frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 21 May 2024
DOI 10.3389/feduc.2024.1354182

OPEN ACCESS

EDITED BY
Angela Jocelyn Fawcett,
Swansea University, United Kingdom

REVIEWED BY
Mindy Bridges,
University of Kansas Medical Center, United States
Tanya Serry,
La Trobe University, Australia

*CORRESPONDENCE
Gail Gillon
✉ gail.gillon@canterbury.ac.nz

RECEIVED 12 December 2023
ACCEPTED 15 April 2024
PUBLISHED 21 May 2024

CITATION
Gillon G, McNeill B, Scott A, Gath M, Macfarlane A and Taleni T (2024) Large scale implementation of effective early literacy instruction. *Front. Educ.* 9:1354182. doi: 10.3389/feduc.2024.1354182

Large scale implementation of effective early literacy instruction

Gail Gillon*, Brigid McNeill, Amy Scott, Megan Gath, Angus Macfarlane and Tufulasi Taleni

Child Well-Being Research Institute, Faculty of Education, University of Canterbury, Christchurch, New Zealand

Introduction: Large-scale implementation of Multi-Tiered Systems of Support (MTSS) to enhance children's early literacy success is critically important to address global literacy challenges. This paper describes one such initiative, the Better Start Literacy Approach (BSLA), which was specifically designed for large-scale implementation in New Zealand.

Methods: Between February 2020 and May 2023 over 3,000 teachers in 819 schools across New Zealand implemented BSLA, with baseline data available for 29,795 5-year-old children. Teachers implemented novel online assessments to monitor children's early literacy growth.

Results: In comparison to an internal control group, accelerated progress in children's phonic, and phoneme awareness skills was evident after 10 weeks of BSLA Tier 1 (universal) teaching. After 30 weeks, there were significant gains in word reading, spelling, listening comprehension and oral narrative abilities; growth did not differ based on gender, ethnicity, or socioeconomic status. Children with greater learning needs who received BSLA Tier 2 (small group targeted teaching) caught up to their peers in word reading and spelling skills.

Conclusions: With appropriate resourcing, planning and engagement with communities, successful large-scale implementation of evidenced-based early literacy approaches is possible within a relatively short time. Culturally responsive implementation within MTSS frameworks holds much promise for raising literacy achievement for all children.

KEYWORDS

literacy, reading, phoneme awareness, multi-tiered systems of support, teaching, intervention, implementation

Introduction

The World Health Organization (WHO) has highlighted the critical importance of literacy to reducing current global health inequities. It has called upon governments worldwide to increase investments in building their population's comprehensive literacy skills, particularly within diverse communities (WHO, 2020). This call to action is reinforced by the United Nations Sustainable Development Goals for 2030. Ensuring equitable and quality education for all is one of 17 goals agreed to by contributing partner countries to ensure the prosperity of our people and our planet (United Nations, 2015). Improved literacy, education and health outcomes in a population begins with early literacy success. Strong foundational linguistic and early reading skills underpin later reading abilities (Cunningham and Stanovich, 1997; Tunmer et al., 2006; Sparks et al., 2014; Zubrick et al., 2015; Lepola et al., 2016; Su et al., 2017; Peng et al., 2019). In turn, reading competency supports the development of comprehensive literacy skills, such as the ability to evaluate, critique, and analyze written information in its many forms. Those children

with strong early reading skills are more likely to experience cumulative advantages compared to children who struggle with early reading (Juel, 1988; Stanovich, 2009; Ritchie and Bates, 2013).

Despite our advanced knowledge of facilitators of children's early literacy success, there remains wide variability in children's literacy achievement. Only modest gains (if any) are being reported in reducing educational inequities for indigenous populations, children raised in poverty or children with learning challenges (Ferrer et al., 2015; Schulte et al., 2016; Stone et al., 2017; Hanushek et al., 2022). For some children with lower level literacy skills the achievement gap compared to their peers over time may be widening (D'Agostino and Rodgers, 2017). In addition, the COVID-19 world pandemic has led to severe disruptions to children's learning (Relyea et al., 2023) having a profound adverse effect for some of our diverse communities. Urgent collective effort is therefore required to ensure all children develop the necessary foundational literacy skills that support comprehensive literacy abilities to achieve our global literacy aspirations.

Understanding pathways that lead to the successful wide-scale implementation of research informed literacy teaching is critically important. Insights gained from how differing countries or regions address such issues (e.g., Piper et al., 2018; Ramacciotti et al., 2023) may inform practices internationally to more effectively address our global literacy challenges. This paper presents a case study of an early literacy initiative introduced by the New Zealand Government to reduce persistent education and health inequities. The initiative aims to advance strong foundation literacy skills for all learners with a particular focus on supporting young Māori (New Zealand's indigenous population) and children from Pacific Peoples[1] ethnic group for whom disparity in education outcomes is greatest (May et al., 2019). Factors that enabled an early literacy approach referred to as the *Better Start Literacy Approach (BSLA)* to extend from development phase, through controlled research trials (Gillon et al., 2019, 2020, 2023a,b) to national implementation are described.

Multi-Tiered Systems of Support

Multi-Tiered Systems of Support (MTSS) are advocated to ensure all children succeed in their learning. MTSS replace Response to Intervention frameworks (RTIs) through taking a more comprehensive and systems wide perspective to remove barriers to students' learning (Averill et al., 2011; Pullen et al., 2018; Freudenthal et al., 2023). MTSS may be used to support the academic, social, emotional or behavior development of children and youth (Averill et al., 2011). In relation to children's reading development, MTSS are based on the "science of reading"—cumulated knowledge gained over many years through robust scientific methods related to effective reading instruction and reading interventions. MTSS are specifically designed to alleviate the adverse behavioral, education and health outcomes frequently associated with reading failure to focus instead on ensuring each child experiences reading success (Gonzalez et al., 2022). The multiple tiers of support within such frameworks begin with research-based reading instruction for all children (or Tier 1 quality universal reading instruction). Children who don't respond as well to this first Tier of reading instruction (identified through valid and reliable assessment methods) are provided with supplementary tiers of support such as Tier 2 support (typically involving small group targeted instruction) or Tier 3 support (individualized support to meet a specific learner's literacy needs).

Research evidence supports the effectiveness of MTSS frameworks in developing children's early literacy skills (for review, see Fien et al., 2021). Overall, the evidence shows Tier 1 teaching approaches that provide explicit and systematic instruction in phonics, phoneme awareness, and fluency are effective at promoting early literacy development (Fuchs and Deshler, 2007; Al Otaiba et al., 2019; Petscher et al., 2020). Relatively few studies have focused on the impact of structured and systematic teaching in vocabulary and oral language comprehension within Tier 1 teaching, although these are also considered important elements of quality class-wide literacy teaching (Petscher et al., 2020). Wanzek et al. (2016) conducted a meta-analysis to evaluate outcomes for children's foundational literacy skills from a pool of 72 experimental or quasi-experimental studies that investigated the effectiveness of Tier 2 type interventions for children in kindergarten to grade 3. Their analyses indicated moderate positive effects of Tier 2 type interventions on foundational reading skills (e.g., phonological awareness and recognition skills) as evidenced by improvements on both standardized and non-standardized assessment tasks. Small to moderate effects were found for children's language comprehension on standardized measures with large effect sizes reported for non-standardized language measures. There was no difference in outcomes between individual or small group interventions (e.g., up to five children in a group) nor in length of session (e.g., 30 min compared to 1-h sessions). The type of professional implementing the intervention (e.g., researcher, class teacher, special educator) also did not influence outcomes (Wanzek et al., 2016). These analyses suggest that large scale implementation of Tier 2 teaching implemented in relatively cost-effective ways in regular education settings is feasible. Tier 2 may be particularly effective when it is in addition to quality Tier 1 teaching (rather than replacing it) and when it is implemented early in children's first year at school (Al Otaiba et al., 2014).

Despite the evidence-base documenting positive effects of literacy teaching within MTSS (and Response to Intervention) frameworks, the majority of published research in this area falls into the categories of exploration, development, and initial efficacy studies (Solari et al., 2020). There is a need to build toward clinical implementation studies (evaluation of effective interventions within real world settings) and public health studies (i.e., evaluation of population/widespread outcomes of effective interventions) (Fuchs and Fuchs, 2017; Solari et al., 2020). This current paper advances knowledge of MTSS through examining data collected from both Tier 1 and Tier 2 literacy teaching in a large-scale implementation of a comprehensive early literacy approach within the New Zealand education system.

1 Pacific Peoples ethnicity refers to Pacific Peoples groups (including: Sāmoa, Tonga, the Cook Islands, Niue, Tokelau, Tuvalu, and other smaller Pacific Peoples nations) who are now living in New Zealand (NZ Stats 2018 Census).

Assessment within MTSS

A key feature of MTSS for early literacy instruction is the regular monitoring of children's progress in foundational literacy skills. The aims of these assessments are to (a) monitor children's progress in response to evidenced based literacy teaching in order to determine next steps for teaching and (b) identify those children who require additional, instructional support (either Tier 2 or 3 of the MTSS). MTSS are potentially useful for the early identification of children with word reading problems such as dyslexia. Carefully monitoring children's response to literacy instruction based on the science of reading in their first year at school may be more helpful than many dyslexia screening assessments given the psychometrics limitations of such screening assessments (Catts and Hogan, 2020).

To be effective, however, monitoring assessments within MTSS need to be both rigorous and efficient (Petersen and Stoddard, 2018). Most studies that include progress monitoring assessment data within MTSS literacy frameworks focus on reliable measures related to word recognition (e.g., phonic, phonological awareness, word reading fluency). Freudenthal et al. (2023) summarized 22 MTSS studies between 2017 and 2020. These studies included a range of researcher and teacher-implemented interventions and measures. However, none of these studies included a monitoring measure of children's oral language or listening comprehension. Given that both word recognition and language comprehension are important for skilled reading comprehension (Hoover and Tunmer, 2018), understanding valid and reliable language monitoring assessments for class teachers to use is important as we advance our understanding of the benefits of MTSS frameworks.

This current study extends previous research in oral narrative assessments and teaching within classroom contexts (Israelsen-Augenstein et al., 2022; Petersen et al., 2022; Gillam et al., 2023). It also extends recent research in the potential benefits of digital technologies and artificial intelligence to improve efficiency in literacy monitoring assessments (Adlof and Hogan, 2019; Fox et al., 2022; Bright et al., 2023) We examine data collected through a large scale implementation of teachers' administrating a series of online tasks to 5–6 year old children as part of their regular teaching practices within a MTSS framework. The tasks included assessment of children's oral narrative (story retell), listening comprehension, phonic, phoneme awareness, and non-word reading and spelling.

Change processes

It is widely recognized that there are significant barriers to achieving sustained change in teachers' literacy assessment and teaching practices (Le Fevre, 2014). To acknowledge and overcome such barriers the implementation of large-scale literacy initiatives requires well-planned change processes at multiple levels (Solari et al., 2020). Understanding the systems changes necessary and effective methods to support the uptake of evidenced based literacy teaching practices into routine class teaching is examined within the fields of Implementation Science and Translational Science. Emerging research and discussions in translational science related to reading highlight the need for a strong communication strategy, engagement with community and a team of researchers with the skills and dispositions necessary to facilitate the successful translation of research evidence into literacy teaching practice (Solari et al., 2020). A partnership approach with school leaders, researchers, policy makers, community leaders, cultural and indigenous leaders all working together in co-constructed processes to support change is critically important (Macfarlane and Macfarlane, 2019). Provision of quality professional learning and development for teachers, ensuring teaching fidelity is aligned to evidence-based practices, valuing and celebrating teachers' success in uplifting their children's literacy achievement, adequate resourcing and time for teachers to enhance their assessment and teaching strategies all need to be addressed. Flexibility within literacy approaches to ensure literacy practices meet the needs of diverse learners and acknowledge local cultural contexts are also important to consider (Macfarlane et al., 2016; Hoover and Soltero-González, 2018). This paper describes the development of an early literacy approach specifically designed for large scale implementation based on these principles of facilitating research evidence into changing or enhancing teacher's current literacy teaching practices.

