufficient number of LanguageLift lessons (preferably close to 50 lessons):

- Receptive vocabulary and narrative comprehension and production (Peabody Picture Vocabulary Test, 5th ed.; PPVT-5; Dunn, 2018; and Test of Narrative Language, 2nd ed.; TNL-2; Gillam & Pearson, 2017)
- Grammatical competence
    - Students in Foundation (subtests of the Clinical Evaluation of Language Fundamentals – Preschool, 2nd ed.; CELF P-2; Wiig et al., 2006)
    - Students in Years 1 and 2 (subtests of the Clinical Evaluation of Language Fundamentals, Australian and New Zealand 5th ed.; CELF-5 A&NZ; Wiig et al., 2017)
- Reading comprehension – students in Year 2 only (Neale Analysis of Reading Ability, 3rd ed.; NARA-3; Neale, 1999).
- Social emotional and behaviour functioning (MultiLit-developed survey)

The guidelines accompanying these tests for identifying children at potential risk of language difficulty were used to define a set of children who then completed the LanguageLift Placement Test. Results of the Placement Test were used to establish the final group of students who would benefit from working on the specific skills targeted in the program.

To examine the difference between pre- and post-test outcomes, assessment data from all 45 students were collated and statistically analysed together. The main assessment measures used to evaluate the program's efficacy were related to oral language, since this is what LanguageLift is intended to directly target. Results of the reading and pragmatic communication assessments are discussed below as secondary outcome measures.

Did language skills improve over the duration of instruction?

To determine whether students showed oral language improvements over the course of the intervention, the differences between pre- and post-test raw scores were first examined. As shown in Table 2, the students made statistically significant gains (with p-values <0.05) on all the assessed areas of oral language. Based on the effect sizes (Cohen's *d*), these gains were also substantial.

Table 2. Raw score means (and standard deviations) and the resultant gains on language assessment measures

Assessment measure	*n*	Pre-test Raw score (SD)	Post-test Raw score (SD)	Gain			
				Raw score (SD)	*t*	*p*	Cohen's *d*
CELF P-2 Sentence Structure	22	13.82 (3.05)	16.86 (2.92)	3.05 (2.89)	4.95	**<.001**	1.06 (L)
CELF P-2 Word Structure	22	11.32 (5.10)	15.23 (4.25)	3.91 (3.31)	5.54	**<.001**	1.18 (L)
CELF-5 Sentence Comprehension	23	18.61 (4.01)	20.78 (3.42)	2.17 (3.30)	3.16	**.005**	0.66 (M)
CELF-5 Word Structure	23	16.91 (5.90)	23.48 (5.28)	6.57 (3.89)	8.09	**<.001**	1.69 (L)
TNL-2 Comprehension	45	13.69 (6.92)	20.22 (6.20)	6.53 (3.77)	11.63	**<.001**	1.73 (L)
TNL-2 Production	41*	17.37 (8.29)	25.24 (10.88)	7.88 (7.70)	6.56	**<.001**	1.02 (L)
PPVT-5	45	100.11 (22.48)	113.20 (19.83)	13.09 (14.52)	6.05	**<.001**	0.90 (L)

Note: When interpreting Cohen's *d* effect sizes, a small (S) effect is 0.2; a medium (M) is 0.5; and a large (L) effect is 0.8 (although see Kraft (2020) for less conservative interpretations based on educational interventions). *4 students were excluded because they did not respond to at least one Production subtest item at pre-test.

Did language skills improve compared to standard scores?

The program trials involved only participants who were receiving LanguageLift (i.e., there were no comparison or control groups). As such, there is no experimental evidence to answer the question of how the cohort's progress compared with age-based expectations. However, to see whether the language skills of students involved in the research trial also improved *beyond* what might be expected given the (approximately) six-month duration between pre- and post-test time points, we can look at the change in average standardised scores (i.e., age equivalent, percentile, scale and standard scores) for each assessment measure.

The results presented in Table 3 indicate that students who received LanguageLift (over the course of approximately six months) experienced accelerated growth in most skill areas beyond what would be expected in a six-month timeframe. As indicated in the table, the improvement in skills was significant and large. One exception was the CELF-5 Sentence Comprehension subtest, on which the Year 1 and 2 cohort's average improvement was equivalent to the average intervention duration (see below for more on this).

