Reading and Writing
https://doi.org/10.1007/s11145-019-09945-8

Using the MULTILIT literacy instruction program with children who have Down syndrome

Lisa Lim[1] · Joanne Arciuli[1] · Natalie Munro[1] · Linda Cupples[2]

Abstract
We evaluated the MULTILIT—"Making Up Lost Time in Literacy"—instruction program for individuals with Down syndrome. The 12-week program was administered on a 1:1 basis to participants who were assessed at three timepoints: Baseline, pre-instruction, and post-instruction. Participants were allocated non-randomly to two groups that had comparable abilities prior to intervention. These groups differed in the duration of the control period between baseline and pre-instruction: 12 weeks for Group 1 ($n=8$, mean age = 168.88 months) and 24 weeks for Group 2 ($n=7$, mean age = 164.71 months). Improvements in literacy skills were evaluated within participants (each participant serving as her/his own control), and the influence of control periods was evaluated between participants. Phonological awareness, word reading accuracy, and word spelling accuracy all improved significantly from pre- to post-instruction, with large effect sizes; whereas no statistically significant changes were observed during the control period (between baseline and pre-instruction assessments) regardless of the duration of this control period.

Keywords Literacy · Instruction · Reading · Spelling · Phonological awareness · Intellectual disability

Introduction

Reading and writing are essential life skills for all individuals, including people with Down syndrome (DS). When it comes to teaching these skills, a practical approach is to focus initially on the ability to read and spell single words accurately (i.e., 'cracking the orthographic code'). While there has been debate about the best way to help children with DS learn to read and spell single words, there is a body

✉ Joanne Arciuli
joanne.arciuli@sydney.edu.au

1 Faculty of Health Sciences, The University of Sydney, Sydney, Australia

2 Department of Linguistics, Macquarie University, Sydney, Australia

Published online: 21 March 2019

of research that supports inclusion of a phonics component. In reading, phonics involves mapping letters onto sounds (decoding), whereas in spelling, sounds must be mapped onto letters (encoding). Other researchers have argued, however, that a sight-word approach is more appropriate for teaching children with DS, because of their well-documented weakness in tasks that rely on phonological short-term memory (i.e., the ability to process and store sound-based information for short periods of time). We provide a brief overview of this research in relation to children with DS before reporting on an intervention study using MULTILIT (2007a)—"Making Up Lost Time in Literacy"—a commercially available instructional program that combines phonics and sight-word reading approaches.

Single word reading and spelling acquisition in DS

In children with typical development, the ability to decode words and nonwords relies on precursor skills such as phonological awareness (PA) (e.g., Oakhill & Cain, 2012; Wagner, Torgesen, & Rashotte, 1994). PA is reflected in the conscious knowledge that spoken sentences, phrases, and words are comprised of smaller phonological units. Some examples of PA include the ability to identify the number of syllables in a word, its onset (initial consonant or consonant cluster) and rime (the middle vowel and final consonant of a monosyllabic word), and the individual phonemes it contains. The ability to reflect on individual phonemes (phonemic awareness) is more strongly associated with alphabetic reading development than other forms of PA, such as rime awareness (e.g., Melby-Lervåg, Lyster, & Hulme, 2012).

Confusion often surrounds the terms PA and phonics. Whereas PA is an auditory skill that is assessed without reference to letters of the alphabet, phonics is an approach to reading instruction that incorporates the teaching of letter-sound associations as well as PA. A phonics approach is designed to teach children to read by converting written letter strings into spoken form (decoding), and to spell by converting spoken forms into letter strings (encoding). In the early stages of learning to read, phonics approaches are generally considered superior to non-phonics approaches for children with typical development (e.g., National Institute of Child Health & Human Development, 2000). An advantage of phonics is that it provides children with the skills needed to accurately identify unfamiliar written forms. Furthermore, according to the self-teaching hypothesis (e.g., Share, 1995, 1999), children's successful decoding attempts form the basis for establishing a set of word-specific orthographic representations. These orthographic representations are essential for fast and accurate sight word reading; that is, recognition of written words without the need for decoding.

Initial controversy surrounded the importance of PA in learning to read for children with DS. Cossu, Rossini, and Marshall (1993) reported that children with DS could read real words and nonwords at a level commensurate with a group of typically developing peers, despite having significantly poorer PA skills. They concluded that PA was not essential for reading in children with DS. Over time however, the role of PA has been reconsidered and is now believed by many to be a critical skill in learning to read for children with DS, especially in decoding unfamiliar words (e.g., Cupples & Iacono, 2002; Goetz et al., 2008). In accordance with this view, and on the basis that children with DS can benefit from PA instruction, in some cases to a greater extent than children with typical development, Naess (2016) recommended that children with DS should receive explicit and systematic PA training early in reading development.

Although fewer studies have investigated spelling than reading in children with DS, some evidence also suggests that children's spelling attempts reflect use of a phonic strategy. For example, Cardoso-Martins, Peterson, Olson, and Pennington (2009) reported that children with DS in their study made spelling attempts such as 'MK' for 'make' and 'LT' for 'light.' Although they constitute errors, these spelling attempts suggest the ability to identify some component sounds in the target words and map them onto corresponding letters. In addition, Lim, Arciuli, Rickard Liow, and Munro (2014) found that some children with DS were able to spell the first letters of target words; some wrote similar sounding words in their attempts (e.g., 'made' for 'maid'), and others provided phonologically plausible, but incorrect, responses (e.g., 'beaff' for 'beef'). These results are consistent with the proposal that phonics skills (i.e., PA and letter-sound correspondence) are a worthy potential intervention target for spelling in children with DS.

In sum, PA and/or phonics based instructional approaches, may contribute to successful reading and spelling development in children with DS. However, such instruction might also present some challenges for these children, who are known to have impairments in phonological short-term memory; that is, the ability to process and store spoken information for short periods of time (e.g., Cupples & Iacono, 2000; Laws, 1998; Lanfranchi, Baddeley, Gathercole, & Vianello, 2012). Consequently, they could encounter difficulties in remembering strings of ordered phonemes, as required for successful completion of PA and phonological decoding tasks. It is therefore important to consider additional, complementary approaches to reading instruction, such as sight word (also known as whole word) reading. In particular, it may be advantageous to adopt literacy instruction programs that foster both strategies for these children (Lemons & Fuchs, 2010a; Roch & Jarrold, 2012).

