Preliminary data demonstrating progress with SpellEx

Background

The development of SpellEx Part A began in 2020, and the program underwent field testing in several Australian schools throughout 2021. This process proved quite challenging given the lockdowns, attendance issues and visitor restrictions in place that were all associated with the COVID-19 pandemic. Nevertheless, program content was delivered by classroom teachers over this time, and the schools involved were supported as much as possible by MultiLit's Product Development team. It should be noted that most schools were not able to complete SpellEx Part A in its entirety in 2021, primarily due to the pandemic-related disruptions. These schools continued the program with students as they entered Year 4 in 2022. Teacher observations and feedback were collected over the 2021 school year, as were spelling assessment data pertaining to students receiving the program. This information helped the Product Development team adapt the program to work in real-life classrooms, gauge how successful SpellEx Part A was in meeting its teaching objectives and establish which assessment measures were most suitable for further trialling.

In 2022, SpellEx Part A entered a second year of trialling. During this year, data for a new cohort of students were collected and analysed by the MultiLit Research Unit, to assess the progress made by students receiving the program. As in 2021, schools were able to deliver most (but not all) of the SpellEx Part A program over the year, again primarily due to pandemic-related disruptions. The results from the Year 3 students in this 2022 trial are detailed below.

Trial schools

Schools were specifically selected for the trial to cover different geographical locations across five Australian states (New South Wales, South Australia, Queensland, Western Australia and Victoria), as well as a range of governing school sectors, levels of socio-educational advantage and student language backgrounds. This heterogeneity was considered important for evaluating the outcomes of students from a variety of real-world contexts. In total, there were 13 schools and 604 Year 3 students for whom data on at least one assessment measure were available. On average, the students in the cohort were 8 years and 3 months old at the beginning of the school year. Information about the schools and students is given in Table 1.

The assumed socio-educational and language backgrounds of students involved in the trial (based on each school's general student population) varied substantially between sites. Most had Index of Community Socio-Educational Advantage (ICSEA) scores that were average (900–1100) or above average (>1100), with one school population in the below average range (<900). Some schools had a very low proportion of students with a Language Background Other Than English (LBOTE), while for other schools, these students represented the majority. Many of the 13 schools had implemented at least one year of InitiaLit instruction prior to 2022, indicating that these students had a foundation of systematic synthetic phonics instruction on which to refine their spelling abilities.

Table 1. Descriptive statistics about Year 3 sample.

School	State	ICSEA	% LBOTE[1]	Assessed on	*n* classes[2]	% of sample[3]
1	NSW	Average	15	SAST/CoST	2	7
2	NSW	Above average	50	SAST/CoST	3	11
3	NSW	Above average	30	SAST/CoST	1	3
4	NSW	Above average	30	SAST	4	15
5	NSW	Average	90	SAST/CoST	1	3
6	NSW	Above average	60	SAST	3	10
7	NSW	Above average	15	SAST	2	4
8	SA	Average	5	SAST	1	3
9	Qld	Average	15	SAST/CoST	5	18
10	Qld	Above average	45	SAST/CoST	3	10
11	WA	Below average	50	SAST	1	1
12	WA	Average	25	SAST/CoST	4	10
13	Vic	Average	5	SAST	3	4

Note: ICSEA = Index of Community Socio-Educational Advantage; LBOTE = Language Background Other Than English; SAST = South Australian Spelling Test; CoST = Components of Spelling Test

[1]Rounded to the nearest 5% to preserve schools' anonymity.
[2]Includes Year 3 students from Year 3–4 composite classes (1 class in Schools 1 and 2; 3 classes in School 13).
[3]Rounded to the nearest unit (i.e., not equal to exactly 100%).

Assessment measures

To determine how students receiving SpellEx Part A progressed in their spelling development over the course of a school year, they were assessed in February and November of 2022 (hereafter 'pre-test' and 'post-test', respectively). The assessments included Form B of the South Australian Spelling Test (SAST) (Westwood, 2005) and the Components of Spelling Test: Real-word version (CoST) (Daffern, 2018).