MultiLit

Table 3. Standardised score means (and standard deviations) and the resultant gains on language assessment measures

Assessment measure	*n*	Pre-test Raw score (SD)	Post-test Raw score (SD)	Gain			
				Standard score (SD)	*t*	*p*	Cohen's *d*
CELF P-2 Sentence Structure	22	7.68 (2.28)	9.18 (2.75)	1.50 (2.18)	3.23	**.004**	0.69 (M)
CELF P-2 Word Structure	22	6.82 (3.08)	7.91 (2.64)	1.09 (2.09)	2.45	**.023**	0.52 (M)
CELF-5 Sentence Comprehension	23	7.57 (1.88)	8.17 (2.72)	0.61 (2.44)	1.19	NS	0.25 (S)
CELF-5 Word Structure	23	5.74 (1.89)	8.26 (2.77)	2.52 (2.13)	5.68	**<.001**	1.18 (L)
TNL-2 Comprehension	45	7.82 (2.28	9.56 (1.94)	1.73 (1.71)	6.80	**<.001**	1.01 (L)
TNL-2 Production	41*	7.80 (1.85)	8.95 (2.20)	1.15 (2.30)	3.20	**.003**	0.50 (M)
PPVT-5	45	89.31 (11.75)	93.31 (11.93)	4.00 (9.14)	2.94	**.005**	0.44 (M)

Note: Refer to Note under Table 1. NS = non-significant. Yellow shading indicates non-parametric distribution (Shapiro-Wilk *p* < .05). For this measure, a non-parametric significance test was used to supplement the results of a paired samples *t*-test (Wilcoxon Signed Ranks Test *p* < .001). Scale scores (where 'average' performance = 10) were used for all CELF and TNL-2 tests; standard scores (where 'average' performance = 100) were used for the PPVT-5.

Finally, the results were examined to determine whether students shifted out of the bottom quartile (i.e., 25%) according to the percentile scores for each language measure. This is an important question, because the goal of delivering LanguageLift is to improve the language skills of those students with difficulties, thereby shifting them closer to or within the 'average' range for their age. As shown in Figure 1, this goal was achieved in the program trials. On all language measures, the proportion of students performing at or below the 25th percentile at post-test is lower than the proportion at pre-test.

Figure 1. Proportion of students scoring in the bottom quartile on language measures at pre- and post-test
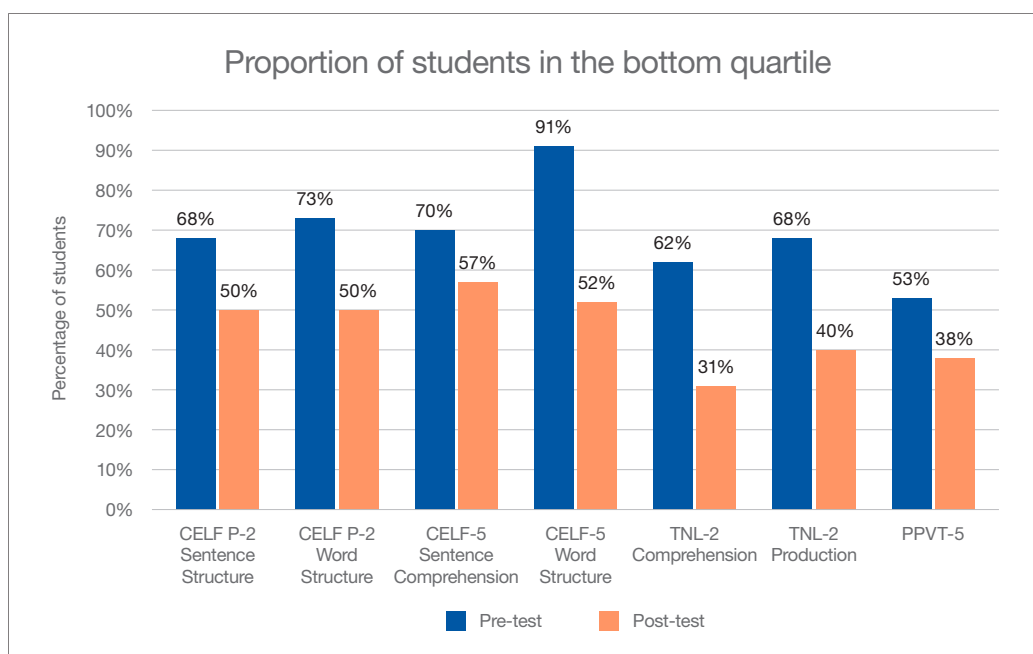
In summary, the results from analyses of raw, scale, standard, age equivalent and percentile scores reveal that students involved in the LanguageLift trial improved in their oral language skills from pre- to post-test. It is likely that these improvements can be tied directly to the three key areas targeted in LanguageLift:

▸ **Vocabulary:** The PPVT-5 results indicated significant improvements in vocabulary, a skill which is explicitly targeted in the program using rich instruction methods. The fact that on average, scores improved on a standardised test of vocabulary not linked to the specific words taught in the program is striking. Because not every word a child needs to know can be taught explicitly, one of the aims of rich vocabulary instruction is to stimulate more general vocabulary growth by strengthening and extending semantic networks, and by encouraging children to pay attention to words themselves. As words become increasingly well-connected in a child's mind, deducing and acquiring the meanings of other unfamiliar words independently should become easier. Our PPVT-5 results provide some support for this hypothesis.

▸ **Grammar:** The CELF P-2 and CELF-5 Word Structure subtests indicated that all students made significant improvements in word-level grammar (e.g., use of verb suffixes like 'ing'). The CELF P-2 Sentence Structure subtest results also indicated that Foundation students made significant improvements in sentence-level grammar (e.g., understanding simple sentences like 'The boy has a ball'). These skills are heavily targeted in the early part of the program using explicit word and sentence-building instruction supported by icons, sentence boards and other visual aids, along with frequent teacher modelling and opportunities for child production of structures.

The results of the CELF-5 Sentence Comprehension subtest administered to students in Years 1 and 2 indicated that there was a smaller proportion of students in the bottom quartile at post-test (57%) than at pre-test (70%). Overall, the cohort also showed improvement to a degree that was equivalent to the number of months that passed. Given that the students were entered into the program because they were observed to be falling behind their peers, it is also a pleasing result that the students kept up with age-based norms. In addition, perhaps growth was less rapid here because, in the later part of the program though sentence-level skills continue to be addressed explicitly, scaffolding is reduced and the focus switches to retelling stories using a range of different sentences.

▸ **Story skills:** The TNL-2 results indicated that all students made significant improvements in their story retell and comprehension abilities. These skills were targeted in the program through repeated exposure to and discussion of various stories and their narrative structure. Students were also supported in applying their knowledge of narrative structure to retell stories themselves. It is pleasing to have seen such large gains in this area, as this is a particular focus of the program.

Did social, emotional and behavioural skills related to communication improve over the duration of intervention?

To evaluate whether social, emotional and behavioural skills related to communication also improved in response to LanguageLift, teachers of students participating in the program were given a survey to complete at pre- and post-test[1]. Teachers were asked to indicate on the survey form how well, on a scale from 1 to 7 (i.e., 'not well' to 'very well'), they thought a number of given statements described the child. There were 13 questions (four tapping social adjustment, four tapping emotional adjustment, four tapping behavioural adjustment and one tapping overall confidence). Higher scores represented more positive communication behaviours.

---

[1]A child-friendly version of the survey was also given to the participating students. However, there was a clear ceiling effect at pre-test, indicating that the students were unwilling or did not know how to respond to the task. The difference between pre- and post-test scores on this survey was not statistically significant.

Survey results were available for 44 students. The average (mean) score at pre-test was 56.23 (SD = 17.42) and at post-test, the average (mean) score was 69.34 (SD = 9.76). The difference between these scores was statistically significant (t = 5.99, p < .001). The distributions of scores at pre- and post-test are shown in Figure 2. These results indicate students made significant positive changes in their classroom communication behaviours. A further analysis of scores in the three sub-areas of social, emotional and behavioural adjustment showed that improvements were significant across all three areas. Given the strong relationship between language abilities and social, emotional and behavioural functioning in a school context (Snow, 2016), the findings may be attributable to students' participation in LanguageLift. Because positive behaviour management is an aspect of LanguageLift, it is also possible that the improvements in student functioning may have been due to this aspect of the intervention (along with the extra attention afforded to children in a small group) rather than their language gains.