Reading and spelling instruction for individuals with DS

Some previous studies have documented the success of instructional programs whose primary focus was PA and/or phonics, for reading and/or spelling in children with DS. For example, Cologon, Cupples, and Wyver (2011) conducted a 10-week treatment study for 7 children with DS, ages 2; 11 to 10; 8 (years; months). Children were taught to read 50 target words (5 per week) using activities that included oral reading, word-picture matching, oral blending, and sentence completion. Children showed significant improvements in PA (phoneme blending and segmentation), alphabet knowledge (letter sounds), and the reading of both real words and nonwords from pre- to post-instruction but effect sizes were not reported. These improvements were maintained for 6 months. In a study that focused on teaching the PA skill of blending, Burgoyne, Duff, Snowling, Buckley, and Hulme (2013) described treatment outcomes for 10 children with DS, ages 6; 11 to 10; 6. Children received daily instruction for 6-weeks. They showed significant improvements in sound blending and word reading, but no significant improvements on sound isolation, nonword reading, or word spelling. Effect sizes were not reported. These studies are consistent with the view that instruction in PA and/or phonics may serve to boost word reading in children with DS. However, with respect to generalisability of intervention effects, Burgoyne et al. (2013) reported no significant change in children's spelling.

Other treatment studies have focussed more broadly on PA, phonics, sight-word teaching, and language skills (including vocabulary) in their literacy instruction for children with DS. Goetz et al. (2008) reported on a study of 15 school-aged children with DS (ages 8; 3 to 14; 6) who were taught targeted letter sounds, PA, sight word reading, book reading, and oral motor exercises to improve articulation. Intervention outcomes were evaluated by analysing results obtained over the initial 8-weeks of the study. During this period a group of children who received immediate instruction made significantly greater gains in early word recognition and knowledge of letter sounds than a delayed-intervention (control) group who had not yet received instruction. Effect sizes were large.

Baylis and Snowling (2012) evaluated a literacy program that covered a range of diverse skills, including: Letter sound knowledge, onset-rime awareness, sight vocabulary, word decoding, text reading accuracy, comprehension, word spelling, and writing. Instruction was delivered twice weekly over a 10-week period to a group of 10 children with DS (ages 9; 10 to 14; 10). Large and statistically significant improvements in word reading were evident after instruction for the group as a whole and maintained for 3 months. There was also evidence of enhanced phonic decoding in nonword reading for 4 of the 10 children.

A study by Lemons and Fuchs (2010b) reported on 24 children with DS (6–16 years) who received individual instruction 5 days per week for 6 weeks in PA (sound segmentation, sound blending), letter sounds, sight-word reading, and reading of decodable words, nonwords, and connected text. After intervention, 23 of the 24 participants demonstrated significant improvements in sight word reading and letter sounds. A subset of 16 children improved significantly in their reading of decodable words, with 9 of those children also showing significantly improved reading of nonwords.

The only larger scale reading and language intervention study conducted in this area, one which incorporated phonics and sight word approaches, was reported by Burgoyne and colleagues (Burgoyne et al., 2012). Fifty-seven children with DS (ages 5; 2 to 10; 0) were allocated to an intervention group ($n = 29$) or a wait control group ($n = 28$). For the intervention, children received individual instruction in letter sounds, phoneme blending, sight-word reading, text reading, and vocabulary knowledge for 40 min daily. Results showed that, compared to the wait control group, the intervention group made small-to-medium but statistically significant gains in single word reading, letter-sound knowledge, phoneme blending, and taught expressive vocabulary after 20 weeks of intervention. The children appeared unable to generalise their newly acquired skills to areas not directly targeted during intervention, such as word spelling.

This review of previous intervention studies is encouraging in showing that children with DS responded positively to instruction encompassing a diverse range of PA, phonics, and whole word components. Nevertheless, two shortcomings can be noted. First, there is limited evidence for generalisability of taught skills to related tasks (e.g., from reading aloud to spelling). Second, most studies utilised researcher-developed literacy programs, which are not easily accessed by a wider range of educators and clinicians. The question arises as to whether positive outcomes would be obtained in this population using a readily available program with some evidence of benefit for low progress readers.

In this respect, our approach in the current study is similar to that of Lemons, Mrachko, Kostewicz, and Payterra (2012), who described an evaluation for 15 children with DS (ages 5–13 years) using two commercially available programs, one targeting PA and simple decoding, and the other targeting accurate and fluent word identification. Unlike Lemons et al., however, who assigned participants to different treatment programs, our participants all received the same intervention program, but were assigned to different groups in order to manipulate the length of the no instruction (control) period. Another point of contrast is that we assessed a different program to those examined by Lemons, Mrachko, Kostewicz, and Paterra (2012)—the MULTILIT literacy program—which combines PA, phonics, and whole word approaches, in materials and training.

The MULTILIT program

Research has shown that MULTILIT can improve the reading and/or spelling skills of low progress readers (e.g., Wheldall, Beaman, & Langstaff, 2010; Buckingham, Beaman, & Wheldall, 2012), including some children with learning disabilities or mild intellectual disability (Wheldall & Beaman, 2000). In a study reported by Wheldall et al. (2010), participants received instruction 5 days a week over a 20-week period, with each session lasting approximately 3 h. Results showed statistically significant and large improvements in the reading of single words and nonwords, word spelling, and reading comprehension for a group of 34 low progress readers (mean age 11; 4) who scored at or below the 25th percentile on a standardised passage reading test prior to intervention.

Further evidence comes from a study by Buckingham et al. (2012) who reported on 44 elementary school children, from Year 3 to 6, with low average reading scores on the Wheldall Assessment of Reading Passages (Wheldall & Madelaine, 2006). Matched pairs of participants were randomly allocated to a control group or an experimental group. Children in the experimental group received the MULTILIT program 4 days per week, for a period of 18 weeks, with each session lasting approximately 1 h. Following instruction, a large and statistically significant difference in nonword reading favoured the experimental group, but no statistically significant group differences emerged on other measures, including word reading and spelling. The researchers suggested that factors such as fidelity of treatment across different MULTILIT instructors, and the relatively large teaching group size (of six per group) may have influenced the results.

As far as we are aware there has been no previously published study in which MULTILIT has been trialled with children who have DS. In addition, since much of the available evidence regarding MULTILIT has come from the team that created the program, the current study is one of the few independent assessments (in that none of the researchers associated with the current study has any commercial or other involvement with the program).