Study aims

Specifically, the aims of this paper are to:

1. Present a case study within the New Zealand education context of the large scale implementation of the Better Start Literacy Approach (BSLA). Key features of the approach and its development from controlled research trials (2016–2019) to a government funded national implementation in over 46% of the country's primary schools (from 2020 to 2023) are described.
2. Examine data routinely collected as part of BSLA large scale implementation to provide insight into the benefits of such implementation for enhancing children's foundational literacy skills.
3. Analyze teacher self-reported fidelity data to gain insight into levels of adherence to evidence-based practices within large scale implementation.

Methods

The data collection method and process reported were approved by the University of Canterbury's Human Ethics Committee (application 2021/06/LRPS). The approval allowed the researchers to analyze de-identified assessment data routinely collected within the national implementation of BSLA for secondary analyses to evaluate the impact of the Better Start Literacy Approach. The committee determined this was low risk and individual consent for participation was not required. However, as children entered BSLA, parents were provided with an option to exclude their child's anonymized data from the database used for any secondary analyses.

New Zealand education context

New Zealand is a Commonwealth country of ~5 million people with 86% of the population living in or near a major city. Māori, around 15% of the population, are the indigenous peoples of New Zealand. European settlers exceeded the number of Māori in the nineteenth century and in the 2018 census 74% of the population identified as being of European decent, 11% identified as Asian and 8% from the Pacific Peoples islands. English is the main language spoken but both Te reo Māori and NZ Sign Language are official languages of the country.

Education is free in New Zealand for children aged 5–19 with it being compulsory for children aged 6–16 years to attend school. Prior to school entry 95% of children will have attended some form of early childhood education. The Government currently provides up to 20 h a week of free early childhood education for children aged 2–5 years. The ECE sector's curriculum framework is Te Whāriki (Ministry of Education, 2017). This promotes a holistic, socio-cultural perspective in learning and teaching across across five learning strands (wellbeing, contribution, belonging, communication, and exploration). Teachers use a learning progressions framework to monitor children's growth across these strands. There are currently no required formal assessment measures of emerging literacy skills during early childhood or as children transition to school.

Children often start school on or close to their 5th birthday and formal literacy instruction typically begins from the outset of their attendance at school. In relation to literacy instruction, teachers working in English medium schools (96% of all primary schools) implement the New Zealand English curriculum (Ministry of Education, 2007). This provides a broad framework and learning progression indicators to monitor children's oral and written language development. Schools are self-governing and although the New Zealand Ministry of Education provides Government funded schools with several quality literacy resources, schools and teachers can make their own decisions about the types of literacy approaches and interventions they implement. Reading Recovery has been the main government funded approach to support young struggling readers (Clay, 1990). However, given evidenced informed criticism of both the theoretical underpinning of the approach and longer term reading outcomes of children who have received Reading Recovery (Tunmer et al., 2013; Nicholas and Parkhill, 2014; Chapman and Tunmer, 2016; D'Agostino et al., 2017; May et al., 2023), there has been a steady decline in the implementation of Reading Recovery in NZ schools since 2012. Only 41% of eligible schools implemented Reading Recovery in 2021 down from 51% in 2019. Outside of Reading Recovery there is considerable variability in literacy assessments and interventions used in New Zealand for struggling readers (Arrow et al., 2022).

New Zealand's literacy rankings

Data from the 2021 Progress in International Literacy Study (PIRLS) shows that in 4th grade the average reading achievement score for the New Zealand participating children was significantly above the center point of the PIRLS scale (CenterPoint = 500) (Mullis et al., 2023). New Zealand's average scaled reading score was 521 (Std. error 2.3) which resulted in a rank of 27th out of 57 participating countries. Since 2006, there has been a consistent and small decline in children's reading achievement in New Zealand (with a drop from a scaled score of 532 in 2006). Despite performing relatively well on average, the wide variability in students' reading achievement performance is an area of persistent and ongoing concern. In both the 2016 and 2021 PIRLS, variability between New Zealand's high and low achieving readers by 4th grade is larger than any other country scoring above the PIRLS Centrepoint on reading achievement (Mullis et al., 2017, 2023). Like most countries, girls performed significantly higher in 2021 (scaled score of 531) than boys (scaled score of 512) with the average difference of 19 scale score (Mullis et al., 2023).

In recognizing the persistent inequities in education outcomes in New Zealand, the Government launched a Literacy and Communication and Math Strategy (Ministry of Education, 2022) to uplift foundational skills for all children. As part of an early literacy initiative aligned with this strategy the Government has funded the implementation of the *Better Start Literacy Approach (BSLA)* (Gillon et al., 2023a). From 2020 to May 2023 over 3,000 junior schoolteachers and literacy specialists have been trained in the BSLA in 819 schools The next section describes the development of BSLA and its pathway from controlled trials to this large scale national implementation.

Better Start Literacy Approach

Development phase

The BSLA emerged from the first phase of a 10-year Government funded programme of research focused on child and youth wellbeing. This program of research is referred to as the—The Better Start National Science Challenge (Maessen et al., 2023). It involved a multidisciplinary team of research leaders across New Zealand engaging with community to co-construct research that would lead to sustained change and positively impact communities. Following extensive community consultation, three key areas for research were identified: successful learning, healthy weight, and youth resilience. As part of the successful learning theme, a 5-year programme of research was funded to uplift children's early literacy success.

Key aspects of this research planning involved engaging with community and co-constructing evidenced based literacy teaching instruction for children's first year at school that could be implemented at scale across school communities. A MTSS framework with strengths-based and culturally responsive literacy teaching practices was proposed based on research findings and community feedback.

Cultural framework

A *He Awa Whiria* or "braided rivers" framework (Gillon and Macfarlane, 2017) was used as a framework to consider the braiding of research evidence from the scientific literature (inclusive of research findings involving diverse populations, and differing research methodologies) together with Māori knowledge

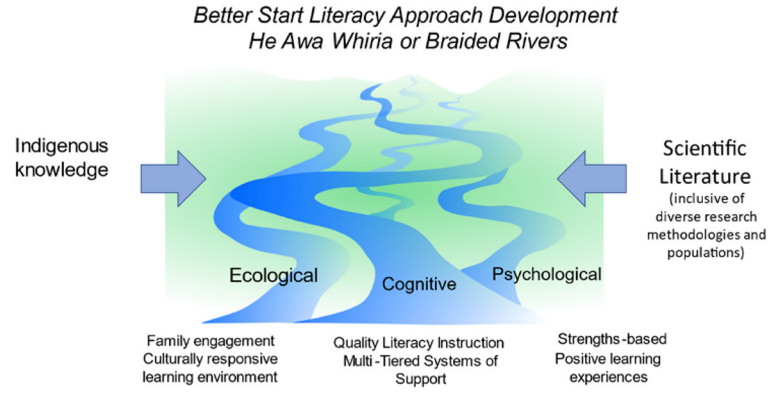

FIGURE 1
Highlights the *He Awa Whiria* framework for BSLA of braiding together multiple influences on children's early literacy development.

and evidence indicating facilitators of education success for Māori and Pacific Peoples (e.g., Berryman and Eley, 2017; Webber and Macfarlane, 2020; Alansari et al., 2022). Our *He Awa Whiria* framework for literacy acknowledges the importance of research related to necessary cognitive skills to literacy learning, but also acknowledges the influence of ecological, and psychological factors in children's reading development (Aaron et al., 2008) (see Figure 1). The braided river metaphor conceptualizes that just as the streams of a braided river may sometimes flow apart so too does indigenous knowledge and domain specific research knowledge have its own integrity and importance. However, when streams of water braid together and form a larger body of water, the river gains force and strength. So too through the braiding of differing knowledge streams the potential to ensure all children experience early literacy success gains momentum. The schema used within BSLA to support teachers in developing culturally responsive literacy teaching practices within the He Awa Whiria framework is *The Hikairo Schema for Primary Teachers* (Ratima et al., 2020). Three core principles within this schema include:

1. *Relevance:* Teachers provide learning experiences that are relevant for children's cultural and personal identities;
2. *Balance of power:* Mutual care, trust and respect between teachers and learners is fostered through the learning experiences; and
3. *Scaffolding:* Teachers scaffold learning and provide necessary supports to ensure successful learning outcomes are in the grasp of all learners.

Key features of BSLA content

There are several key features of BSLA teaching content that stem from the *He Awa Whiria* framework (Gillon et al., 2024). These are summarized in the following section.

Cognitive stream of knowledge

The structured literacy approach to teaching within BSLA is aligned to the Simple View of Reading (see Hoover and Tunmer, 2018). Teachers use explicit teaching techniques to develop children's skills that support both their word recognition and their language comprehension skills. Consistent with MTSS frameworks, structured daily 30-min lesson plans are provided for Tier 1 class teaching that develop children's skills in three areas:

1. Vocabulary, oral narrative and listening comprehension through shared book reading activities.
2. Phonic, phoneme awareness and morphological awareness activities following our BSLA phonic scope and sequence (see Supplementary material: in the first 10 weeks of teaching, new phonic patterns are introduced at a rate of between 2 and 4 new patterns per week with flexibility for teachers to revise target phonic patterns as suitable to their learners).
3. Activities to support children to transfer their developing phonic and phoneme awareness skills to the reading and writing process.

In addition to these 30 min class lessons, children also participate in a 10–15 min small group instructional reading sessions using a newly developed decodable text reading series that the BSLA leaders were involved in developing called Ready to Read Phonics Plus (RRPP) (see Arrow et al., 2021 for details of reading series). The phonic scope and sequence for BSLA and the RRPP reading series are aligned. This alignment aims to maximize children's exposure to target phonic and phonological patterns from class activities to small group reading with connected text.

For children identified as requiring Tier 2 support teachers implement 40 small group lesson plans (usually four 30-min session over a 10 week period). These lessons are focused on strengthening children's phonic, phoneme awareness, word reading and spelling skills. This support is typically supplementary to children participating in Tier 1 class activities. There is strong alignment between Tier 1 and Tier 2 teaching activities to maximize teaching intensity for learners who require additional support.

With the support of literacy specialists or speech-language therapists, teachers adapt BSLA assessments and Tier 1 and Tier 2 activities for children with complex communication needs who require Tier 3 level support (Clendon et al., 2022). This support is provided at an individual level and often involves teacher assistants who are supporting these learners.

Important teaching elements within BSLA include the following:

- BSLA teaching time is focused on phonological awareness skills at the phoneme level (identifying, blending, segmenting, and manipulating phonemes). Syllable and rhyme awareness are not teaching targets within the 30-min daily lessons. This is based on evidence that phoneme skills are more strongly associated with early reading and spelling (Hulme et al., 2002) and that teaching to improve children's syllable and rhyme awareness does not necessarily transfer to reading improvements (Nancollis et al., 2005).
- Based on evidence from earlier experimental studies (e.g., Hatcher et al., 1994; Gillon, 2000) many of the BSLA teaching activities integrate phoneme-grapheme knowledge phoneme awareness and the reading and spelling of real words concurrently to maximize teaching efficiency.
- The pace of teaching through the phonic scope and sequence is aligned to the Self Teaching Hypothesis of learning to read (Share, 1995). That is, the explicit teaching of phonic patterns and decodable text within the readers is not too tightly constrained to slow learning down. Rather, it allows children to discover a limited number of untaught orthographic patterns alongside the explicitly taught patterns, particularly after the first few weeks of BSLA teaching.

Ecological stream of knowledge

A key aspect of BSLA is supporting teachers to implement BSLA in culturally responsive ways. This involves supporting teachers to develop their own personal cultural competence (Ratima et al., 2020) as well as encouraging them to relate the stories and activities in culturally relevant ways to their learners. In addition, content for family workshops, family reports and weekly updates about BSLA teaching are provided as a guide for teachers to engage families in their children's literacy learning. Typically teachers followed these guides to offer two sessions for children's parents and family members during the first 10 week period of implementing BSLA.

Psychological stream of knowledge

BSLA is developed around strengths-based principles (Gillon et al., 2023a). Teachers are encouraged to focus on what children can achieve and identify their next steps for learning. The assessment data are used and presented in positive ways to focus on students' growth and celebrate their learning and effort. Each child is expected to succeed. The use of children's assessment data to inform teaching along with scaffolding and explicit teaching techniques are promoted to ensure all children experience success in their learning. The consistency and routine embedded within the lesson plans help support children's confidence in engaging in the activities and reduces cognitive load to support learning success.