Both assessments are standardised measures of spelling achievement, which means they contain stimuli that may not be targeted in the SpellEx Part A program. In this way, they differ to SpellEx's in-program assessments, which measure how well students have mastered the spelling patterns and conventions that are taught within SpellEx Part A's scope and sequence. As such, the SAST and CoST may be considered to represent the success with which students generalise strategies they have learned in the program when spelling unfamiliar words.

The SAST and CoST were both administered to whole classes of students at a time. The SAST was administered using its standard paper record form, while (except for five students) the CoST was delivered and scored using the online version of the test. All assessment administration was conducted by teachers employed at the trial schools. Paper record forms were then posted to the MultiLit Research Unit where they were scored, double-checked and compiled into a database by a number of experienced researchers. Online CoST responses were digitally scored and sent automatically to the MultiLit Research Unit.

Both the SAST and CoST are measures of real-word spelling proficiency, and students complete them by writing or typing the dictated word. The tests differ from one another in the types of scores that are generated for interpretation or analysis.

The CoST is used to generate phonological, orthographic and morphological subscale scores, as well as a raw whole word score, which is based on the student's overall word spelling accuracy. The subscale scores provide insight into the strategies that students can use when spelling, which can help teachers plan for instruction:

- The phonological subscale captures skills in spelling graphemes that consistently correspond to specific phonemes (e.g., the 'm', 'a' and 'sh' in 'mash').
- The orthographic subscale captures knowledge of orthographically legal versus illegal letter sequences (e.g., 'boyl' is not a legal spelling of 'boil' because 'oy' only appears at the end of English words or syllables).
- The morphological subscale captures knowledge of how prefixes (e.g., 're-') and suffixes (e.g., '-ed') are represented in print. All three types of word form awareness contribute to overall spelling ability (Daffern, 2017).

Raw scores from all three subscales can also be converted to standardised scores (i.e., *z*-scores and percentile ranks) that allow for comparison with grade-based norms.

The SAST can only be used to generate a score that is based on the student's word spelling accuracy (like the CoST's whole word score). This raw score can be converted to a spelling age equivalent.

It was not possible for every class to receive both the SAST and CoST due to time and technological constraints. Some data were also incomplete and were therefore excluded from analyses. Of the original 604 students in the sample, complete SAST and CoST data were available for 256 students, while only SAST data were available for 293 students and only CoST data were available for 55 students. These factors resulted in the sample sizes described in the next section.

Did spelling skills improve over the year?

To determine whether students showed improvements in spelling skills over the course of a year's instruction with SpellEx Part A, the differences between pre- and post-test raw scores were statistically analysed and reported as a whole group.

Table 2. Means (and standard deviations) and the resultant gains on measures of spelling (raw scores) for Year 3 students.

Assessment metric	*n*	Pre-test raw score (SD)	Post-test raw score (SD)	Gain			
				Raw score (SD)	*t*	*p*	Cohen's *d*
CoST Whole word score	311	23.39 (11.74)	32.86 (13.10)	9.47 (6.53)	25.597	<.001	1.45 (L)
CoST Phonological subscale	311	19.65 (4.09)	22.19 (4.18)	2.54 (2.82)	15.905	<.001	0.90 (L)
CoST Orthographic subscale	311	16.86 (7.82)	21.95 (6.78)	5.09 (4.26)	21.065	<.001	1.19 (L)
CoST Morphological subscale	311	11.51 (6.67)	18.11 (9.19)	6.60 (5.67)	20.511	<.001	1.16 (L)
SAST	549	35.79 (9.88)	42.99 (9.85)	7.21 (4.43)	38.127	<.001	1.63 (L)

Note: CoST = Components of Spelling Test; SAST = South Australian Spelling Test. 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) is 0.5; and a large (L) effect is 0.8 (although see Kraft [2020] for less conservative interpretations based on educational interventions).

As shown in Table 2, the Year 3 students made statistically significant gains in their CoST and SAST raw scores. Based on the effect sizes, these gains were also substantial (Cohen's $d \geq 0.90$).