The current study

Much of the previous research aimed at improving the ability of children with DS to read aloud and/or spell single words has suffered from a lack of demonstrated generalisability, especially between related tasks (e.g., from blending to segmenting phonemes, or from reading to spelling). In the current study we addressed the issue of generalisability by assessing a variety of outcomes and using tests that were entirely independent of learning materials. In addition, to assist the literacy instruction practices of educators and clinicians as well as to inform researchers working in this field, we avoided researcher-developed teaching materials, assessing instead the MULTILIT program, an evidence-based literacy program developed at Macquarie University, where materials and training are commercially available.

The MULTILIT program is an ideal candidate in the context of this research for a number of reasons. It specifically targets teaching of PA skills at the phoneme level (i.e., phoneme blending and segmentation) and letter-sound correspondences to assist with spelling and word decoding. In addition, phonics skills are taught alongside recognition of sight words, a combination of techniques which may be especially beneficial for children with DS.

In the current study, we examined children's responses to MULTILIT training delivered on a 1:1 basis. Three specific research questions were addressed as follows:

1. Do children with DS show greater improvements in PA following MULTILIT intervention than a no-intervention control condition implemented within participants?
2. Do children with DS show greater improvements in word reading accuracy following MULTILIT intervention than a no-intervention control condition implemented within participants?
3. Do children with DS show greater improvements in word spelling accuracy following MULTILIT intervention than a no-intervention control condition implemented within participants?

We hypothesised that children with DS would show larger improvements in PA, word reading, and word spelling skills following MULTILIT than a no-intervention control condition.

Method

Participants

Ethics approval was granted by the University of Sydney Human Research Ethics Committee and informed written parental consent was obtained for all participants. All participants in the study were residents of Singapore. A total of 18 school-aged children with DS (ages 9; 2 to 17; 8) were recruited through the Down syndrome Association, Singapore, and a school for special needs in Singapore. One participant dropped out of the study after a family member passed away; two others had difficulty committing to weekly home visits for the duration of the program. These omissions resulted in a final participant sample of 15.

The inclusion criteria were: A diagnosis of DS, an absence of serious visual impairment and/or severe sensorineural hearing loss as reported by parents (no audiological testing was conducted), and that the child's dominant language spoken at home and at school was English. Fourteen participants were diagnosed with Trisomy 21 and one participant had a mosaic form of DS. According to parental report, three participants had a mild conductive hearing loss. The remaining 12 participants had hearing within normal limits. Parents reported that English was spoken at home at least 80% of the time for all participants. Three participants were also exposed to a language other than English (one to Mandarin and two to Malay), but parents reported that these participants did not use their second language to the extent of being considered bilingual or fluent in both. For all children, school lessons were delivered in English. Fourteen of the children attended a special school, while one child was in a mainstream school.

After baseline measures were collected, participants were allocated to one of two groups that varied only in the length of the 'no-intervention control period' between baseline and pre-instruction assessments: 12 weeks for Group 1 versus 24 weeks for Group 2. A nonrandomised allocation ensured comparable abilities prior to intervention across the two groups. Participants were matched group-wise on the three outcome measures used in the study, namely: PA (three measures—sound deletion, sound isolation, and sound blending), word reading accuracy, and word spelling accuracy. Of these variables, the weakest match was on spelling accuracy. In this case, the group difference and associated effect size were inflated by a single high-scoring participant in Group 1 (P5, see "Appendix"). Regardless, the group difference was not statistically significant. The two groups were also matched on a set of control variables that might reasonably be expected to associate with concurrent reading ability and/or changes in reading skill (e.g., see Hulme et al., 2012; Lemons & Fuchs, 2010b; Mengoni, Nash, & Hulme, 2014), namely: Chronological age, phonological short-term memory, nonverbal intelligence, receptive vocabulary, and passage reading comprehension. A more detailed account of the matching procedure is provided in Table 1. See the Measures section below for assessment details. Group 1 comprised eight participants (seven males) and Group 2 comprised seven participants (two males). As can be seen from Table 1, there were no significant differences between the groups on

Table 1 Mean raw scores, standard deviations and independent *t* test results for baseline measures as a function of participant group

Variable (maximum raw score)	Group 1 ($n=8$) (12-week control) M (SD)	Group 2 ($n=7$) (24-week control) M (SD)	t ($df=13$)[a]	*Sig.* level (p)	Cohen's d[b]
Age in months	168.88 (16.20)	164.71 (44.37)	.24	.82	.12
CELF–4 digit span[c] (16)	3.75 (2.66)	4.00 (1.63)	.22	.83	.11
Raven's CPM[d] (36)	12.38 (4.47)	14.00 (7.66)	.51	.62	.26
SBVT vocabulary[e] (70)	41.88 (11.90)	38.86 (5.76)	.64	.54	.32
Passage reading comprehension[f] (16)	2.88 (3.00)	2.00 (2.31)	.63	.54	.33
PA total[g] (36)	11.25 (5.47)	11.43 (7.93)	.00	.96	.03
WRAT word reading[h] (55)	12.88 (8.82)	12.43 (7.72)	.10	.92	.05
WRAT spelling[i] (20)	8.00 (4.81)	5.29 (3.04)	1.28	.22	.67

The procedure for matching groups 1 and 2 was as follows. First, all 15 participants were ordered from high to low according to their WRAT word reading scores. Then, beginning with the highest scoring participant, individuals were allocated to alternate groups. Finally, 3 pairs of participants with very similar scores were swapped between groups in order to optimise matching across *all* relevant variables

[a]Levene's test for equality of variances revealed a significant difference between groups for Age in months and SBVT vocabulary—degrees of freedom were reduced to 7.4 and 10.4 accordingly

[b]Cohen's *d* measures effect size for comparisons involving two independent groups—small, medium, and large effects are reflected in d-values of .20, .50, and .80 respectively

[c]Clinical Evaluation of Language Fundamentals-Fourth Edition, Digit Span

[d]Raven's Coloured Progressive Matrices

[e]Singapore Bilingual Vocabulary Test

[f]York Assessment of Reading for Comprehension (YARC), Passage Reading Test

[g]PA Total (summed raw scores for sound isolation and sound deletion from YARC Early Reading and sound blending from Burgoyne et al., 2012)

[h]Wide Range Achievement Test-Fourth Edition, Word Reading

[i]Wide Range Achievement Test-Fourth Edition, Spelling

the matching variables, with all p values clearly non-significant (i.e., all significance levels > .20).