Professional learning and development (PLD) for teachers

Through an iterative process involving consultation and feedback from teachers and literacy specialists, we developed a series of online BSLA micro-credentials for class teachers, literacy specialists (e.g., teachers with advanced qualifications in children's literacy development and speech-language therapists) and teacher aides (Scott et al. submitted[2]). Micro-credentials provide teachers with flexibility to engage in PLD in specific learning areas and to gain recognition for their learning (DeMonte, 2017). The online content was available through the University's student learning management system and provided a high-quality online learning experience. The content included theoretical content related to the science of reading (modified for teacher aides), activities to develop professionals' personal linguistic knowledge, access to the Tier 1 and Tier 2 daily lesson plans, a variety of video demonstrations of teaching lesson with commentary around effective teaching strategies, training in the assessment tasks and data interpretation (modified for teacher aides), content related to developing culturally responsive teaching practices, access to materials to support family engagement and access to adapted teaching and assessment content to support children with complex communication challenges. Each professional enrolling in the BSLA micro credential was required to successfully complete online assessment tasks and online quizzes to gain their micro-credential. In addition, class teachers were required to complete the teaching of at least 10 weeks (or 40 lesson plans) of BSLA. Literacy specialists and teacher aides were required to provide support to class teachers during these 10 weeks of teaching. Teachers had continuous access to this online learning content (even after they had completed their micro-credential) to enable them to revise content, review lesson demonstrations or engage more deeply in learning content as they progressed with BSLA implementation.

As well as online independent learning teachers engaged weekly in live (or recorded) Zoom sessions with the BSLA leadership team for the first 10 weeks of teaching BSLA. As part of the BSLA micro-credential for literacy specialists, participants also attended a 2-day live workshop with the BSLA co-developers and leadership team. This workshop focused on their role as mentors and coaches for class teachers as well as more in depth training in key areas such as BSLA assessments and data interpretation.

Online assessments designed for large scale implementation

A bespoke assessment website was developed to accommodate the online BSLA assessment tasks (Scott et al., 2022; Gillon et al., 2023b). Each teacher has login access to their own class data held within the assessment website. An assessment dashboard provided teachers with details of each child's assessment task performance and alerted teachers to when the next round of monitoring assessments for a particular child was due and what assessment tasks required completing. The online assessment tasks that children completed via iPad or computer ensured consistency of task presentation, reduced teacher workload, and animated characters engaged young learners in the tasks.

2 Scott, A., Gillon, G., McNeill, B., and Gath, M. (submitted). Facilitators of success of a microcredential model for professional learning and development to enhance teachers' literacy practices.

Automatic scoring and automated reporting, features supported large scale implementation.

Research trials to inform large scale implementation

Following the design and development of BSLA, the approach was first trialed in new entrant and Year 1 classrooms (5- to 6-year-old children) from seven schools in areas of high social-economic deprivation. Data from this quasi-experimental research trial suggested that BSLA was more effective in accelerating children's foundational literacy skills than the classroom literacy approaches teachers were currently implementing (Gillon et al., 2019, 2020). The Ministry of Education funded a replication of this research trial. The replication involved 14 schools (seven of which were in a different regional area than the first trial), from across socio-economic deprivation deciles (Gillon et al., 2023a; McNeill et al., 2023). The data showed significant positive effects for both Tier 1 and Tier 2 BSLA implementation and demonstrated that BSLA was equally effective across schools of differing socio-economic deprivation deciles. Following a competitive Government procurement process, the BSLA research team were selected to provide PLD to support teachers and literacy specialists to implement BSLA across the country. It also extended the development of BSLA from supporting children in new entrant and Year 1 classes to supporting children in Year 2 classes.

Multidisciplinary team

The importance of a building a team with a variety of discipline knowledge, leadership skills and dispositions to support large scale implementation of evidenced based literacy intervention and systems change has been identified (Solari et al., 2020). The co-lead developers of BSLA (first two authors of this paper) had a strong research and leadership background in speech-language therapy, teacher education and more broadly children's wellbeing. The academic leadership team for BSLA includes leaders and experienced practitioners in junior school class teaching, Māori education, education psychology, Pacific Peoples education and health, speech-language therapy, developmental child psychology, and specialists in children's learning and behavior. This team collectively support the teachers and literacy specialists through the PLD and in school implementation of BSLA. The team engaged with a software company in the design of the assessment website as well as online learning experts in the development of the PLD micro-credential. In addition, the large-scale implementation was supported by an advisory group comprising international leaders in the science of reading, children's language development and cultural leaders in Māori and Pacific Peoples communities.

Participant data included in analyses

This paper describes data routinely collected through the national implementation of BSLA. By default, all assessment data is transferred to an anonymized research dataset; however, all parents are given the option to have their child's data excluded from the research data base. We have used this anonymized research dataset for our analyses.

As at May 2023, assessment data were available for 29,795 five-year-old students (aged 5 y 0 m to 5 y 11 m) at the Baseline assessment point (prior to teachers implementing BSLA). Further assessment data were available for the subsequent monitoring assessment points within BSLA for the following numbers of students:

- 16,497 students at the Ten Week assessment point. Teachers administer these assessments after children have received 10 weeks of BSLA Tier 1 teaching or ∼40 BSLA lessons.
- A subset of 7,876 children from the larger cohort who were identified as the "school entry cohort" as they were aged 5 y 0 m to 5 y 3 m at baseline assessment and therefore received BSLA from the start (or within a couple of months from the start) of their formal literacy learning.
- 5,495 students with greater learning needs who were assessed at the Twenty Week assessment (i.e., after they had received 10 weeks of BSLA Tier 2 teaching).
- 2,313 students at the Thirty Week assessment. All children participate in the Thirty Week assessment round after they have received ∼30 weeks of BSLA teaching. At the time of this analysis, data was available for 2,313 children who had received both their Baseline assessment and Thirty Week assessments.

The students included in the data analyses were from 819 schools, spread across all 16 Regional Council areas in New Zealand. The full sample for this analysis had a mean age of 64.3 months (SD = 3.3) at the Baseline assessment point. Children's ethnicities were as follows: New Zealand European (48.3%), Māori (22.1%), Pacific Peoples (9.1%), Asian (12.7%). Student gender: 49.0% female, 51.0% male, and 0.01% other.

Students in the dataset came from all levels of socio-economic deprivation based on the New Zealand Index of Deprivation (Atkinson et al., 2019) (see Supplementary material for details). A Kolmogorov-Smirnov test indicated significant skew to the distribution of socioeconomic deprivation ($D = 0.11$, $p < 0.01$; skewness = −0.20, kurtosis = −1.06). The skewed distribution was due to the Ministry of Education initially prioritizing schools in areas of higher socio-economic deprivation for this Early Literacy Initiative opportunity.

From teacher report, 3,328 (11.2%) children were classified as English Language Learners (ELL). These children were predominantly of Asian ethnicity (51.0%) or Pacific Peoples ethnicity (21.5%). Gillon (2023) provides a discussion of these learners' progress within BSLA.

Assessment measures

All tasks (except for the non-word spelling task) were presented via a touch screen iPad or laptop. An animated character (male voice with New Zealand accent) spoke to the children for the

phoneme awareness and phoneme grapheme matching tasks. The children were required to touch the screen (iPad), or the teacher used a computer mouse to select the child's response. The children's responses were automatically recorded and a detailed response analysis for each child's attempt was immediately available to the child's teacher. Previous research established both the test-retest reliability for the phoneme awareness and phoneme grapheme matching tasks (Carson et al., 2015) as well as the internal consistency of the task items (Gillon et al., 2023a). Gillon et al. (2023b) describe the validity of the online story retell assessment. The tasks on the assessment website were organized in the following order, but teachers could choose to administer the tasks in any order.

Initial phoneme identity

In this task, the animated character asked children to select one of three pictures shown on the screen or iPad that started with the target sound. For example: "*Dog likes words that start with a /d/ sound. Can you help dog find words that start with his favourite sound. Which word starts with /d/*?" The phonemes tested were: /m/, /s/, /k/, /b/, and /f/ (two items tested each phoneme). Proficiency on this task was set at 8 or more items correct out of 10. Cronbach's alpha across the 10 items was 0.85, indicating high reliability.

Phoneme blending

In this task, an animated character instructed the children: "*I am going to say one of these words very slowly. Click on the picture you think I am saying.*" For example, children were shown the images of a cake, a cape, and a ring, while listening to the character say the individual phonemes of /k//ei//k/. They were then required to click the image of the word they thought the character was saying. Children completed 10 test items. Proficiency on this task was set at 8 or more items correct out of 10. Cronbach's alpha across the 10 items was 0.79, indicating good reliability.

Letter-sound knowledge

In this task children were required to select the letter (grapheme) that matched the speech sound (phoneme) that they heard by tapping on the screen. This letter-sound matching was assessed with two sets of items (eight items in Set 1; m, d, c, t, s, l, n, p), 17 items in Set 2 (Set 2: ch, b, i, f, r, g, e, sh, k, u, j, w, o, a, v, th, h), with children only progressing onto Set 2 if they achieved proficiency at Set 1. Proficiency on Set 1 of this task was set at 6 or more items correct out of 8, and proficiency on Set 2 was set at 13 or more out of 17. Cronbach's alpha was 0.84 for Set 1 and 0.83 for Set 2, indicating high reliability.

Non-word reading

Students were asked to read sets of non-words and their responses were audio-recorded and scored. Children received two points for every correct grapheme within the non-word, with one point given for a partially correct grapheme (such as an added sound or a b/d or p/q reversal error). A total of four sets of non-words (each set containing 10 words) were used, each with an increasing level of difficulty (level 1 example: wep; level 2 example: flom; level 3 example: gace; level 4 example: spream). Students started at Set 1 of non-word reading and only progressed to subsequent sets if they were proficient (scored 80% or higher) on the previous set. A total non-word reading score was calculated as the sum of all four sets (with 0's imputed for any sets that students did not progress to) out of a maximum of 264 points. Interrater reliability was assessed by having a single independent coder score a random sample of 200 non-word reading assessments. The two-way random effects intraclass correlation was 0.92 across the 10 items across sets, indicating high interrater reliability.

Non-word spelling

The non-word spelling task was a parallel task to non-word reading, with students asked to spell the same four sets of non-words presented in the non-word reading task. The teacher read aloud the non-word and asked the child to write the word onto a record form. Students received a score of 0, 1, or 2 for each grapheme depending on the spelling accuracy. A total non-word spelling score was calculated as the sum of all four sets (with 0's imputed for any sets that students did not progress to) out of a maximum of 264. Interrater reliability was assessed by having a single independent coder score a random sample of 200 non-word spelling assessments. The two-way random effects intraclass correlation was 0.93 across the 10 items across sets, indicating high interrater reliability.

Oral narrative

Students' oral narrative skills and listening comprehension were assessed through a novel story retell task presented digitally on an iPad or computer (Gillon et al., 2023b). In this task, students listened to a story with accompanying pictures on the screen. Following the presentation of the story, children were prompted to retell the story in their own words. The recordings of the children's retells were then transcribed automatically using speech-to-text software (Scott et al., 2022), and analyzed using Systematic Analysis of Language Transcripts (SALT) software (Miller and Chapman, 2004). For each student's transcript of their story retell, the following quantitative data was automatically generated for teachers: number of words used, number of different words used, number of nouns, verbs, adjectives, and adverbs used, and percentage of intelligible utterances. In addition, the research team had access to more detailed SALT analyses including mean length of utterance.

Following the retell, children were asked five comprehension questions which were presented on the screen for the teacher to read aloud. Three of the questions were factual questions tapping understanding of character identities and actions. Two were inferential questions tapping comprehension of story context and motivation for character's actions. Teachers scored children's responses online (usually in real time). Each response was scored as 2 for correct, 1 for partially correct, or 0 for incorrect. Interrater reliability of comprehension scoring was assessed by having a single independent coder score a random sample of 200 comprehension tasks. The two-way random effects intraclass correlation was 0.52 across the five items, indicating moderate interrater reliability.