The improvements made by students who received the program can also be seen in Figures 1–3. These graphs show the proportion of students in the sample who obtained CoST test scores that were in the lower quartile (i.e., below the 25th percentile), the middle range (i.e., between the 25th and 75th percentile), and the upper quartile (i.e., above the 75th percentile) at pre- and post-test. On all three subscales, there was a visible shift in percentile distributions from pre- to post-test, representing the movement of students from the lower quartile into the middle range and upper quartile.

It should be noted that these results do not represent a shift in spelling ability *relative* to same-aged peers. This is because the percentile ranks have been converted from raw scores based on a normative sample who were assessed at only one time point in the school year (i.e., Term 3). Nevertheless, the graphs do show that the students' *absolute* spelling performances have improved substantially from pre- to post-test.

Figure 1. Percentage of students scoring in the lower quartile, middle range and upper quartile on the CoST Phonological subscale at pre- and post-test.

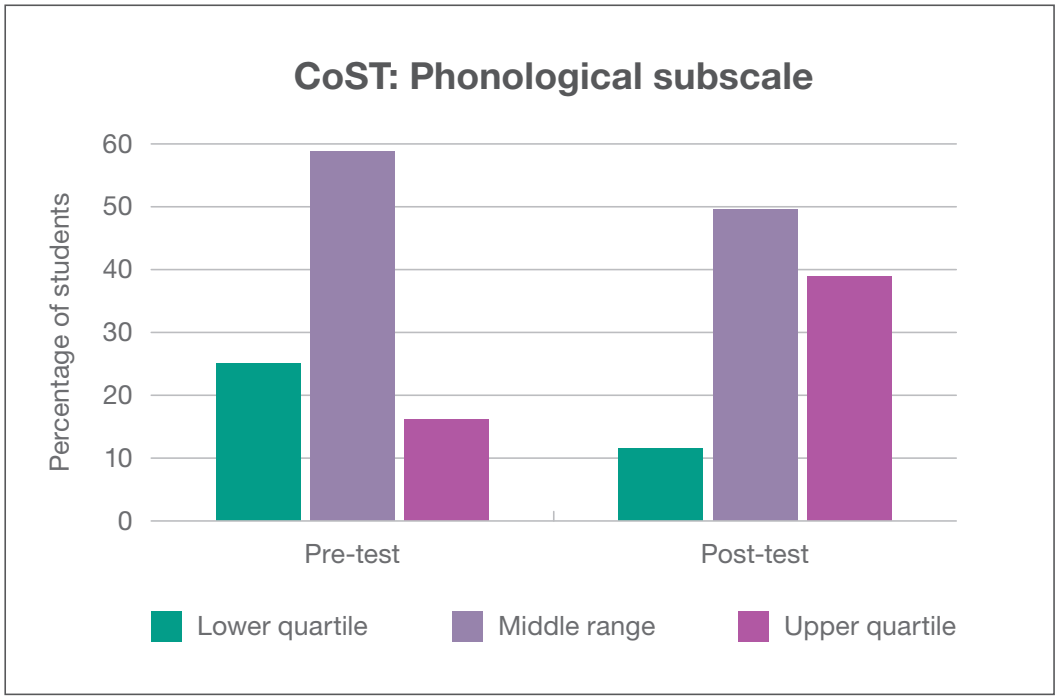


Figure 2. Percentage of students scoring in the lower quartile, middle range and upper quartile on the CoST Orthographic subscale at pre- and post-test.