Table 1 reveals that, as expected, the children in this study performed well below the levels expected of typically developing children at a similar age on the primary dependent variables of reading and spelling. Thirteen of the 15 participants achieved word reading scores that fell within the bottom 1% of the population on the Wide Range Achievement Test-Fourth edition (WRAT-4; Wilkinson & Robertson, 2006), while the remaining 2 participants scored within the bottom 4%. Spelling was equally or more delayed, with all participants scoring in the bottom .5% of the normative distribution. Non-verbal intelligence was assessed using the Raven's Coloured Progressive Matrices (Raven's CPM; Raven, Court, & Raven, 1995). Children achieved a mean raw score of 13.1 correct (out of a possible 36), with an interquartile range (IQR) from 10 to 14. This level of performance placed them at or below the normative level expected at

6 years of age. These consistently low achievement levels were evident for even the oldest participants in the sample, ages 17; 6 and 17; 8.

Procedure and assessments

All assessment sessions and all MULTILIT instructional sessions were delivered by the same clinician who, prior to the start of the study, attended the recommended MULTILIT training workshop in Sydney, Australia. The program was delivered sequentially to two groups of participants. This allowed us to assess any improvement in literacy skills that might have occurred without the MULTILIT instructional program, over a longer time period for the second intervention group. Children in both groups were assessed using the same test battery on three occasions: Baseline, pre-instruction and post-instruction. For Group 1, pre-instruction testing took place 12 weeks after baseline, followed by 12 weeks of instruction and an immediate post-instruction assessment (at 24 weeks after baseline). The only difference for Group 2 was that pre-instruction testing took place 24 weeks after baseline to provide a longer control period. Both groups eventually received the instruction, which consisted of the MULTILIT Sight Words program (MULTILIT, 2007b) and MULTILIT Word Attack Skills program (MULTILIT, 2007c).

Assessments and subsequent teaching sessions were conducted at a convenient location for each child (such as at the premises of the Down syndrome Association, a quiet room at the child's home, or in a room at a speech pathology clinic). These sessions were supplementary to ordinary schooling activities. During both treatment and control periods it was business as usual for all participants.

Treatment fidelity

Fidelity can be assessed using compliance, context or competence measures (O' Hare & Doell, 2015). For the current study, compliance fidelity measures included use of the MULTILIT manual, with adherence to all instructions contained therein. The MULTILIT program also provides online help, workbooks, and teaching videos to support clinicians and educators. Finally, all children attended from one to three individual treatment sessions per week. To ensure context fidelity, the MULTILIT program was delivered by a clinician who underwent the recommended 1-day training workshop in the MULTILIT Reading Tutor Program. This training included detailed explanations and practical exercises relating to all aspects of the program. Although no competency measures were obtained, the clinician who administered the MULITILIT program is a qualified speech pathologist who had worked with children with DS for several years in the area of literacy skills.

Measures

Baseline measures

Participants' abilities were assessed at the outset of the study. The following baseline measures were used.

Phonological short-term memory

The digit span subtest of the Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003) was administered. The examiner read out a series of digits arranged in increasing length (from a minimum of two to a maximum of nine) with two trials presented at each span length. Participants were asked to repeat the digits in the same order as presented (i.e., forward digit span). A time limit of 10 s was applied for each item. A score of 1 was given for a correct response. If a participant did not respond or gave the wrong answer, a score of 0 was given. According to Semel et al., test–retest reliability is .80 for typically developing individuals and individuals with educational disabilities, ages 5–21 years.

Nonverbal intelligence

The Raven's CPM (Raven et al., 1995) comprising 36 items was administered to assess nonverbal intelligence. For each stimulus item, participants were presented with an incomplete matrix, and asked to choose from six alternatives the correct matching piece to complete the pattern. A time limit of 10 s was applied for each stimulus item. Participants achieved a score of 1 for a correct response while an incorrect response or no response yielded a score of 0. Cronbach's alpha = .94 for a normative sample, ages 4 to 11 years (Raven et al., 1995).

Receptive vocabulary

A total of 70 items from the Singapore Bilingual Vocabulary Test (SBVT; Rickard Liow & Tng, 2003) was administered in English. The SBVT is a culturally appropriate test, which has been used with the local Singaporean population (see Lim et al., 2014) based on the British Picture Vocabulary Scale. Participants listened to a target word spoken aloud by the examiner and pointed to the correct picture from a choice of four. A time limit of 10 s was applied for each item. A score of 1 was awarded for a correct response. If a participant did not respond or gave the wrong answer, a score of 0 was given. Cronbach's alpha = .81 for a sample of children with DS, ages 7–13 years (Lim et al., 2014).

Reading comprehension

The passage reading test from the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2009) was administered to measure reading comprehension abilities. Minor substitutions were made to enhance cultural appropriateness;

for example, “Today is a very special day for Mrs Johnson’s class” was changed to “Today is a very special day for Mrs Lee’s class.” The test consists of a Beginner Level passage where the examiner and the child take turns reading a short paragraph aloud. The remaining passages from Level 1 to Level 6 are read aloud solely by the participant. A total of sixteen comprehension questions was administered by the examiner. Verbal clarifications such as “Tell me more?” were used to facilitate responses from the participants. A correct response yielded a score of 1 while an incorrect or no response yielded a score of 0. Cronbach’s alpha = .55 to .77 for a normative sample ages 5–11 years (Snowling et al., 2009).

Outcome measures

We included three outcome measures. They were PA (sound isolation, sound deletion, and sound blending), single word reading, and spelling.

Phonological awareness

PA was assessed using two subtests from the YARC—sound isolation and sound deletion—and a sound blending task adapted from Burgoyne et al. (2012). Each task contained 12 test items, with a time limit of 10 s applied for each item. In the current research, a composite PA score was computed for each participant by tallying the number of correct responses made in the three tasks.

Sound isolation. To explain phoneme isolation, three felt squares were displayed to symbolise the phonemes /s/, /a/, /n/. In response to the question, “What is the first sound in san?” the examiner pointed to the first felt square and said, “/s/”. Six test items targeted identification of initial phonemes, and six items targeted identification of final phonemes. All test items were nonwords (e.g., “mig”, “swib”). Cronbach’s alpha = .93 for a normative sample, ages 4–7 years (Snowling et al., 2009).

Sound deletion. To explain sound deletion, three felt squares were displayed to symbolise the phonemes /m/, /i/, /t/. In response to the question, “What do you get when you take /m/ away from /mit/?”, the examiner removed the felt square corresponding to the deleted initial phoneme and said /it/. All test items were real words (e.g., “boat, “house”). Cronbach’s alpha = .93 for a normative sample, ages 4–7 years (Snowling et al., 2009).