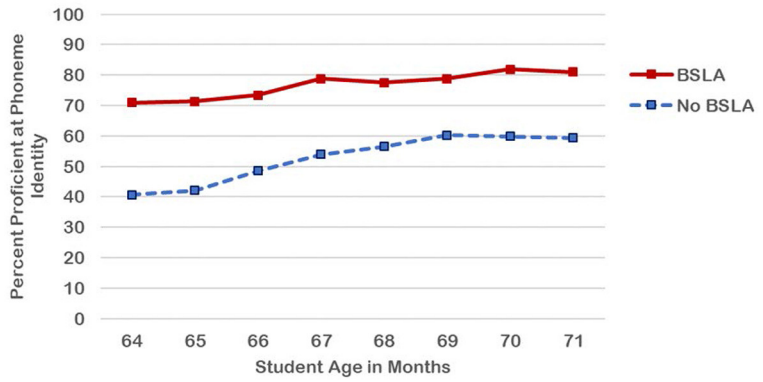

FIGURE 2
Percent of students proficient at initial phoneme identity before and after 10 weeks of BSLA teaching.

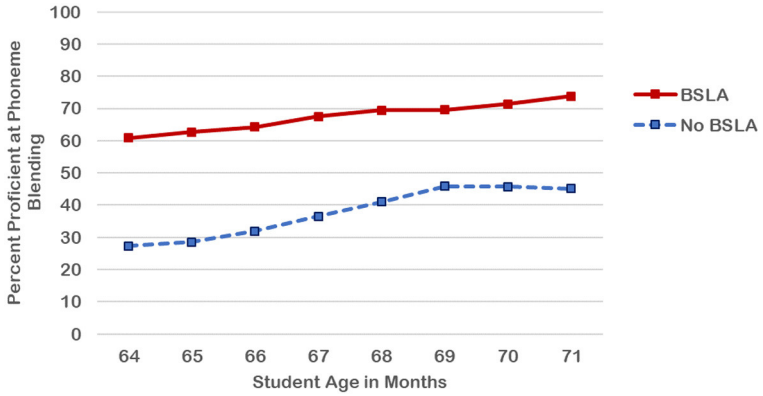

FIGURE 3
Percent of students proficient at phoneme blending before and after BSLA teaching.

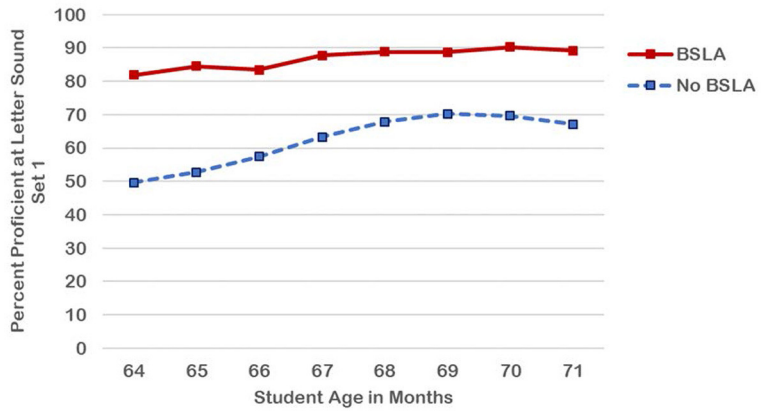

FIGURE 4
Percent of students proficient at Set 1 letter-sound knowledge before and after BSLA teaching.

Scores used in analysis

On all measures other than the oral narrative, analysis was undertaken on both proficiency levels and raw scores. Proficiency was defined as scoring 80% or higher on the task. When using raw scores, data was imputed at subsequent assessment points once students reached proficiency on the task (as tasks were not repeated once proficiency was demonstrated) (see notes in Supplementary material).

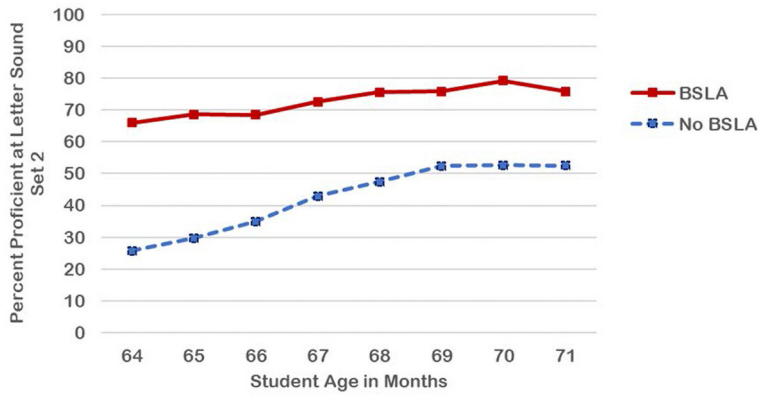

FIGURE 5
Percent of students proficient at Set 2 letter-sound knowledge before and after BSLA teaching.

BSLA teaching fidelity data

Teachers were requested to complete online fidelity checklists during their first 10 weeks of BSLA teaching. On these checklists, teachers reported on student engagement, total instructional time for class/large group lessons and small group reading, and the activities that were included in the large and small group sessions by selecting all that applied from a list of potential activities.

Results

Skill progression before and after BSLA teaching

Our first set of analyses used a subset of student Baseline data as an internal control group through a comparison of students of the same month of age who had not yet received BSLA teaching (Baseline assessment internal) against those who had received approximately 10 weeks of BSLA teaching (Ten Week assessment). Figures 2–5 plot the percent of students proficient on each assessment task by student month of age at the Baseline and the Ten Week assessment points. Note that these figures begin at 64 months of age to allow for enough children per cell at the Ten Week assessment point. Thus, the figures include 15,380 children at Baseline (internal control group) and 10,532 children at Ten Week assessment (BSLA treatment group).

After 10 weeks of BSLA teaching, students of all months of age showed a higher rate of proficiency on all tasks than students of the same age who had not yet received BSLA. The gap is particularly large for the youngest students (e.g., ages 64–66 months). Effect sizes of the difference between the BSLA group and the no-BSLA group at each month of age on each task indicated medium to large effects of BSLA teaching (initial phoneme identity, Hedges' g range 0.38–0.64; phoneme blending, Hedges' g range 0.48–0.76; letter-sound knowledge set 1, Hedges' g range 0.42–0.73; letter-sound knowledge set 2, Hedges' g range 0.49–0.90). See Supplementary Table 1 for means, standard deviations, and effect sizes.

Growth in assessment scores by ethnicity

We next examined the change in raw scores from Baseline to the Ten Week assessment for each ethnic group. Large effect sizes ranging from 0.63 to 1.81 indicated all groups scores increased in a meaningful way. Effect sizes were higher for Māori and Pacific Peoples relative to NZ European and Asian students on all tasks aside from phoneme blending (see Supplementary Table 2).

Repeated measures ANOVAs with a within-subjects factor of time and between-subjects factor of ethnicity were used to determine whether growth differed significantly by ethnicity. For this analysis, students could not be included in multiple ethnic groups, and we have coded a prioritized ethnicity variable using the following prioritized order: Pacific Peoples, Māori, Asian, NZ European. Any significant Time*Ethnicity interactions (indicating different growth patterns by ethnicity) were followed up with Bonferroni *post-hoc* testing. Table 1 provides the results of this analysis, with differing subscripts indicating significantly different growth between ethnic groups.

The overall effect of differing growth by ethnicity was minimal for initial phoneme identity and phoneme blending tasks. However, meaningful effects were found on both phoneme-grapheme tasks (equivalent to $d = 0.29$ for set 1 and $d = 0.20$ for set 2). In both cases Māori and Pacific Peoples groups showed more growth than Asian and NZ European students.

Figure 6 illustrates the growth in Phoneme-Grapheme knowledge (set 2) from Baseline to the Ten Week assessment by ethnicity. The figure demonstrates a reduction in the gap between ethnicities at Ten Week as compared to Baseline.

School entry cohort

This set of analyses focuses on the 7,883 students who were aged 5 y 0 m to 5 y 3 m at the Baseline assessment and who were re-assessed after 10 weeks of BSLA teaching. This represents children who had just commenced school when BSLA teaching was introduced. We examined the change in scores as well as changes in proficiency rates from school entry (Baseline) to

TABLE 1 Interactions between time and ethnicity in repeated measures analysis.

Assessment	Repeated ANOVA Time*ethnicity effect	Estimated marginal mean growth (Std. error) Pacific peoples	Māori	Asian	NZ European
Initial phoneme identity	$F_{(3,14,580)} = 13.66$, $p < 0.001$; $\eta_p^2 = 0.003$	2.41 (0.08)[a]	2.40 (0.05)[a]	1.97 (0.07)[b]	2.14 (0.03)[b]
Phoneme blending	$F_{(3,13,692)} = 7.55$, $p < 0.001$; $\eta_p^2 = 0.002$	2.17 (0.07)[b]	2.25 (0.05)[b]	2.39 (0.06)[ab]	2.46 (0.03)[a]
Letter-sound knowledge (Set 1)	$F_{(3,15,888)} = 94.89$, $p < 0.001$; $\eta_p^2 = 0.02$	2.07 (0.06)[a]	2.24 (0.04)[a]	1.29 (0.05)[c]	1.70 (0.02)[b]
Letter-sound knowledge (Set 2)	$F_{(3,10,470)} = 35.14$, $p < 0.001$; $\eta_p^2 = 0.01$	9.65 (0.21)[ab]	10.11 (0.12)[a]	8.04 (0.17)[c]	9.14 (0.08)[b]

Differing superscripts indicate significant differences at alpha = 0.05. η_p^2 effect sizes are 0.01 = small, 0.06 = medium, 0.14 = large.

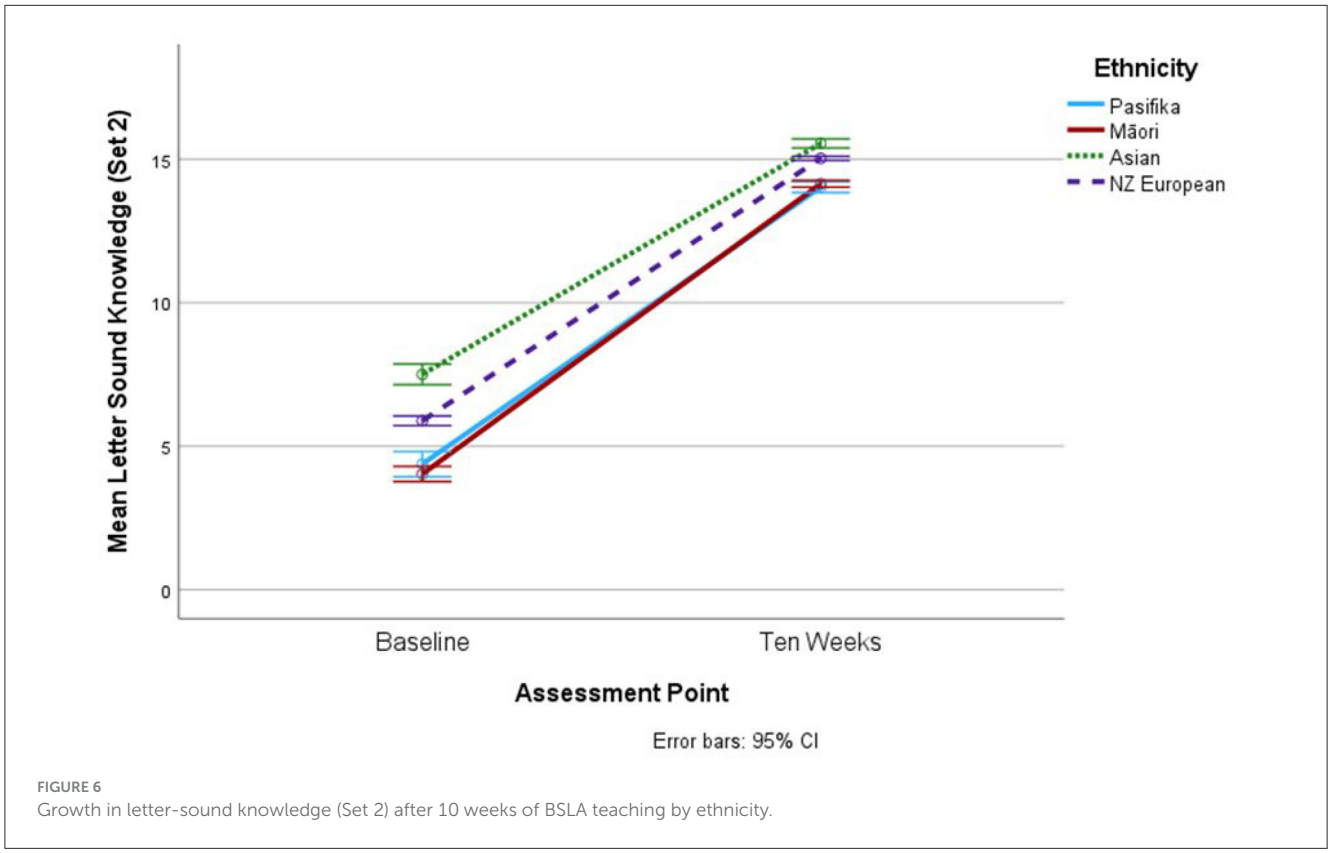

FIGURE 6
Growth in letter-sound knowledge (Set 2) after 10 weeks of BSLA teaching by ethnicity.

the Ten Week assessment point within this group of students. Table 2 provides descriptive statistics for initial phoneme identity, phoneme blending, and letter-sound knowledge for the school entry cohort at Baseline and Ten Weeks. Paired samples *t*-tests indicated significant change from Baseline to Ten Weeks on all tasks (all *t*'s > 85.62, all *p*'s < 0.001).

