Preliminary data demonstrating student progress with LanguageLift

Background

The development of LanguageLift began in 2019. In this first year and throughout 2020, the program underwent field testing in several Australian schools. During these years, the program was delivered by members of MultiLit's product development team, as well as classroom teachers. Observations and feedback given by the teachers were collected, and these helped in adapting the program to work in real-life classrooms.

In 2021, a more rigorous trial of the program was conducted by the MultiLit Research Unit, in order to determine what effects the program had on students' language development. The trial sample consisted of 45 students from five schools, all located in the Australian city of Perth. (A further 40 students from two schools in New South Wales were recruited for the trial but were forced to withdraw following extended school closures attributed to the COVID-19 pandemic.) The level of socio-educational advantage in the general student populations of the five schools (as measured by comparing to an average rating of 1000 on the Index of Community Socio-educational Advantage; ICSEA) 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.

All students in the first three years of school at the five participating schools were screened on their language skills. Two different screeners were used:

- Wheldall Sentence Comprehension Screener (WSCS; Wheldall et al., 2022)
- TILLS Student Language Scale (TILLS SLS; Nelson et al., 2015)

Students in the first year of school (hereafter referred to as 'Foundation') were assessed using the WSCS, suitable for children up to age five, while Year 1 and 2 students were assessed using the TILLS SLS, suitable for use with children aged six years and above. 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.

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-focused speech pathology intervention were not included in intervention groups.

Forty-five students (15 females; 30 males) were identified across the five schools for participation in the program. This total comprised 22 students in Foundation, 15 students in Year 1, and 8 students in Year 2. On average, the age of the cohort was 5 years, 10 months (i.e., 5;10).

These students' oral language skills were assessed before and after they had received a sufficient number of LanguageLift lessons (preferably close to 50 lessons) to see whether their language skills improved after receiving the program. 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).

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.

Assessment measures

The assessment measures administered to each student at pre- and post-test depended on their year level.

- To measure receptive vocabulary and narrative comprehension and production, all students received the Peabody Picture Vocabulary Test (5th ed.; PPVT-5; Dunn, 2018) and the Test of Narrative Language (2nd ed.; TNL-2; Gillam & Pearson, 2017).
- To measure grammatical competence, younger students in Foundation completed subtests of the Clinical Evaluation of Language Fundamentals – Preschool (2nd ed.; CELF P-2; Wiig et al., 2006) appropriate for children aged three to six, while older students in Years 1 and 2 completed subtests of the Clinical Evaluation of Language Fundamentals (Australian and New Zealand 5th ed.; CELF-5 A&NZ; Wiig et al., 2017), appropriate for children from age five.

- Assessing reading comprehension accurately in children in the first two years of school is difficult. To measure reading comprehension in our study, only children in Year 2 were given the Neale Analysis of Reading Ability (3rd ed.; NARA-3; Neale, 1999).

All students' teachers were also asked to complete a survey to assess the children's social, emotional and behavioural functioning pre- and post-intervention. A summary of the language assessment measures used is given in Table 2.

Table 2. List of language assessment measures

Domain	Grade	Assessment measure	Skills measured
Oral language	F	CELF P-2 Sentence Structure and Word Structure subtests	Sentence comprehension, grammar and morphology
	1–2	CELF-5 Sentence Comprehension and Word Structure subtests	Sentence comprehension, grammar and morphology
	F–2	TNL-2	Narrative comprehension and production
		PPVT-5	Receptive vocabulary
Reading	2	NARA-3	Reading accuracy and comprehension
Pragmatic communication	F–2	Teacher survey	Social, emotional and behavioural functioning

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 track student progress 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.

Results

Did language skills improve over the duration of intervention?

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 (the values in the 'Raw score (SD) Gain' column of Table 3). As shown in the bolded column 'p', the students made statistically significant gains (with p-values <0.01) on all the assessed areas of oral language. Based on the effect sizes (Cohen's *d*), these gains were also substantial.

Table 3. 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.

To evaluate whether the language skills of students involved in the program trial also improved *beyond* what might be expected given the (approximately) 6-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.

With respect to age equivalent scores, the students in the trial made gains between pre- and post-test time points that were, on average, equivalent to:

- 12 months on the CELF P-2 Sentence Structure subtest
- 11 months on the CELF P-2 Word Structure subtest
- 6 months on the CELF-5 Sentence Comprehension subtest
- 18 months on the CELF-5 Word Structure subtest
- 13 months on the TNL-2 Comprehension score
- 12 months on the TNL-2 Production score
- 8 months on the PPVT-5

These results 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 time frame. 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).