Sound blending. To explain sound blending, two felt squares were displayed to symbolise the phonemes /b/, /i/. In response to the question, “What do you get when you join /b/ and /i/ together?” the examiner pointed to the two squares and said “bee.” All test items were real words (e.g., “bat,” “doll”). Cronbach’s alpha = .66 for children with DS, ages 6–10 years (Burgoyne et al., 2012).

Word reading

The word reading subtest from the WRAT-4 (Wilkinson & Robertson, 2006) was administered. This test comprises 55 words that increase in complexity from simple (e.g., ‘milk’) to more complex (e.g., ‘rudimentary’). A time limit of 10 s was applied for each item. If a participant made no response or read the word incorrectly, a score of 0 was given. Cronbach's alpha = .92 for a normative sample, ages 5–94 years (Wilkinson & Robertson, 2006).

Word spelling

The first 20 words from the spelling subtest of the WRAT-4 were administered individually. Each word was read aloud by the clinician, and then read aloud in the context of a sentence before being read in isolation once more. Participants wrote their answers down and were instructed to try their best to spell all the target words. Spelling accuracy was defined as the number of whole words spelled correctly, each one achieving a score of 1. A time limit of 15 s to commence the spelling was applied for each item. If a child did not write anything or spelled the word incorrectly, a score of 0 was given. For the complete (42-word) subtest, Cronbach's alpha = .90 for a normative sample, ages 5–94 years (Wilkinson & Robertson, 2006).

MULTILIT instructional program

Two components of the MULTILIT Reading Tutor Program were used for instruction purposes: MULTILIT Sight Words (MULTILIT, 2007b) and MULTILIT Word Attack Skills (MULTILIT, 2007c). Individual sessions lasting 45–60 min in duration were conducted outside school hours up to three times per week for 12 consecutive weeks. Frequency of attendance varied across children as a consequence of their individual circumstances and preferences. For Group 1 (12-week control period), four children attended sessions three times per week, three children attended sessions twice per week and one child attended sessions once per week. For Group 2 (24-week control period), one child attended sessions three times per week, three children attended sessions twice per week and three children attended sessions once per week.

Word attack

The MULTILIT Word Attack Skills program comprises three parts: Word attack accuracy, word attack fluency, and word spelling. All parts consist of 13 levels of increasing difficulty. Level 1 requires children to name consonants and single vowels, and their corresponding sounds (e.g., /b/, /i/), whereas Levels 11–13 require children to read progressively longer words (e.g., "fright", "storm") with various letter combinations (e.g., "igh", "or"). As per MULTILIT guidelines, before commencing instruction, the MULTILIT Word Attack Skills placement test was administered to determine each child's entry level for instruction. A failed level occurred where the child made two or more errors. Testing stopped when the child failed three MULTILIT Word Attack Skills levels.

Word attack accuracy

Children received instruction in reading accuracy using letter sound correspondences and decoding strategies. A list of words to be taught was presented for oral reading. For each misread word, the clinician assisted by pointing to each letter with her finger, and sounding out the word's individual letters (e.g.,/s//a//d/). The child was asked to perform the step with the clinician, then independently. Following this procedure, and in line with instructions in the MULTILIT manual, 10 or more words were then chosen from the teaching page in random order. Cues were faded to allow the child to perform the strategy on his or her own; for example, "Now I want you to read the whole word. When I point to the word, I want you to sound out the word in your head like we've just done until I say, 'What word?'." If the child made one or more errors, teaching took place again. The criterion for progressing to the next MULTILIT Word Attack Skills level was 100% accuracy on the test page consisting of 15 words. To aid with generalization of the taught words to a meaningful context, the child read a short passage consisting of previously learnt sounds or words at the end of each skill level. There was no criterion for the short passage reading.

Word attack fluency

Children read the same set of words taught in word attack accuracy within a time limit. The imposition of a time limit helped to train fluency and automaticity in reading (MULTILIT, 2007c). The recommended time limit of 15–20 s was modified to a time limit of 60–90 s for the children in this study.

Spelling

A spelling component for the decoded words was also included in treatment but no criterion was used for progression; that is, children were not required to spell a specified number of words correctly before moving to the next level. During each teaching session, children were asked to write five words, saying the sound of each letter aloud as they wrote in order to reinforce letter sound correspondences.

Sight words

The MULTILIT Sight Words program involves teaching of 200 high frequency sight words encountered in children's books such as "school", "them", "fast" (MULTILIT, 2007a). As per MULTILIT guidelines, a placement test was administered first to determine the child's appropriate level for instruction. Testing stopped when the child made 10 reading errors. Teaching steps for each misread word were (a) the clinician wrote the word in an exercise book and spelled it out loud, (b) the child was asked to read the word, (c) the child copied the word and spelled it out loud while writing, (d) the child was asked to read the word again, (e) the word was covered and the child was asked to write it out from memory, and then say it, and (f) the child was asked to construct a spoken sentence using the word. If the child did not know how to form a sentence with the misread word, a model sentence was provided by the clinician. Children moved on to the next level when 100% accuracy was achieved for each list consisting of 10 words. Children also revised lists from the previous week at each session.

Data analysis

All participants in Groups 1 and 2 were assessed at three time points. Hence, the design was a 2 (Group) × 3 (Time) fully-crossed factorial, with non-repeated measures for Group (12-week control period vs. 24-week control period) and repeated measures for Time (baseline vs. pre-instruction vs. post-instruction). Analyses of variance (ANOVAs) were used to address each of the three research questions outlined above. Planned pairwise comparisons were conducted on the main effect of Time and the Time x Group interaction, because predictions for change differed across the various assessment sessions. Scores were expected to remain stable for both groups from baseline to pre-instruction, with significant change predicted to occur from pre-instruction to post-instruction. ANOVAs and pairwise comparisons were conducted using SPSS version 22.

Results

The sphericity assumption, which is associated with the use of repeated measures designs, was not met for analyses of PA or word reading, thus increasing the likelihood of a Type 1 error (incorrect rejection of the null hypothesis). To counteract this problem, the Huynh-Felt correction was applied to the degrees of freedom for the omnibus F-tests in these cases. All other assumptions were met.

Did MULTILIT instruction improve PA, word reading and word spelling?