As seen in Table 2, shifts in both mean and median scores are evident after 10 weeks of BSLA teaching. As an illustration of this rapid growth Figures 7, 8 show the distribution of the number of children who reached proficiency levels (8, 9, or 10 items correct) for the initial phoneme identity task at Baseline and Ten Week assessment, respectively. Table 3 provides the proficiency rates for the school entry cohort at Baseline and at Ten Week assessment.

Non-word reading and non-word spelling

This analysis focused on the 2,313 students who completed the Thirty Week assessment. Only 12.3 and 1.7% of students completed the non-word reading and non-word spelling tasks at Baseline assessment, respectively. For any students who did not attempt these tasks at Baseline we have imputed a score of 0 at Baseline.

Repeated measures ANOVAs were used to examine growth over time in non-word reading and non-word spelling at the Baseline, Ten Week, and Thirty Week assessments. We examined the between-subjects factors of ELL status, gender, Māori ethnicity, Pacific Peoples ethnicity, and socioeconomic deprivation to determine whether growth differed based on these factors.

There was significant growth in non-word reading scores from Baseline to Thirty Weeks [$F_{(2,4,108)} = 426.48$, $p < 0.001$; $\eta_p^2 = 0.17$], with significant differences between scores at each of Baseline, Ten Weeks, and Thirty Weeks (p's < 0.001). There were no interactions with gender, ethnicity, socioeconomic deprivation, or ELL status, indicating similar rates of growth across demographics (all $\eta_p^2 < 0.009$).

There was also significant growth in non-word spelling scores from Baseline to Thirty Weeks [$F_{(2,3,828)} = 375.13$, $p < 0.001$; $\eta_p^2 = 0.16$], with significant differences between scores at each of Baseline, Ten Weeks, and Thirty Weeks (p's < 0.001). There were no interactions with gender, ethnicity, socioeconomic deprivation, or ELL status, indicating similar rates of growth across demographics (all $\eta_p^2 < 0.007$).

TABLE 2 Descriptive statistics for school entry cohort assessments.

	Assessment point	Mean (SD)	Median
Initial phoneme identity/10 items	Baseline	5.35 (2.93)	5
	10 weeks	7.94 (2.64)	9
Phoneme blending/10 items	Baseline	5.21 (2.69)	5
	10 weeks	7.48 (2.57)	8
Letter-sound knowledge (Set 1)/8 items	Baseline	4.32 (2.51)	4
	10 weeks	6.90 (1.75)	8
Letter-sound knowledge (Set 2)/17	Baseline	4.22 (5.98)	0
	10 weeks	14.43 (3.01)	15

BSLA Tier 2

There were 2,355 children who were identified as receiving 10 weeks (or 40 lessons) of supplementary BSLA Tier 2 and who were assessed at the Twenty Week assessment point. This equates to 14.3% of students assessed at Ten Weeks who received Tier 2. There was no difference in the percentage of English Language Learners in the Tier 2 (9.9% ELL) and non-Tier 2 students [10.8% ELL; $\chi^2(1) = 1.64$, $p = 0.20$], nor in the gender distribution of Tier 2 (52.7% male) and non-Tier 2 students [50.4% male; $\chi^2(1) = 3.65$, $p = 0.06$]. Tier 2 and non-Tier 2 students also didn't differ in their age (Cohen's $d = 0.09$) or socioeconomic deprivation (Cohen's $d = 0.12$).

We examined growth for the Tier 2 children across the Baseline, Ten Week, and Twenty Week assessment points using repeated measures ANOVAs (see Table 4). There was significant growth for Tier 2 children on initial phoneme identity, phoneme blending and both letter-sound tasks. Bonferroni *post-hoc* tests indicated significant differences between Baseline, Ten Weeks, and Twenty Weeks on all tasks (all p's < 0.001).

Growth in non-word reading and spelling in response to Tier 2

To determine the impact of Tier 2 teaching on students word decoding and encoding ability, we selected a matched control group of demographically similar students but who did not require Tier 2 teaching during their first year at school. There were 432 students with data available at the Thirty Week assessment who had received at least 10 weeks of Tier 2 teaching. Our control group was selected from the remaining 1,881 students with data at the Thirty Week assessment. Case control matching was used to select a matched non-Tier 2 student for each Tier 2 student based on gender, age in months (±1 month), and socioeconomic deprivation (plus or minus one decile), resulting in 432 matched controls. There were 189 females (43.8%) in both the Tier 2 and non-Tier 2 matched

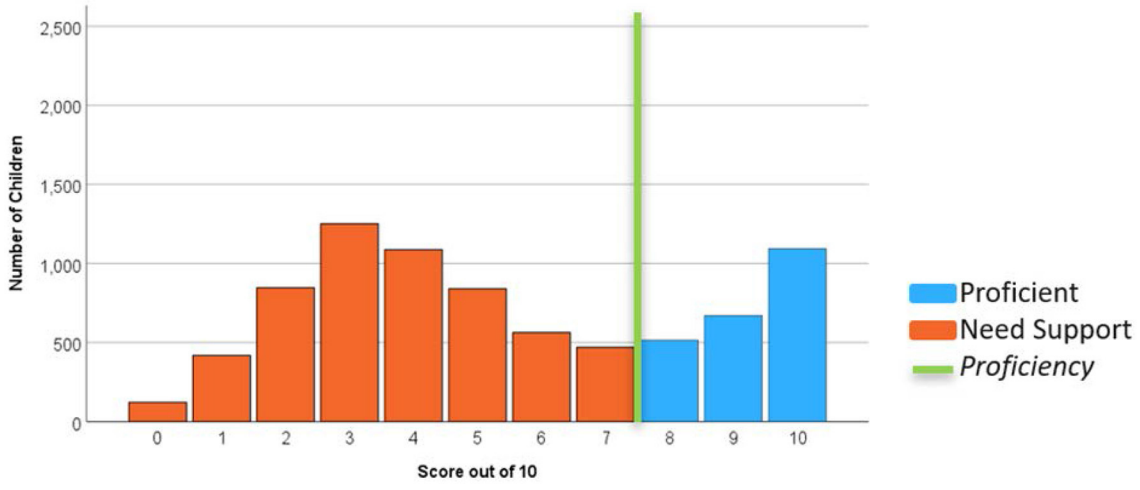

FIGURE 7
Initial phoneme identity score distribution at baseline: school entry cohort.

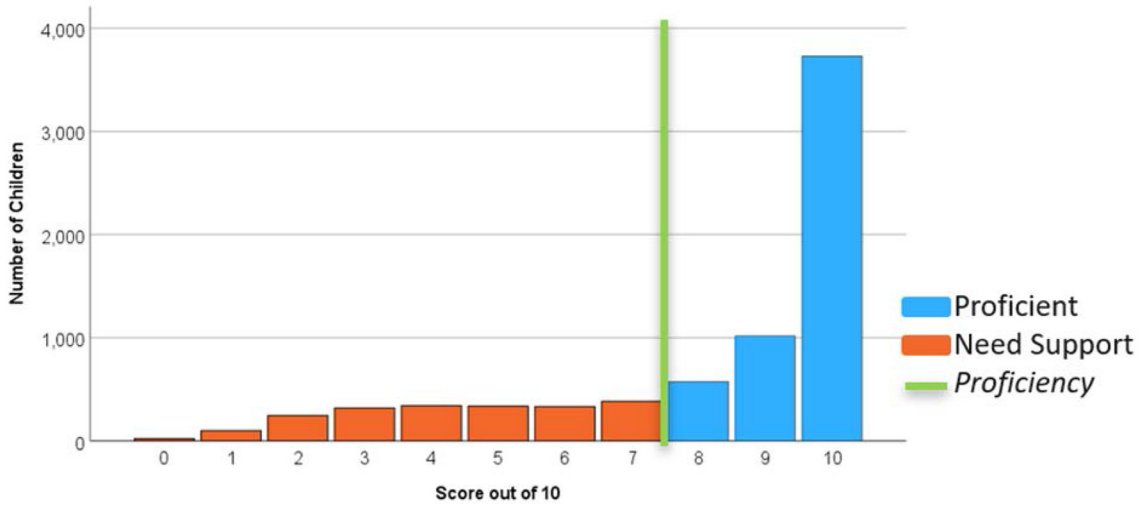

FIGURE 8
Initial phoneme identity score distribution at 10 weeks: school entry cohort.

TABLE 3 Proficiency rates for school entry cohort at baseline and 10 weeks.

	Percent proficient at baseline	Percent proficient at ten weeks
Initial phoneme identity	28.9%	72.4%
Phoneme blending	21.9%	62.8%
Phoneme-grapheme (Set 1)	36.4%	84.1%
Phoneme-grapheme (Set 2)	15.9%	68.2%

groups. Age in months did not differ significantly between the groups [$t_{(862)} = 0.39$, $p = 0.70$; $d = 0.03$]; however, the control group scored higher in socioeconomic deprivation than the Tier 2 group, although this was a very small effect [$t_{(862)} = 2.05$, $p = 0.04$; $d = 0.14$].

We examined growth for these two groups (ie children who received Tier 2 and children who did not require Tier 2) in both non-word reading and non-word spelling across 30 weeks of BSLA teaching.

Figures 9, 10 show the growth in non-word reading and non-word spelling, respectively, from Baseline to Thirty Week assessment. Note that children in Tier 2 completed an additional post-Tier 2 assessment (at Twenty Week assessment).

After 10 weeks of BSLA teaching there was a substantial difference between Tier 2 students and non-Tier 2 students, indicating that teachers had correctly identified students requiring further support after 10 weeks of teaching [non-word reading: $t_{(789)} = 5.03$, $p < 0.001$, $d = 0.36$; non-word spelling: $t_{(754)} = 6.14$, $p < 0.001$, $d = 0.44$]. At the 30 Week assessment, however, once Tier 2 students had completed Tier 2 teaching, there was no longer significant differences between these students and their matched controls on non-word reading [$t_{(816)} = 1.41$, $p = 0.16$, $d = 0.10$] or non-word spelling [$t_{(784)} = 1.81$, $p = 0.07$; $d = 0.13$], indicating Tier 2 students had caught up to their peers.

Tier 2 students' task response times

Mean response times were computed for each online assessment task (initial phoneme identity, phoneme blending, and letter-sound task) across task items. Change scores, reflecting the reduction in mean response time from Ten Week to Twenty Week assessments, were then computed.

We first compared data for the 864 Tier 2 students and matched non-Tier 2 controls at the Ten Week assessment point. There was a trend for Tier 2 students to take longer on all tasks compared to non Tier 2 matached controls at the Ten Week assessment (ie., pre-Tier 2 implementation). The response time differences were statistically significant ($p < 0.05$) for phoneme blending and letter-sound matching (Set 2) tasks (results are shown in Supplementary Table 3).

We next examined growth from Ten Weeks to Twenty Weeks (i.e., pre- to post-Tier 2) in the 2,355 students who had completed BSLA Tier 2 using repeated measures ANCOVAs, with change in response time as the covariate. For these analyses, only students who completed the task both pre- and post- Tier 2 are included. On all tasks other than phoneme blending there was a significant interaction between time and change in response time (see Table 5). In other words, the more students' response times reduced between the pre- and post-Tier 2 assessments, the greater their growth in scores over this time on initial phoneme identity and letter-sound task.