A similar pattern may be observed in Table 5, which shows the results from statistical analyses examining the difference between pre- and post-test scale or standard scores. In most cases, the improvement between average pre- and post-test scores was significant, as shown by p-values <0.05. Moreover, these gains were substantial, showing medium to large effect sizes. Note, however, that this was not the case for the CELF-5 Sentence Comprehension subtest.

Table 4. Standardised score means (and standard deviations) and the resultant gains on language assessment measures

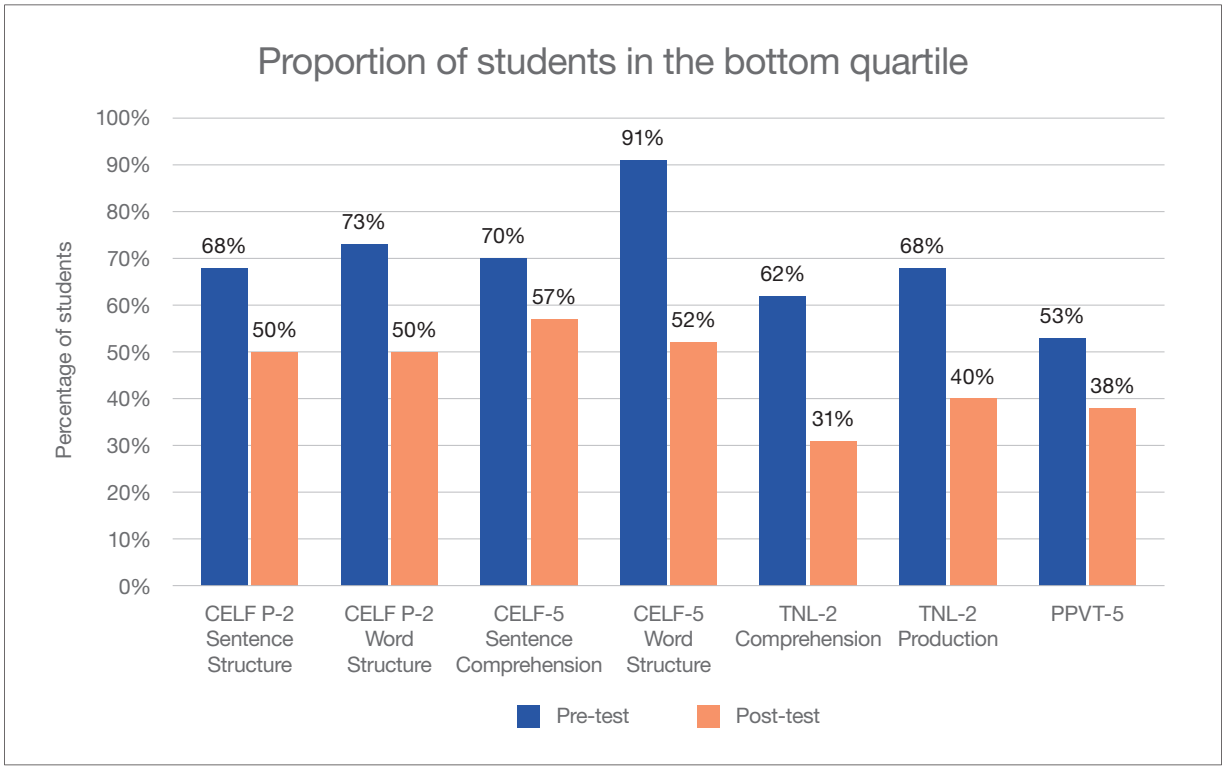
				Gain			
Assessment measure	***n***	**Pre-test Raw score (SD)**	**Post-test Raw score (SD)**	**Standard score (SD)**	***t***	***p***	**Cohen's *d***
CELF P-2 Sentence Structure	22	7.68 (2.28)	9.18 (2.75)	1.50 (2.18)	3.23	**.004**	0.69 (M)
CELF P-2 Word Structure	22	6.82 (3.08)	7.91 (2.64)	1.09 (2.09)	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. On all language measures, the proportion of students performing at or below the 25th percentile at post-test is visibly 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 a pleasing result that the students kept up with age-based norms. Growth may have been less rapid here, because in the later part of the program, while 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.

  An example of this progress can be clearly seen in these two versions of the same story used in the LanguageLift Placement Test, told by a Year 2 child after hearing an adult tell them the story, before and after participating in intervention. The child had not heard this story in the intervening period.

Figure 2. Example story retells before and after LanguageLift intervention

February, 2021

Um he's making pancakes for Father's Day and … and he … he was trying to make pancakes for his dad And when he finished making the pancakes his dad said, "Thank you."