Table 2 shows the mean PA scores achieved by participants in both groups at baseline, pre-instruction and post-instruction. Individual participants' raw scores are shown in "Appendix". Raw PA scores showed evidence of excellent test–retest reliability across the full participant sample prior to instruction (Pearson's $r = .953$ between scores at baseline and pre-instruction). An ANOVA revealed that the main effect of time was significant [F (1.5, 26) = 21.83, $p < .001$, $\eta_p^2 = .63$]; and planned comparisons confirmed that, as expected, significant improvements in PA occurred from pre-instruction to post-instruction [F (1, 13) = 58.04, $p < .001$, $\eta_p^2 = .82$], but not from baseline to pre-instruction [F (1, 13) = 3.03, $p = .106$, $\eta_p^2 = .19$]. Although the number of weeks that intervened between baseline and pre-instruction assessments was greater for Group 2 (24 weeks) than for Group 1 (12 weeks), there was no indication that either group of participants improved in their PA prior to receiving MULTILIT instruction. In fact, both groups showed a slight reduction in their

Table 2 Mean raw scores and standard deviations for the dependent variables of PA Total, WRAT word reading and WRAT spelling as a function of participant group and time period

Dependent variable (maximum possible score)	Group 1 (12-week control; $n=8$)			Group 2 (24-week control; $n=7$)		
	Baseline	Pre-instruction (12 weeks after baseline)	Post-instruction (24 weeks after baseline)	Baseline	Pre-instruction (24 weeks after baseline)	Post-instruction (36 weeks after baseline)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
PA total[a] (36)	11.25 (5.47)	10.00 (4.57)	15.25 (5.34)	11.43 (7.93)	10.86 (7.52)	14.14 (8.59)
WRAT Word reading[b,c] (55)	12.88 (8.82)	12.88 (9.82)	15.13 (10.95)	12.33 (8.45)	12.17 (8.70)	15.50 (12.37)
WRAT spelling[d] (20)	8.00 (4.81)	7.13 (5.00)	8.38 (4.96)	5.29 (3.04)	5.00 (3.42)	6.43 (3.41)

[a]PA Total (summed raw scores for sound isolation and sound deletion from the York Assessment for Early Reading and sound blending from Burgoyne et al., 2012; max. score = 36)

[b]Wide Range Achievement Test-Fourth Edition, Word Reading

[c]Group 2 means and SDs computed across 6 participants-outlier omitted

[d]Wide Range Achievement Test-Fourth Edition, Spelling

mean PA Total scores from baseline to pre-instruction. Group 1 showed an average reduction of approximately 3.5% (1.25 out of 36) over 12 weeks, whereas Group 2 showed an average reduction of 1.6% (.57 out of 36) over 24 weeks (see Table 2). Confirming the similarity between groups, neither the main effect of group nor the interaction between Time and Group approached significance (both $F\text{s} < 1$).

The number of single words read correctly by individual participants at each assessment point is shown in “Appendix”. Raw scores showed evidence of excellent test–retest reliability across the full participant sample prior to instruction (Pearson’s r = .946 between reading scores at baseline and pre-instruction). Prior to analysis, data were omitted from a single outlying participant in Group 2 (see “Appendix”, Participant 15) who read approximately three times as many words correctly at baseline (13 words correct) as at pre- or post-instruction (4 and 5 words correct respectively). Table 2 shows the mean number of single words read correctly by the remaining participants in both groups at baseline, pre-instruction, and post-instruction.

As per the PA data, the largest amount of improvement in word reading occurred from pre-instruction to post-instruction. Thus, although the main effect of time was significant [F (1.4, 24) = 4.75, $p = .033$, $\eta_p^2 = .28$], planned comparisons confirmed a significant improvement in word reading from pre- to post-instruction [F (1, 12) = 6.72, $p = .024$, $\eta_p^2 = .36$], but not from baseline to pre-instruction ($F < 1$). Once again, the two participant groups showed a similar pattern of results despite the marked difference in the length of time that intervened between baseline and pre-instruction assessments. In particular, there was no suggestion in the data that either group’s reading improved during the no-intervention control period: Group 1 participants read the same number of words correctly, on average, at baseline and pre-instruction assessments (12 weeks intervening); whereas Group 2 participants read .16 words fewer, on average, at pre-instruction than baseline (24 weeks intervening). Neither the main effect of group nor the interaction between time and group approached significance (both $F\text{s} < 1$). To ensure that the statistical results were not unduly biased by our omission of the outlying participant mentioned above, the analysis was repeated using the full sample of 15 participants. The essential pattern of results did not change. Although the overall effect of time was not quite significant [F (1.54, 26) = 3.05, $p = .081$, $\eta_p^2 = .19$], planned comparisons confirmed a significant improvement in word reading from pre- to post-instruction [F (1, 13) = 6.84, $p = .021$, $\eta_p^2 = .35$], but not from baseline to pre-instruction ($F < 1$).

The mean number of words spelled correctly by participants in both groups at baseline, pre-instruction, and post-instruction are shown in Table 2, with individual participants’ raw scores shown in “Appendix”. Test–retest reliability was once again excellent across the full participant sample prior to instruction (Pearson’s r = .946 between spelling scores at baseline and pre-instruction). Consistent with the other outcome measures, ANOVA revealed a significant overall effect of time [F (2, 26) = 6.24, $p = .006$, $\eta_p^2 = .32$], reflecting once again an improvement in children’s performance from pre- to post-instruction [F (1, 13) = 15.00, $p = .002$, $\eta_p^2 = .54$] but not from baseline to pre-instruction [F (1, 13) = 2.49, $p = .139$, $\eta_p^2 = .16$]. Neither the main effect of group nor the interaction between group and time approached significance, with F (1, 13) = 1.10, $p = .313$, $\eta_p^2 = .08$ and $F < 1$ respectively.

Finally, correlational analyses using *Pearson's* coefficient revealed no statistically significant association between attendance (total number of sessions over the 12 weeks) and any of the outcome measures (i.e., PA, word reading, and word spelling), with all p values > .05. However, this result needs to be interpreted within the context of minimal variability in attendance in our study. A larger study with greater variability in attendance may reveal a relationship between attendance and literacy outcomes.

Discussion

The present study evaluated outcomes after a 12-week MULTILIT program for developing PA, word reading accuracy, and word spelling accuracy skills in children with DS. Two groups of children were matched for average performance on a range of relevant variables at the outset of the study and were assessed at three time points. The key difference between the groups was the duration of the no-intervention control period. Group 1 received the MULTILIT program following a no-intervention control period of 12 weeks, whereas Group 2 received the intervention after a corresponding period of 24 weeks.

A similar pattern of results was obtained for the three dependent variables of PA, single word reading, and single word spelling. Whereas neither group of children improved significantly during the control period (from baseline to pre-instruction testing), both groups made significant improvements on all outcome measures after receiving the MULTILIT program. All statistically significant effects were large, with the largest value of η_p^2 being .82 for PA. Moreover, these effects were obtained using tests that were entirely independent of the materials taught during the literacy instruction program.