Oral language and listening comprehension

Students were assessed on their oral narrative ability and listening comprehension at the Baseline and Thirty Week

TABLE 4 Repeated measures analysis of growth for Tier 2 students.

	Estimated marginal means (std error)			Repeated measures ANOVA		Effect size	
	Baseline	10 weeks	20 weeks	*F*-statistic	*p*-value	η^2	Cohen's *d*[a]
Initial phoneme identity/10	5.22 (0.09)	7.75 (0.08)	8.91 (0.06)	1,131.92	<0.001	0.48	1.92
Phoneme blending/10	4.64 (0.07)	6.79 (0.08)	8.27 (0.07)	1,297.26	<0.001	0.52	2.08
Letter-sound Set 1/8	4.34 (0.7)	6.87 (0.5)	7.54 (0.05)	1,705.45	<0.001	0.54	2.17
Letter-sound Set 2/17	4.44 (0.23)	14.38 (0.13)	16.10 (0.07)	2,261.35	<0.001	0.76	3.56

[a]Eta-squared effect sizes (η^2) have been converted to Cohen's d using formulas available in Cohen (1988). Raw scores (number correct) are used in the analyses.

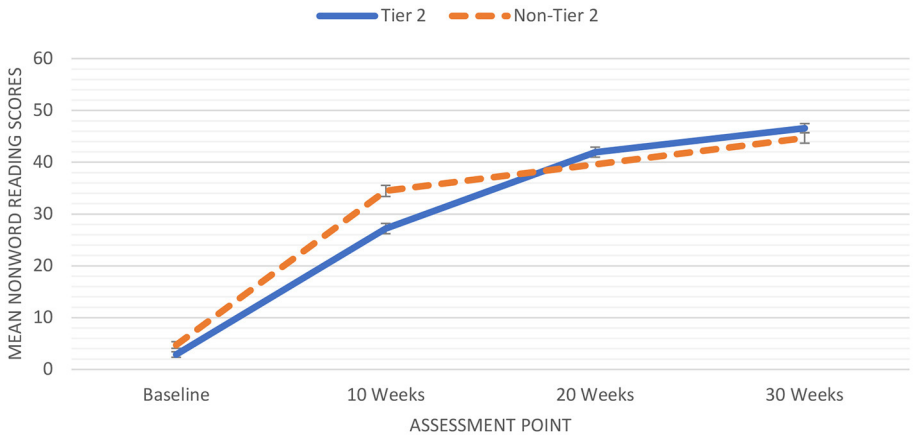

FIGURE 9
Non-word reading scores over 30 weeks of BSLA teaching for Tier 2 students and controls.

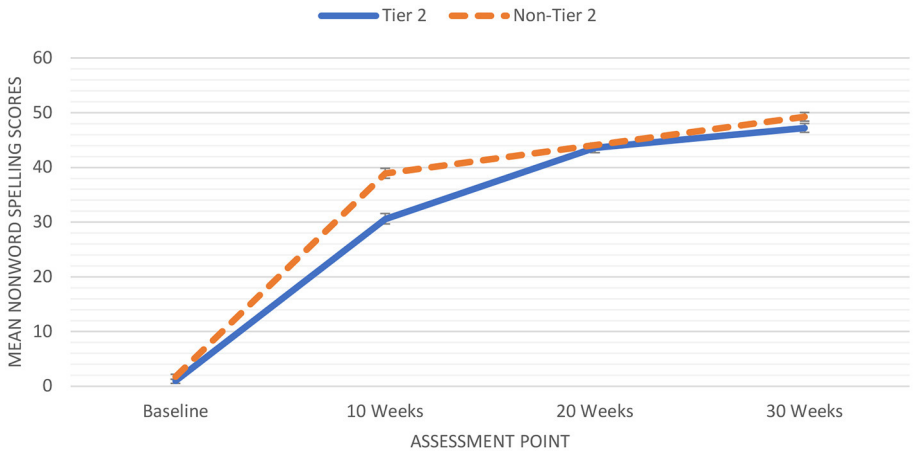

FIGURE 10
Non-word spelling scores over 30 weeks of BSLA teaching for Tier 2 students and controls.

assessment points. Repeated measures analyses of variance were used to examine growth over Thirty Week assessment of BSLA teaching in the 1,518 students assessed at Baseline and Thirty Week on the story retell task. In addition to examining within-subjects growth over time, we also examined the between-subjects factors of gender, ethnicity, and socioeconomic deprivation [coded as high deprivation (deciles 6 through 10) or low deprivation (deciles 1 through 5)] to determine whether growth differed based on these factors. We used an effect size threshold of partial eta-squared (η_p^2) >0.01 (equivalent to Cohen's $d = 0.2$) to identify significant effects.

Significant growth over 30 weeks of BSLA teaching was observed in listening comprehension [$F_{(1,1,399)} = 31.71$, $p < 0.001$; $\eta_p^2 = 0.02$], number of words [$F_{(1,1,336)} = 31.76$, $p < 0.001$; $\eta_p^2 = 0.02$], number of different words [$F_{(1,1,336)} = 27.09$, $p < 0.001$; $\eta_p^2 =$

TABLE 5 ANCOVA results predicting growth pre- to post-Tier 2.

Task	Effect	*F*-test
Initial phoneme identity	Time	$F_{(1,589)} = 523.96, p < 0.001, \eta_p^2 = 0.47$
	Time*Response time change	$F_{(1,589)} = 15.96, p < 0.001, \eta_p^2 = 0.03$
Phoneme blending	Time	$F_{(1,921)} = 881.37, p < 0.001, \eta_p^2 = 0.49$
	Time*Response time change	$F_{(1,921)} = 1.24, p = 0.27, \eta_p^2 = 0.001$
Letter-sound knowledge (Set 1)	Time	$F_{(1,609)} = 292.32, p < 0.001, \eta_p^2 = 0.32$
	Time*Response time change	$F_{(1,609)} = 61.47, p < 0.001, \eta_p^2 = 0.09$
Letter-sound knowledge (Set 2)	Time	$F_{(1,359)} = 312.32, p < 0.001, \eta_p^2 = 0.47$
	Time*Response time change	$F_{(1,359)} = 44.62, p < 0.001, \eta_p^2 = 0.11$

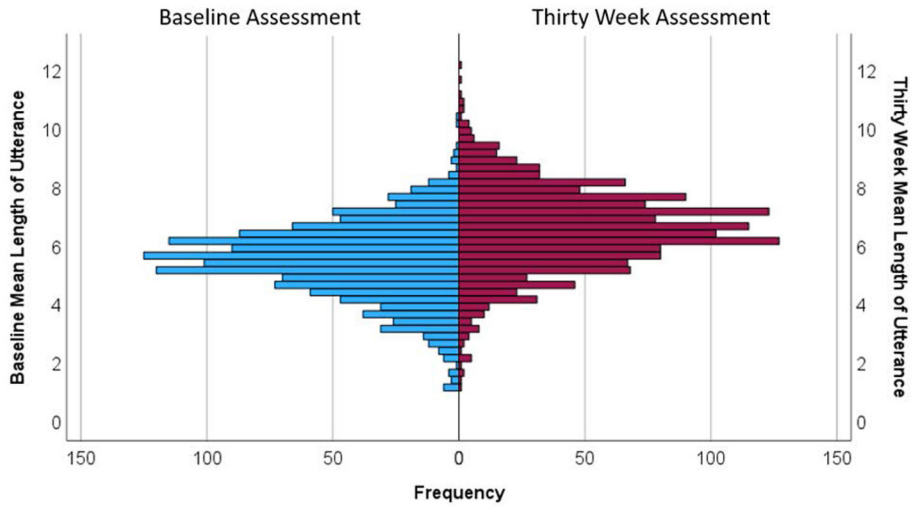

FIGURE 11
Distribution of mean length of utterance scores at Baseline and Thirty Week assessments.

0.02], and mean length of utterance [$F_{(1,1,305)} = 95.27, p < 0.001; \eta_p^2 = 0.07$]. While growth in intelligibility was statistically significant, it did not meet our effect size threshold for significant growth [$F_{(1,1,336)} = 9.59, p = 0.002; \eta_p^2 = 0.007$]. Thus, the largest growth effect was seen for mean length of utterance ($d = 0.54$), with more modest effects for listening comprehension, number of words, and number of different words (all $d = 0.29$). Figure 11 shows the shift in the distribution of mean length of utterance scores from Baseline to Thirty Weeks.

There were no significant interactions with any between-subjects factors (all $\eta_p^2 < 0.006$), indicating that patterns of growth in oral narrative abilities and listening comprehension were comparable across gender, ethnicity, and socioeconomic deprivation.

BSLA teaching fidelity

A total of 2,805 fidelity checklists were completed by BSLA teachers during the first 10 weeks of BSLA teaching. In terms of student engagement with BSLA lessons, teachers reported that overall, most students' engagement with the activities was high (65.2%), with a smaller percentage of lessons where engagement was rated as average/variable (32.3%) or low (2.5%). The majority of teachers (70%) reported spending 30 min or more in implementing the BSLA daily lesson plans with 26% indicating they spent between 15 and 25 min daily and 4% indicating they spent <15 min. See Table 4 in Supplementary material for the percentage of lessons that included each of the potential Tier 1 BSLA activities and small group reading instruction activities.

Overall, on average teachers included 6.4 (SD = 1.5) different activities in their daily 30 min large group lessons and 4.2 (SD = 1.1) activities in their daily small group instructional reading lessons.

Discussion

Large scale implementation of effective early literacy teaching approaches is critically important as we strive toward education

equality (United Nations, 2015) and address current global literacy challenges (UNESCO, 2019). Insights gained from how differing countries or regions engage in such implementation and the benefits realized from such initiatives, are useful to guide effective literacy teaching practices internationally.

In this study, we presented data from a large-scale implementation of an early literacy approach referred to as the Better Start Literacy Approach (BSLA) within the New Zealand education context. The BSLA was specifically designed for large scale implementation. Its pathway from initial design to large scale implementation included several features that are consistent with recommendations from other research groups (Hoover and Soltero-González, 2018; Solari et al., 2020). First, there was strong investment in the design and development phase (over a 4-year period). The approach was based on the science of reading and braided together important research findings related to cognitive, ecological (including cultural), and psychological factors that influence children's literacy development. Its development involved consultative processes between a multidisciplinary team of researchers, practitioners, Māori and Pacific leaders, school leaders and policy makers. The implementation plan was co-constructed with class teachers and evolved over time in response to teacher feedback. Controlled research trials were implemented to establish the approach was more effective than teachers' existing early literacy teaching practices. Significant attention was given to the development of online PLD using a micro-credential professional learning framework to support teachers across the country to implement the new approach. Novel online assessments for reliable and efficient ways for teachers to monitor student growth and guide teaching content relevant to local cultural contexts were developed.

Investment in this development phase allowed for the rapid large scale implementation of BSLA across the country. Funded by The New Zealand Ministry of Education, within a 2.3 year time period (February 2000–May 2023) over 3,000 teachers from 819 schools were involved in implementing BSLA in junior school classrooms. The online model of PLD proved particularly useful as this implementation phase occurred during periods (e.g., 2–3 month periods) of school closures due to COVID-19 pandemic disruptions.

Findings from analyses using data that teachers routinely collected online as part of BSLA are very encouraging. After the teachers had implemented just 10 weeks (or 40 Tier 1 daily lesson plans) 5-year-old children (n = 10,532) showed a higher rate of proficiency on phonic and phoneme awareness measures than students of the same age who had not yet engaged in BSLA teaching (n = 15,380). Effect sizes of the difference between the BSLA group (n=10,532 children) and the no-BSLA group (N = at each month of age between 64 and 70 months on each task indicated medium to large effects sizes favoring BSLA teaching. Children commencing school at 5 y 0 m−5 y 3 m and received BSLA from the outset of their literacy teaching showed remarkable growth in letter-sound knowledge, phoneme identity and phoneme blending skills following 10 weeks of BSLA. The majority of these learners reached proficiency on these tasks at the Ten Week monitoring assessment.