Figure 2. Box and whisker plots showing distributions of scores at pre- and post-test
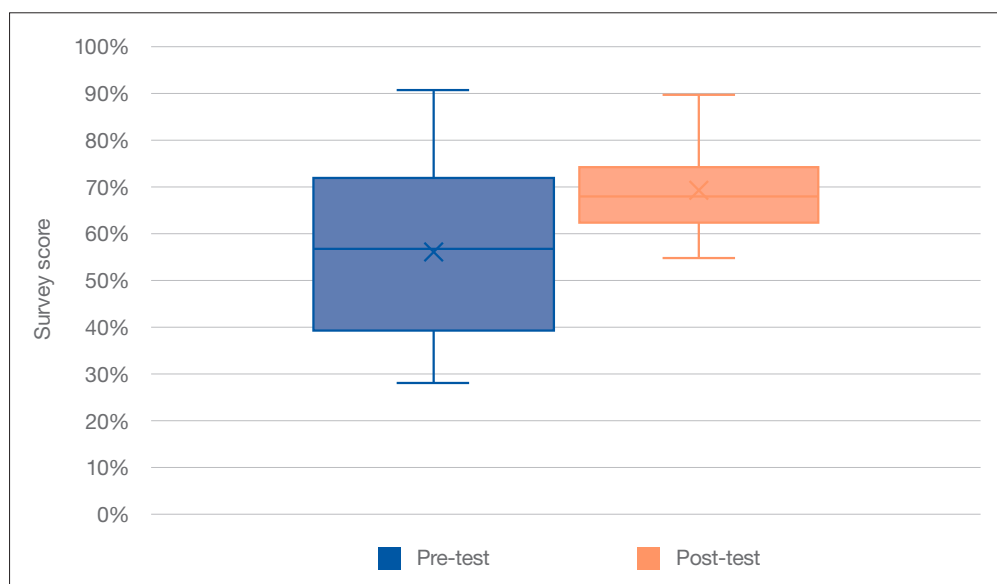
Note: 'Whiskers' above and below the box mark maximum and minimum scores respectively; upper box boundary marks the 75th percentile; lower box boundary marks the 25th percentile; horizontal line within box boundary marks the median; cross (×) marks the mean.

Did reading skills improve over the duration of intervention?

As oral language skills are strongly correlated with reading, a passage reading task was administered to students in the trial who were in Year 2 (*n* = 8). Expectations around how students would perform on this outcome measure were tentative, since the sample size was small[2] and written language skills were not directly targeted in LanguageLift.

Based on the results of the NARA-3, the students improved at a similar rate to what would be expected for the duration of instruction (seven months gain in reading accuracy and six months gain in reading comprehension). This steady gain stands in contrast to the large oral language improvements discussed previously, and serves to highlight the difference observed in skills directly targeted by the program, versus those only indirectly targeted. Nonetheless, it is pleasing that students with low oral language skills on commencement of intervention did not fall further behind age-based norms for reading comprehension during the intervention period. Although this small sample size does not currently speak to a flow-on effect to reading skills, it remains possible that LanguageLift might be shown to indirectly and positively affect reading comprehension over a longer duration between pre- and post-testing.

_____

[2]Twelve additional Year 2 students were lost to the trial due the extended NSW school closures in 2021 attributed to the COVID-19 pandemic.

Conclusion

Results from this 2021 trial clearly demonstrate that students make progress when participating in LanguageLift. Students in Foundation, Year 1 and Year 2 with language difficulties showed significantly improved oral language skills following their participation in the program. According to their teachers, the students also showed significantly more positive pragmatic communication behaviours in the classroom. Ultimately, and in alignment with the Simple View of Reading, reading comprehension improvements may result from participation in the program, although this could not be determined unequivocally in the present trial.

References

Dunn, D. M. (2018). *Peabody picture vocabulary test* (5th ed.). Pearson.

Gillam, R. B., & Pearson, N. A. (2017). *Test of narrative language* (2nd ed.). Pro-ed.

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

Nelson, N., Plante, E., Helm-Estabrooks, N., & Hotz, G. (2015). *TILLS Student Language Scale.* Brookes Publishing.

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*(3), 216–228. https://doi.org/10.3109/17549507.2015.1112837

Wheldall, K., Bell, N., & Notley, A. (2022). *The Wheldall Sentence Comprehension Screener.* MultiLit Pty Ltd.

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical evaluation of language fundamentals* (Australian and New Zealand 2nd ed.; CELF P-2 A&NZ). Pearson.

Wiig, E. H., Semel, E., & Secord, W. A. (2017). *Clinical evaluation of language fundamentals* (Australian and New Zealand 5th ed.; CELF-5 A&NZ). Pearson.