Our findings are in line with previous research showing that children with DS can benefit from instruction that incorporates PA and/or phonics (e.g., Goetz et al., 2008; Burgoyne et al., 2013), and single word reading (e.g., Cologon et al., 2011; Burgoyne et al., 2012; Burgoyne et al., 2013). Nevertheless, some differences are evident between the current study and previous published research. The results of Burgoyne et al.'s (2013) intervention study showed gains in PA (sound blending) and word reading measures but not for word spelling skills. By contrast, findings of the current study show significant improvements in word spelling skills as well as PA and word reading. This difference in generalisation might be due, in part, to the longer intervention period in the current study (12 weeks compared to 6 weeks). The current results also differ from those reported by Baylis and Snowling (2012), whose post-instruction results revealed a small-to-medium and non-significant effect on spelling, compared to the current study in which children's spelling showed a large and statistically significant improvement over the instruction period. It is possible that the current study's emphasis on phoneme-level PA was important in this regard. By contrast, Baylis and Snowling explicitly targeted onset-rime awareness.

We found that there was no significant association between children's intervention outcomes and 'dose' (i.e., the number of sessions they attended per week). Previous research suggests, however, that a longer study with greater variability in attendance may reveal a relationship with literacy outcomes for children with DS. Burgoyne et al. (2012) reported a positive association between growth in reading and number of treatment sessions attended over a 40-week period. In a related vein, Yoder, Woynaroski, Fey, and Warren (2014) reported that communication intervention produced significantly greater growth in spoken vocabulary over a 9-month period in children with DS who received five 1-h treatment sessions per week compared to a single 1-h treatment session per week, after controlling for children's general intellectual ability. Importantly, the current study was not designed to examine the question of the relationship between dose and outcomes. Before drawing any strong conclusions in this regard, a targeted study is necessary comparing closely-matched groups of children who differ systematically in the intensity or dose of their intervention while all other aspects are held constant.

We report statistically significant and large effect sizes after MULTILIT instruction for children with DS. The large effect sizes are especially noteworthy in being observed relative to a lengthy no intervention control period of almost 6 months (i.e., 24 weeks) for half of the children; and the statistical significance of our findings suggests that a relatively small sample size of 15 is sufficient to detect real intervention effects when assessed within participants. We acknowledge, however, that while the sample size reported here compares favourably with previous research in this area (e.g., Goetz et al., 2008, and Lemons et al., 2012, included 15 participants in each of their studies while Cologon et al., 2011, and Burgoyne et al., 2013 included fewer participants) future studies attracting larger samples will improve research quality in the field.

Limitations and future directions

Some limitations of the current study include the fact that no measures of compliance fidelity were collected regarding delivery of the MULTILIT program. Also, children were assessed immediately after instruction but not over the longer term. Future research on the use of MULTILIT with children who have DS would be strengthened by tracking children's progress for several months or even longer after instruction to determine whether observed improvements are maintained.

Another avenue for future research is to explore the use of MULTILIT for children with DS in terms of reading comprehension. The primary reading outcome measure in this study was children's ability to read single words aloud accurately, rather than their comprehension. This approach falls squarely within the context of the simple view of reading, which characterises reading comprehension as the product of decoding and listening comprehension (Gough & Tunmer, 1986). An important aim of future research is to investigate the proposed link between decoding and reading comprehension more directly for these children. To this end it would be useful to include the MULTILIT reinforced reading component in future studies. This component targets daily book reading activities between parents and children using pausing, prompting and praising techniques.

Regardless of potential areas for improvement, the current study provides positive evidence that children with DS can benefit from targeted instruction using the MULTILIT program. On the basis of the current results we cannot, of course, generalise to the wider population of children with DS, because it is an extremely heterogeneous group in terms of nonverbal ability, speech, and language skills, to name a few. Nevertheless, the results obtained here indicate that MULTILIT could be implemented for children with DS by trained professionals, including teachers who work in special education, therapy assistants, parents, or tutors who undergo relevant MULTLIT training. In this way, there is a possibility of helping to develop and improve their PA, reading, and spelling skills for more individuals with DS.

Conclusion

Our findings indicate that the MULTLIT literacy program can assist children with DS to achieve substantial improvements in PA, word reading accuracy, and word spelling accuracy. As far as we are aware, this study is the first to use the MULTILIT program with children with DS. Importantly, this study is a trial of the MULTILIT program that is independent of the program's creators. It is hoped that research such as ours, which assesses literacy interventions that are accessible to clinicians, educators, and researchers alike, will improve literacy acquisition practices for children with DS.

Acknowledgements Lisa Lim was supported by an Australian Government APA Award. Joanne Arciuli was partly supported by a mid career research Future Fellowship granted by the Australian Research Council. Lisa Lim completed this research as a Ph.D. student under primary supervision of Joanne Arciuli. We would like to thank all the children and their families who participated in this research.

Appendix

See Table 3.

Table 3 Participants' ages and raw scores at each time period for PA, word reading, and spelling measures

Variable	Group 1 (12-week control; $n=8$)								Group 2 (24-week control; $n=7$)						
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
Age in months	159	186	138	186	160	173	179	170	125	212	189	186	121	110	210
PA total[a]: baseline	10	5	12	16	12	12	20	3	14	1	8	18	13	23	3
Pre-instruction	13	3	10	12	11	13	15	3	13	1	7	16	12	23	4
Post-instruction	21	9	14	19	17	20	16	6	14	2	10	21	18	27	7
Word reading[b]: baseline	15	14	2	21	26	2	17	6	10	0	17	6	19	22	13
Pre-instruction	16	13	3	19	30	1	17	4	7	0	15	8	20	23	4
Post-instruction	21	15	2	21	34	4	19	5	7	2	17	8	35	24	5
Spelling[c]: baseline	7	9	4	10	17	3	11	3	3	0	6	5	8	9	6
Pre-instruction	8	9	0	9	16	3	9	3	1	0	8	5	7	9	5
Post-instruction	10	9	3	13	16	3	10	3	3	1	10	6	7	10	8

[a]PA Total (summed raw scores for sound isolation and sound deletion from the York Assessment for Early Reading and sound blending from Burgoyne et al., 2012; max. score = 36)

[b]Wide Range Achievement Test-Fourth Edition, Word Reading (max. score = 55)

[c]Wide Range Achievement Test-Fourth Edition, Spelling (max. score = 20)

References

Baylis, P., & Snowling, M. J. (2012). Evaluation of a phonological reading programme for children with Down syndrome. *Child Language Teaching and Therapy, 28,* 39–56.