These data add to the evidence base of effective Tier 1 early literacy teaching strategies within MTSS (e.g., Fuchs and Deshler, 2007; Al Otaiba et al., 2019; Petscher et al., 2020). They suggest that when teachers are well supported to provide explicit and systematic instruction in phonics and phoneme awareness as part of Tier 1 universal teaching, most 5-year-old children can master these skills. Given the robust evidence demonstrating proficiency in phoneme awareness is a strong predictor of early reading and spelling success (see Gillon, 2018 for a review) developing these skills early is important. The speed of learning for the school entry cohort (7,876 children aged 5 y 0 m−5 y 3 m), observed in the current study, is particularly impressive. Key teaching aspects within BSLA such as, focusing phonological awareness activities at the phoneme level, integrating letter-sound knowledge with phoneme awareness activities, and the pace of teaching through the BSLA scope and sequence, may have contributed to this rapid acquisition of phonic and phoneme awareness skills.

Teachers selected 14.3% of students after 10 weeks of Tier 1 teaching for additional Tier 2 support. These learners showed significant growth in phonic and phoneme awareness skills in response to 10 weeks of BSLA Tier 2 small group lesson plans. Their task response times were also improving, suggesting that on most tasks they were becoming both more accurate and more efficient in processing phonological information. By 30 weeks there was evidence that children with greater learning needs who received BSLA Tier 2 teaching had caught up to their peers in word reading and spelling skills. This is a rare finding outside of tightly controlled experimental studies.

These data provide further support to the value of MTSS for early literacy teaching. Within BSLA, Tier 1 and Tier 2 teaching activities are aligned. Such alignment provides increased intensity of teaching for those children with greater learning needs. It helps to reduce cognitive load and build learners' confidence through familiarity of tasks and game activities and provides additional time in small group reading instruction. A further design feature of the BSLA implementation was significant Government investment to engage differing professional groups (e.g., class teachers, speech-language therapists, literacy specialists), in the same high quality Professional Learning and Development (PLD) (schools self-funded their teacher aides to enroll in the BSLA Teacher Aide micro-credential). This was intentional to build "professional learning communities" (Vescio et al., 2008) in localized areas where a common practice model in supporting children's early literacy learning can evolve. It also helped ensure that professionals involved in supporting children in differing tiers of support within MTSS were familiar with the same monitoring assessment measures, data analyses, teaching content and evidenced based teaching strategies. This provides consistency in practice for children and their families and maximizes learning opportunities for children.

An important aspect to the BSLA was explicit teaching of skills that supported children's oral language comprehension and vocabulary growth as well as their word recognition skills (as aligned to the Simple View of Reading). Data from a novel online oral narrative assessment (Gillon et al., 2023b) indicated significant growth in children's comprehension of both factual and inferential information when listening to a short story as well as growth

in expressive vocabulary measures when retelling a story. The successful large-scale implementation of this online oral narrative task by class teachers as part of their regular teaching practice is notable. Evidence that teachers are systematically monitoring children's oral narrative, listening comprehension and expressive vocabulary in valid and reliable ways (in addition to phonic and phoneme awareness progress monitoring) is rare in classroom practice. Teachers' use of valid and reliable online assessments that capitalize on latest digital technologies, automated speech transcription and automated recording and reporting features (as used in BSLA monitoring assessments) hold much promise for large scale implementation. Such features enable teachers to monitor each child's growth in foundational literacy skills in efficient ways that can inform their next steps for teaching. When assessments are designed and used in strengths-based ways to celebrate children's learning and are aligned to teaching activities, they may also support the sustainability of enhanced or new teaching practices.

Teacher self-reported fidelity measures indicated teaching fidelity was high during their first 10 weeks of BSLA implementation. Given the supports teachers were provided (weekly Zoom sessions, lesson plans, online video lesson demonstrations, teaching resources and a literacy specialist to coach or support them through these weeks) this is not a surprising finding. Longer term data and independent observation of teachers' implementation is important in future studies to further investigate teaching fidelity and the sustainability of their enhanced literacy teaching. However, the data of students' progress over a 30 week teaching period suggests teachers were implementing BSLA in ways that significantly enhanced children's foundational literacy skills. This finding contrasts with some previous studies of large-scale literacy PLD. For example, Piasta et al. (2017) through a well-designed study did not find any evidence that literacy PLD delivered state-wide in the USA early education context enhanced children's early literacy achievement. The model involved a series of 2-day face to face workshops with a minimum of 14 h PLD. They did find more positive results when teachers received monthly in-class coaching (vs. no-coaching), but overall effects were minimal. The authors raised important considerations around PLD quality measurement challenges when implementing at large scale and the need for PLD to be embedded within curricula. Within BSLA, teachers have continuous access to the online PLD content and video demonstrations of teaching activities prepared and presented by the BSLA research team and experienced teachers. They also have access to weekly online live Zoom sessions with the research team (even after they have completed their BSLA micro-credential). This ongoing access to high quality PLD may support more positive student learning outcomes. Continued research into the facilitators of ensuring enhanced teacher knowledge and practice leads to significant improvements in learners' literacy achievement is necessary.

The call to action from the WHO to direct resources and investment in developing comprehensive literacy skills in our diverse communities, where education inequities are greater, requires new approaches. The integration of both a culturally responsive and strengths-based approach to early literacy teaching within a MTSS framework are an important feature of BSLA. These features specifically responded to this call for action as well as responded to the voices of Māori and Pacific leaders in New Zealand. The BSLA development integrated research findings highlighting the importance of cultural relevance, cultural appropriateness, valuing and engaging families, strengths-based language and reporting to facilitating success for indigenous learners and children with greater learning needs. Children who identified as Māori or Pacific Peoples group showed accelerated growth following 10 weeks of BSLA teaching on some measures, catching up to their Asian and NZ European peers. Following 30 weeks of BSLA teaching, positive change was reported across ethnic groups, across areas of different socio economic deprivations, for both males and females and for English Language Learners. These data suggest that the early accelerated progress for Māori and Pacific learners was followed by sustained growth in similar positive ways as their peers. These findings of large-scale positive impact for 5-year-old children who identify as Māori and Pacific Peoples across New Zealand are unique. These findings, however, are not surprising given the BSLA research team specifically drew upon research evidence and knowledge from other Māori and Pacific scholars that identified aspects of teaching and learning environments that facilitate learning success for Māori and for Pacific learners.

These positive findings for Māori young learners in New Zealand have international significance for other indigenous and culturally diverse communities. In particular, they highlight the critical importance of consulting and involving indigenous leaders in large scale implementation design and attending to findings from strengths-based research led by indigenous researchers. Our global aim is to decrease current literacy inequities. We must consider how new literacy approaches introduced at scale are meeting the needs of all learners. We should ensure teaching resources, game activities, stories, decodable reading series, vocabulary items used in activities and family engagement strategies selected to build children's oral language and early reading and writing skills reflect local cultural contexts. Approaches that help facilitate cultural inclusiveness and culturally responsive teaching practices within large scale early literacy implementation projects may ensure we are not simply advancing foundational skills in a country's dominant language and culture, Rather, we need to ensure we are also meeting the early literacy needs for our indigenous learners and learners from culturally and linguistically diverse communities. This is an important area for more in-depth and ongoing research.

A further consideration for ongoing research is evaluating the wider societal benefits for a region from a large-scale implementation of effective early literacy approach within the region. A preliminary attempt at such an evaluation is reported through an independent analysis of the potential social value return on investment for the national implementation of BSLA within New Zealand (ImpactLab, 2021). The independent group (Impactlab) who had access to our BLSA data from controlled research trials used a methodology of estimating the known longer term positive outcomes from early literacy achievement as well as estimated reduction in adverse behaviors associated with reading failure. This evidence was used alongside the investment made in BSLA implementation and results shown within the

analysis. This analysis suggested that for every dollar invested in BSLA the social value return on investment for New Zealand was $38.20. Such analyses are important to support further Government investment and the sustainability of new literacy approaches introduced.

Summary

This study has highlighted features of a large-scale implementation of an evidenced based literacy approach (BSLA) within the New Zealand education context. Data demonstrating a strong positive response from thousands of 5-year-old children across ethnicities, socio-economic areas and regional areas is very encouraging. The data suggest that when class teachers are well-supported through quality PLD, they can rapidly develop the foundational literacy skills that are critical to reading and writing success during children's first year at school. Significant government investment in both design and implementation phases is beneficial and is potentially a strong social value investment for a country. The longer-term advantages from such investment in early literacy achievement for these 5-year-old learners in New Zealand now awaits detailed analysis. Understanding pathways and strategies for large scale implementation of effective literacy approaches that differing countries and regions adopt is a continued worthwhile pursuit to address our collective global literacy aspirations.

Data availability statement

The datasets presented in this article are not readily available because ethical approval received currently allows for the BSLA research team to have direct access to the data collected. Further inquiries can be directed to the author. Requests to access the datasets should be directed to: gail.gillon@canterbury.ac.nz.

Ethics statement

This study was approved by the University of Canterbury Human Ethics Committee. This study was conducted in accordance with the local legislation and institutional requirements. The data collection method and processes reported were approved by the University of Canterbury Human Ethics Committee (application 2021/06/LRPS). The approval allowed the researchers to analyze de-identified assessment data routinely collected within the national implementation of BSLA for secondary analyses to evaluate the impact of the Better Start Literacy Approach. The committee determined this was low risk and individual consent for participation was not required. However, as children entered BSLA, parents were provided with an option to exclude their child's anonymized data from the database used for any secondary analysis.

Author contributions

GG: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. BM: Conceptualization, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. AS: Conceptualization, Investigation, Project administration, Writing – review & editing. MG: Data curation, Formal Analysis, Writing – original draft, Methodology, Writing – review & editing. AM: Conceptualization, Funding acquisition, Writing – review & editing. TT: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Ministry of Education (Contract No: MOE 08814) and the Better Start National Science Challenge (Contract No: 3710593).

Acknowledgments

The authors are very grateful to the teachers, literacy specialists, families, and children who participated in this project. They also acknowledge the work of the wider Better Start Literacy Approach team in their implementation of this project. The authors would like to acknowledge Ministry of Business, Innovation and Employment, The Better Start National Science Challenge, and the Ministry of Education for funding support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1354182/full#supplementary-material

References

Aaron, P. G., Malatesha Joshi, R., Gooden, R., and Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: an alternative to the discrepancy model of LD. *J. Learn. Disab.* 41, 67–84. doi: 10.1177/0022219407310838

Adlof, S. M., and Hogan, T. P. (2019). If we don't look, we won't see: measuring language development to inform literacy instruction. *Policy Insights Behav. Brain Sci.* 6, 210–217. doi: 10.1177/2372732219839075

Al Otaiba, S., Baker, K., Lan, P., Allor, J., Rivas, B., Yovanoff, P., et al. (2019). Elementary teacher's knowledge of response to intervention implementation: a preliminary factor analysis. *Annal. Dyslexia* 69, 34–53. doi: 10.1007/s11881-018-00171-5

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., et al. (2014). To wait in tier 1 or intervene immediately: a randomized experiment examining first-grade response to intervention in reading. *Excep. Child.* 81, 11–27. doi: 10.1177/0014402914532234

Alansari, M., Webber, M., Overbye, S., Tuifagalele, R., and Edge, K. (2022). *Conceptualising Māori and Pasifika Aspirations and Striving for Success (COMPASS)*. Wellington: New Zealand Council for Educational Research.

Arrow, A., Gillon, G., McNeill, B., and Scott, A. (2021). *Ready to Read Phonics Plus Project*. Available online at: https://www.betterstartapproach.com/our-research

Arrow, A., Neville, A., Denston, A., and Nicholson, T. (2022). Investigating the number and type of literacy assessments and interventions in Aotearoa New Zealand primary schools. *Austr. J. Learn. Diff.* 27, 185–199. doi: 10.1080/19404158.2022.2064522

Atkinson, J., Salmond, C., and Crampton, P. (2019). *NZDep2018 Index of Deprivation. Interim Research Report, December 2019*. Wellington: Department of Public Health, University of Otago.

Averill, O. H., Rinaldi, C., and Urban Special Education Leadership Collaborative (2011). Multi-tier system of supports (MTSS). *District Admin.* 48, 91–95.