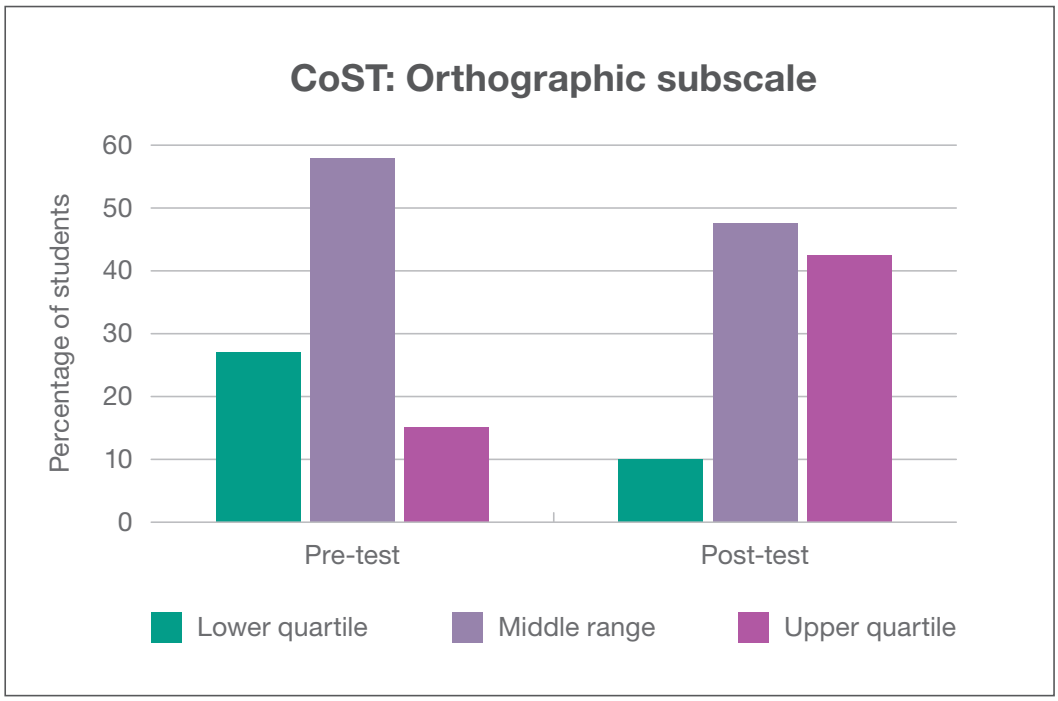
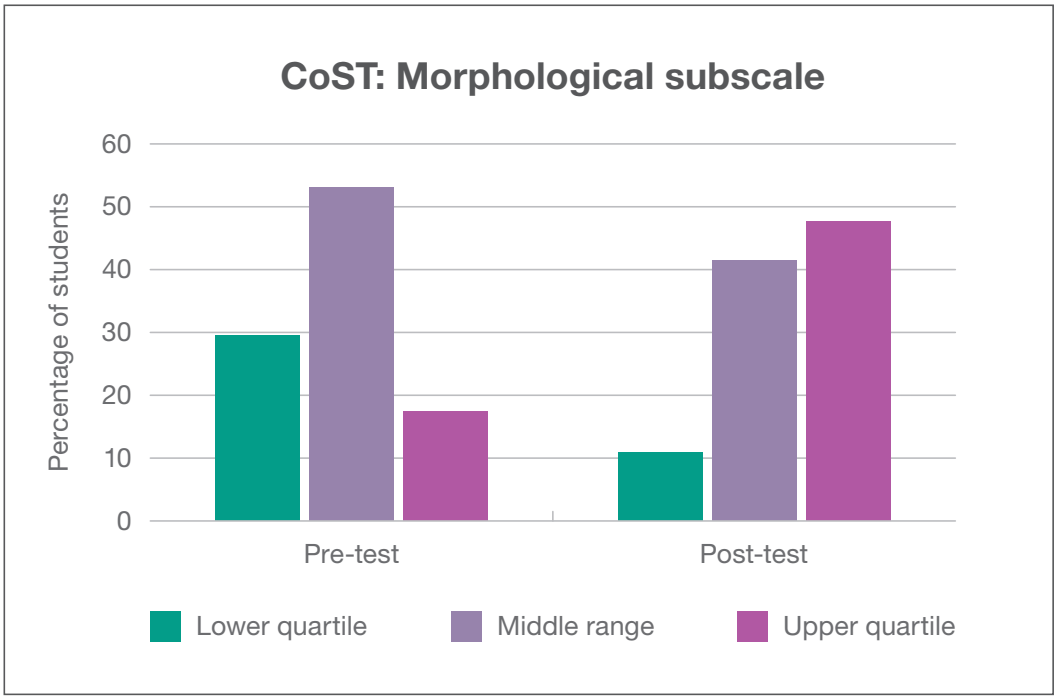


Figure 3. Percentage of students scoring in the lower quartile, middle range and upper quartile on the CoST Morphological subscale at pre- and post-test.