Buckingham, J., Beaman, R., & Wheldall, K. (2012). A randomised control trial of a MULTILIT small group intervention for older low progress readers. *Effective Education, 4,* 1–26.

Burgoyne, K., Duff, F. J., Clarke, P. J., Buckley, S., Snowling, M. J., & Hulme, C. (2012). Efficacy of a reading and language intervention for children with Down syndrome: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 54,* 1044–1053.

Burgoyne, K., Duff, F. J., Snowling, M. J., Buckley, S., & Hulme, C. (2013). Training phoneme blending skills in children with Down syndrome. *Child Language Teaching and Therapy, 29,* 273–290.

Cardoso-Martins, C., Peterson, R. L., Olson, R. K., & Pennington, B. F. (2009). Component reading skills in Down syndrome. *Reading and Writing: An Interdisciplinary Journal, 22,* 277–292.

Cologon, K., Cupples, L., & Wyver, S. (2011). Effects of targeted reading instruction on phonological awareness and phonic decoding in children with Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *116*, 111–129.

Cossu, G., Rossini, F., & Marshall, J. C. (1993). When reading is acquired but phonemic awareness is not: A study of literacy in Down syndrome. *Cognition, 46,* 129–138.

Cupples, L., & Iacono, T. (2000). Phonological awareness and oral reading skill in children with Down syndrome. *Journal of Speech, Language and Hearing Research*, *43*, 595–608.

Cupples, L., & Iacono, T. (2002). The efficacy of 'whole word' versus 'analytic' reading instruction for children with Down syndrome. *Reading and Writing: An Interdisciplinary Journal*, *15*, 549–574.

Goetz, K., Hulme, H., Brigstocke, S., Carroll, M. J., Nasir, L., & Snowling, M. (2008). Training reading and phoneme awareness skills in children with Down syndrome. *Reading and Writing: An Interdisciplinary Journal, 21,* 395–412.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Hulme, C., Goetz, K., Brigstocke, S., Nash, H. M., Lervåg, A., & Snowling, M. J. (2012). The growth of reading skills in children with Down syndrome. *Developmental Science, 15,* 320–329.

Lanfranchi, S., Baddeley, A., Gathercole, S., & Vianello, R. (2012). Working memory in Down syndrome: Is there a dual task deficit? *Journal of Intellectual Disability Research, 56,* 157–166.

Laws, G. (1998). The use of nonword repetition as a test of phonological memory in children with Down syndrome. *Journal of Child Psychology and Psychiatry, 39,* 1119–1130.

Lemons, C. J., & Fuchs, D. (2010a). Phonological awareness of children with Down syndrome: Its role in learning to read and the effectiveness of related interventions. *Research in Developmental Disabilities, 31,* 316–330.

Lemons, C. J., & Fuchs, D. (2010b). Modelling response to reading intervention in children with Down syndrome: An examination of predictors of differential growth. *Reading Research Quarterly, 45,* 134–168.

Lemons, C. J., Mrachko, A. A., Kostewicz, D. E., & Paterra, M. F. (2012). Effectiveness of decoding and phonological awareness interventions for children with Down syndrome. *Exceptional Children, 79,* 67–90.

Lim, L., Arciuli, J., Rickard Liow, S., & Munro, N. (2014). Predictors of spelling ability in children with Down syndrome. *Scientific Studies of Reading*, *18*, 173–191.

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138,* 322–352.

Mengoni, S. E., Nash, H. M., & Hulme, C. (2014). Learning to read new words in individuals with Down syndrome: Testing the role of phonological knowledge. *Research in Developmental Disabilities, 35,* 1098–1109.

MULTILIT. (2007a). *The MULTILIT Reading Tutor Program (revised).* Sydney, Australia: MULTILIT Pty Ltd.

MULTILIT. (2007b). *The MULTILIT Word Attack Skills (revised): Manual*. Sydney, Australia: MULTILIT Pty Ltd.

MULTILIT. (2007c). *The MULTILIT Sight Words (revised): Manual*. Sydney, Australia: MULTILIT Pty Ltd.

Naess, K. A. (2016). Development of phonological awareness in Down syndrome: A meta-analysis and empirical study. *Developmental Psychology, 52,* 177–190.

National Institute of Child Health & Human Development. (2000). *Report of the national reading panel: Teaching children to read*. Washington, D.C.: National Institute of Child Health and Human Development, National Institutes of Health.

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16,* 91–121.

O'Hare, M., & Doell, E. (2015). What supports speech-language pathologists to implement treatments with fidelity? *Journal of Clinical Practice in Speech-Language Pathology, 17,* 140–143.

Raven, J. C., Court, J. H., & Raven, J. (1995). *Coloured progressive matrices*. Oxford: Oxford Psychologists Press.

Rickard Liow, S. J., & Tng, S. K. (2003). Biscriptal literacy development of Chinese children in Singapore. In C. McBride-Chang & H. C. Chen (Eds.), *Reading development in chinese children* (pp. 215–228). Westport, CT: Praeger.

Roch, M., & Jarrold, C. (2012). A follow up study on word and non-word reading skills in Down syndrome. *Research in Developmental Disabilities, 45,* 121–128.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals (CELF-4)* (Fourth ed.). Toronto: The Psychological Corporation.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading. *Cognition*, *55*, 151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of he self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*, 95–129.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., et al. (2009). *York assessment of reading for comprehension*. Great Britain: GL Publishers.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Wheldall, K., & Beaman, R. (2000). *An evaluation of MULTILIT: Making Up Lost Time in Literacy*. Canberra: Department of Education, Training and Youth Affairs.

Wheldall, K., Beaman, R., & Langstaff, E. (2010). 'Mind the gap': Effective literacy instruction for indigenous low-progress readers. *Australasian Journal of Special Education, 34,* 1–16.

Wheldall, K., & Madelaine, A. (2006). The development of a passage reading test for the frequent monitoring of performance of low-progress readers. *Australasian Journal of Special Education, 30,* 72–85.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test* (4th ed.). Lutz, FL: Psychological Assessment Resources.

Yoder, P. J., Woynaroski, T., Fey, M., & Warren, S. (2014). Effects of dose frequency of early communication intervention in young children with and without Down syndrome. *American Journal of Intellectual and Developmental Disabilities, 119,* 17–32.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.