August, 2021

One morning George tiptoed out of his bed, and it was Father's Day. He wanted to make his dad some pancakes. Then he got the eggs and butter. Then … and then … then the flour, and then the flour fell onto the floor. Then he decided to make …. He decided to make bread with eggs. After that he made, he gave …. He walked into his dad's bedroom and then he gave his dad the pancakes and eggs and his dad said, "Thank you, this looks scrumptious."

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] The survey was designed specifically for the program trial. Teachers were asked to indicate on the survey form how well, on a scale of 1–7 (i.e., 'not well' to 'very well'), they thought a number of given statements described the child (see Figure 3). There were 13 questions (four tapping social adjustment, four tapping emotional adjustment, four tapping behavioural adjustment and one tapping overall confidence) so possible scores ranged from 13 to 91. The questions were phrased such that higher scores represented more positive communication behaviours.

Figure 3. Example questions on classroom teacher survey

On a scale of 1–7, where 1 is 'not well' and 7 is 'very well,' how well do you feel this child:

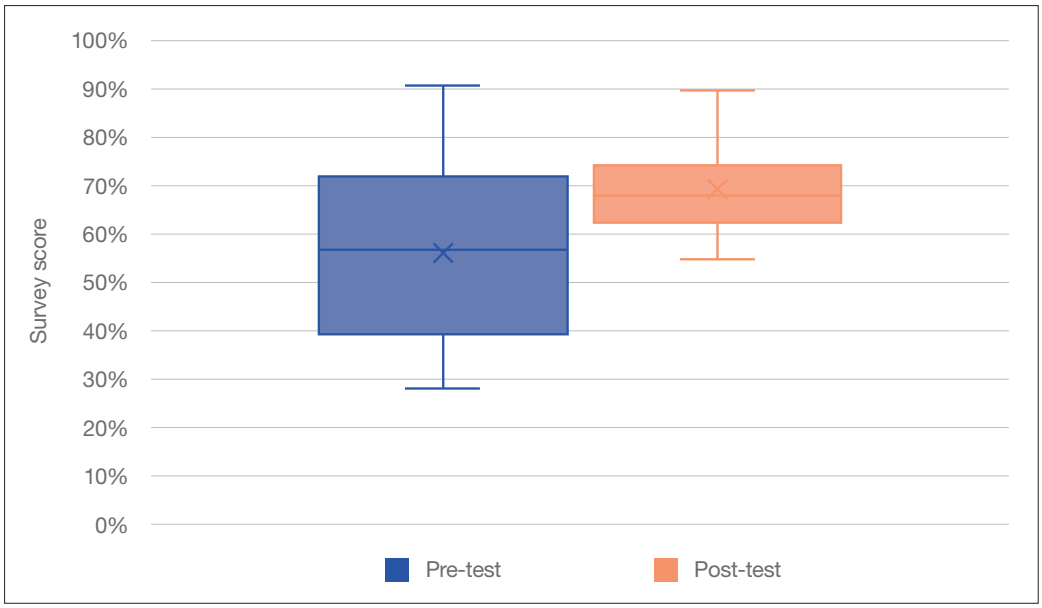
1.	Remains on task during paired, group and whole-class activities?						
	1	2	3	4	5	6	7
2.	Manages emotions appropriately throughout the day?						
	1	2	3	4	5	6	7

Of the 45 students who participated in the trial, 44 had teacher survey results available for analysis (one student's form was not returned). 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 4.

Based on the results of the teacher survey, students showed 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.

[1] A child-friendly version of the survey was also given to the participating students. However, the children tended to rate all survey statements as accurate descriptions of themselves, indicating that they 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.

Figure 4. 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?

Oral language skills are strongly correlated with reading. As such, 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.

The average NARA-3 Reading Accuracy raw scores at pre- and post-test were 25.00 (SD = 15.47) and 33.63 (SD = 16.00), respectively. The average NARA-3 Reading Comprehension raw scores at pre- and post-test were 5.50 (SD = 4.81) and 9.00 (SD = 3.46), respectively. The improvements from pre- and post-test were equivalent to:

- seven months on the NARA-3 Reading Accuracy score
- six months on the NARA-3 Reading Comprehension score.

Based on these results, the students improved at a similar rate to what would be expected for the duration of instruction. 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.

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

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. This expectation is based on the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), which states that reading comprehension is the product of two main areas: word recognition and language comprehension. By targeting the latter area with children who have language difficulties, it is logical that improved reading comprehension should result at some point. (Similarly, by targeting word recognition with children who have difficulties in *that* area, improved reading comprehension should also result.) In future, a longitudinal investigation involving a larger sample of Year 2 LanguageLift participants would be useful in teasing apart the indirect effects that this instructional program has on written language skills.

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.