Did spelling skills improve beyond expected growth for one school year of instruction?

The above results indicate that students' spelling skills improved over the course of a school year. Moreover, they improved across all three component skills that are assessed by the CoST. This was pleasing to note, especially in the context in which this trial took place – that is, with potential disruptions to schooling, attendance and socio-emotional functioning resulting from the COVID-19 pandemic. However, some degree of growth over a school year may be considered unsurprising, given that students' language skills are still maturing and they are interacting with written text over this period.

To evaluate whether the spelling skills of students involved in the program trial improved *beyond* what might otherwise be expected in one school year of instruction, the age equivalent scores corresponding with average SAST raw scores were examined. These indicate that, in the 9- or 10-month duration between pre- and post-test time points, the Year 3 students made the equivalent of 17 months progress.

Did students' spelling ability at pre-test factor into how much progress they made?

Another question that the MultiLit Research Unit sought to address was whether observed progress for the overall cohort was due to a particular group of students improving more than others. To answer this question, the students were divided into three groups according to their SAST spelling scores at pre-test. The 'low scorers' group comprised the lowest-scoring 25% of the total sample. The 'high scorers' group comprised the highest-scoring 25% of the total sample. The 'average scorers' group comprised the 50% of students who performed around the sample average.

Table 3. Means (and standard deviations) and the resultant gains on SAST for differently performing groups.

Assessment metric	*n*	Pre-test raw score (SD)	Post-test raw score (SD)	Gain			
				Raw score (SD)	*t*	*p*	Cohen's *d*
SAST Low scorers	146	23.13 (5.80)	31.41 (7.19)	8.28 (5.11)	19.567	<.001	1.62 (L)
SAST Average scorers	275	36.75 (3.28)	43.97 (5.16)	7.22 (4.29)	27.924	<.001	1.68 (L)
SAST High scorers	128	48.15 (4.45)	54.12 (4.58)	5.97 (3.51)	19.231	<.001	1.70 (L)

Note: The sample sizes in each group represent the closest approximation of intended percentages. See additional notes under Table 1.

As shown in Table 2, similarly large effect sizes (i.e., Cohen's *d* values) were observed across subgroups. This suggests that neither high achievers nor students with spelling difficulties are at any sort of disadvantage, in terms of the progress they can be expected to make over the course of a year's instruction with SpellEx Part A.

Conclusion

Results from the 2022 trial of SpellEx Part A showed that Year 3 students receiving the program made excellent gains in their ability to represent written words accurately between the start and end of one school year. As evidenced by the CoST subscale scores, they also improved in their ability to apply phonological, orthographic and morphological knowledge that underpins spelling development. Given that SpellEx explicitly targets each of these areas in its scope and sequence, it is pleasing to observe the translation of this teaching content into observable score increases. Moreover, the degree of progress made did not appear to vary depending on spelling ability at the outset, meaning that SpellEx Part A can be expected to cater to the variation in skill level that exists within classrooms.

In future years, the MultiLit Research Unit intends to conduct more controlled research trials to evaluate program efficacy. Ideally, this would involve comparing spelling performances of students receiving SpellEx Part A with those receiving different spelling instruction. It would also be interesting to explore what effects SpellEx Part A has on students' broader writing (and even reading) development.

Acknowledgements

Program trials like the one described here would not be possible without the schools involved. We would like to sincerely thank each participating student and member of staff who helped to facilitate this trial of SpellEx Part A.