Berryman, M., and Eley, E. (2017). Succeeding as Māori: Māori students' views on our stepping up to the Ka Hikitia challenge. *N. Z. J. Educ. Stu.* 52, 93–107. doi: 10.1007/s40841-017-0076-1

Bright, R., Ashton, E., Mckean, C., and Wren, Y. (2023). The development of a digital story-retell elicitation and analysis tool through citizen science data collection, software development and machine learning. *Front. Psychol.* 14:989499. doi: 10.3389/fpsyg.2023.989499

Carson, K., Boustead, T., and Gillon, G. (2015). Content validity to support the use of a computer-based phonological awareness screening and monitoring assessment (Com-PASMA) in the classroom. *Int. J. Speech-Language Pathol.* 17, 500–510. doi: 10.3109/17549507.2015.1016107

Catts, H. W., and Hogan, T. P. (2020). Dyslexia: an ounce of prevention is better than a pound of diagnosis and treatment. *Read. League J.* 2, 6–13.

Chapman, J. W., and Tunmer, W. E. (2016). Is reading recovery an effective intervention for students with reading difficulties? A critique of the i3 scale-up study. *Read. Psychol.* 37, 1025–1042. doi: 10.1080/02702711.2016.1157538

Clay, M. (1990). The reading recovery program, 1984-88-coverage, outcomes and education board district figures. *N. Z. J. Educ. Stu.* 25, 61–70.

Clendon, S., Scott, A., Gillon, G., and McNeill, B. (2022). *Assessment for Learning: Adapting Phoneme Awareness Assessment for Children With Complex Communication Needs*. American Speech-Language-Hearing Association Convention New Orleans, USA.

Cunningham, A. E., and Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Dev. Psychol.* 33:934. doi: 10.1037/0012-1649.33.6.934

D'Agostino, J. V., Lose, M. K., and Kelly, R. H. (2017). Examining the sustained effects of reading recovery. *J. Educ. Stud. Placed Risk* 22, 116–127. doi: 10.1080/10824669.2017.1286591

D'Agostino, J. V., and Rodgers, E. (2017). Literacy achievement trends at entry to first grade. *Educ. Res.* 46, 78–89. doi: 10.3102/0013189X17697274

DeMonte, J. (2017). *Micro-Credentials for Teachers*. New York, NY: American Institutes for Research.

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K. E., Michaels, R., Shaywitz, S. E., et al. (2015). Achievement gap in reading is present as early as first grade and persists through adolescence. *J. Pediatr.* 167, 1121–1125. e1122. doi: 10.1016/j.jpeds.2015.07.045

Fien, H., Chard, D. J., and Baker, S. K. (2021). Can the evidence revolution and multi-tiered systems of support improve education equity and reading achievement? *Read. Res. Q.* 56, S105–S118. doi: 10.1002/rrq.391

Fox, C., Jones, S., Gillam, S. L., Israelsen-Augenstein, M., Schwartz, S., Gillam, R. B., et al. (2022). Automated progress-monitoring for literate language use in narrative assessment (LLUNA). *Front. Psychol.* 13:894478. doi: 10.3389/fpsyg.2022.894478

Freudenthal, D. R., Zaru, M. W., and Al Otaiba, S. (2023). "Early literacy, response to intervention, and multi-tiered systems of support," in *Handbook on the Science of Early Literacy,* eds. S. Cabell, S. Neuman, and N. Terry (New York: Guilford Press), 43.

Fuchs, D., and Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learn. Disab. Res. Prac.* 22, 129–136. doi: 10.1111/j.1540-5826.2007.00237.x

Fuchs, D., and Fuchs, L. S. (2017). Critique of the national evaluation of response to intervention: a case for simpler frameworks. *Except. Child.* 83, 255–268. doi: 10.1177/0014402917693580

Gillam, S. L., Vaughn, S., Roberts, G., Capin, P., Fall, A. M., Israelsen-Augenstein, M., et al. (2023). Improving oral and written narration and reading comprehension of children at-risk for language and literacy difficulties: results of a randomized clinical trial. *J. Educ. Psychol.* 115:99. doi: 10.1037/edu0000766

Gillon, G. (2018). *Phonological Awareness: From Research to Practice Second Edition*. Guilford Press.

Gillon, G. (2023). Supporting children who are english language learners succeed in their early literacy development. *Folia Phoniatrica et Logopaedica* 75, 219–234. doi: 10.1159/000531407

Gillon, G., and Macfarlane, A. H. (2017). A culturally responsive framework for enhancing phonological awareness development in children with speech and language impairment. *Speech Lang. Hear.* 20, 163–173. doi: 10.1080/2050571X.2016.1265738

Gillon, G., McNeill, B., Denston, A., Scott, A., and Macfarlane, A. (2020). Evidence-based class literacy instruction for children with speech and language difficulties. *Top. Lang. Disord.* 40, 357–374. doi: 10.1097/TLD.0000000000000233

Gillon, G., McNeill, B., Scott, A., Arrow, A., Gath, M., and Macfarlane, A. (2023a). A better start literacy approach: effectiveness of Tier 1 and Tier 2 support within a response to teaching framework. *Read. Writ.* 36, 565–598. doi: 10.1007/s11145-022-10303-4

Gillon, G., McNeill, B., Scott, A., Denston, A., Wilson, L., Carson, K., et al. (2019). A better start to literacy learning: findings from a teacher-implemented intervention in children's first year at school. *Read. Writ.* 32, 1989–2012. doi: 10.1007/s11145-018-9933-7

Gillon, G., McNeill, B., Scott, A., Gath, M., and Westerveld, M. (2023b). Retelling stories: the validity of an online oral narrative task. *Child Lang. Teach. Ther.* 39, 150–174. doi: 10.1177/02656590231155861

Gillon, G., Smith, J., Maitland, R., Macfarlane, S., and Macfarlane, A. (2024). "He awa Whiria: informing the design of the Better Start Literacy Approach," in *He Awa Whiria: Braiding the Knowledge Streams in Research, Policy, and Practice*, eds. A. Macfarlane, M. Derby, and S. Macfarlane (Christchurch, NZ: Canterbury University Press), 145–168.

Gillon, G. T. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Lang. Speech Hear. Serv. Sch.* 31, 126–141. doi: 10.1044/0161-1461.3102.126

Gonzalez, J. E., Durán, L., Linan-Thompson, S., and Jimerson, S. R. (2022). Unlocking the promise of multitiered systems of support (MTSS) for linguistically diverse students: advancing science, practice, and equity. *Sch. Psychol. Rev.* 51, 387–391. doi: 10.1080/2372966X.2022.2105612

Hanushek, E. A., Light, J. D., Peterson, P. E., Talpey, L. M., and Woessmann, L. (2022). Long-run trends in the U.S. SES—achievement gap. *Educ. Finance Policy* 14, 1–33. doi: 10.1162/edfp_a_00383

Hatcher, P. J., Hulme, C., and Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: the phonological linkage hypothesis. *Child Dev.* 65, 41–57. doi: 10.2307/1131364

Hoover, J. J., and Soltero-González, L. (2018). Educator preparation for developing culturally and linguistically responsive MTSS in rural community elementary schools. *Teach. Educ. Special Educ.* 41, 188–202. doi: 10.1177/0888406417753689

Hoover, W. A., and Tunmer, W. E. (2018). The simple view of reading: three assessments of its adequacy. *Remedial Special Educ.* 39, 304–312. doi: 10.1177/0741932518773154

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., Stuart, G., et al. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *J. Exp. Child Psychol.* 82, 2–28. doi: 10.1006/jecp.2002.2670

ImpactLab (2021). *Better Start Literacy Approach: National Implementation ImpactLab Good Measure Report*. Available online at: https://www.betterstartapproach.com/our-research

Israelsen-Augenstein, M., Fox, C., Gillam, S. L., Holbrook, S., and Gillam, R. (2022). Monitoring indicators of scholarly language: a progress monitoring tool for documenting changes in narrative complexity over time. *Front. Educ.* 22:918127. doi: 10.3389/feduc.2022.918127

Juel, C. (1988). Learning to read and write: a longitudinal study of 54 children from first through fourth grades. *J. Educ. Psychol.* 80:437. doi: 10.1037/0022-0663.80.4.437

Le Fevre, D. M. (2014). Barriers to implementing pedagogical change: the role of teachers' perceptions of risk. *Teach. Teach. Educ.* 38, 56–64. doi: 10.1016/j.tate.2013.11.007

Lepola, J., Lynch, J., Kiuru, N., Laakkonen, E., and Niemi, P. (2016). Early oral language comprehension, task orientation, and foundational reading skills as predictors of grade 3 reading comprehension. *Read. Res. Q.* 51, 373–390. doi: 10.1002/rrq.145

Macfarlane, A., and Macfarlane, S. (2019). Listen to culture: Māori scholars' plea to researchers. *J. R. Soc. N. Z.* 49, 48–57. doi: 10.1080/03036758.2019.1661855

Macfarlane, S., Clarke, T. H., and Macfarlane, A. H. (2016). "Language, literacy, identity and culture: challenges and responses for indigenous Māori learners," in *Multilingualism, Literacy and Dyslexia: Breaking Down Barriers for Educators, 2nd Edn.*, eds. L. Peer and G. Reid (London: Routledge), 74–85.

Maessen, S. E., Taylor, B. J., Gillon, G., Moewaka Barnes, H., Firestone, R., Taylor, R. W., et al. (2023). A better start national science challenge: supporting the future wellbeing of our tamariki E tipu, e rea, mo ngā rā o to ao: grow tender shoot for the days destined for you. *J. R. Soc. N. Z.* 53, 673–696. doi: 10.1080/03036758.2023.2173257

May, H., Blakeney, A., Shrestha, P., Mazal, M., and Kennedy, N. (2023). Long-term impacts of reading recovery through 3rd and 4th grade: a regression discontinuity study. *J. Res. Educ. Eff.* 23, 1–26. doi: 10.1080/19345747.2023.2209092

May, S., Jang-Jones, A., and McGregor, A. (2019). *PISA2018 New Zealand Summary report: System Performance and Equity*. Wellington: Ministry of Education.

McNeill, B. C., Gillon, G., and Gath, M. (2023). The relationship between early spelling and decoding. *Lang. Speech Hear. Serv. Sch.* 54, 981–995. doi: 10.1044/2023_LSHSS-22-00161

Miller, J., and Chapman, R. (2004). *Systematic Analysis of Language Transcripts (SALT, v8, 0.). Computer Software.* Madison, WI: Language Analysis Laboratory, Waisman Center, University of Wisconsin-Madison.

Ministry of Education (2007). *The New Zealand Curriculum. Ministry of Education*. Available online at: https://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/English

Ministry of Education (2017). *Te Whāriki Early Childhood Curriculum. Ministry of Education*. Available online at: https://tewhariki.s3.ap-southeast-2.amazonaws.com/public/Key-documents/Files/Te-Whariki-Early-Childhood-Curriculum.pdf#page=65

Ministry of Education (2022). *Literacy and Communication and Maths Strategy. Ministry of Education*. Available online at: https://assets.education.govt.nz/public/Documents/Ministry/Changes-in-education/ELS-0778-Maths-and-Literacy-Strategies-Doc_web.pdf

Mullis, I. V. S., Martin, M. O., Foy, P., and Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Boston, MA: Boston College, TIMSS and PIRLS International Study Center.

Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., Wry, E., et al. (2023). *PIRLS 2021 International Results in Reading*. Boston, MA: Boston College, TIMSS and PIRLS International Study Center website.

Nancollis, A., Lawrie, B. A., and Dodd, B. (2005). Phonological awareness intervention and the acquisition of literacy skills in children from deprived social backgrounds. *Lang. Speech Hear. Serv. Sch.* 36, 325–335. doi: 10.1044/0161-1461(2005/032)

Nicholas, K., and Parkhill, F. (2014). Is reading recovery sustainable two to four years after discontinuation? *Education* 42, 663–672. doi: 10.1080/03004279.2012.759605

Peng, P., Fuchs, D., Fuchs, L. S., Elleman, A. M., Kearns, D. M., Gilbert, J. K., et al. (2019). A longitudinal analysis of the trajectories and predictors of word reading and reading comprehension development among at-risk readers. *J. Learn. Disab.* 52, 195–208. doi: 10.1177/0022219418809080

Petersen, D. B., Staskowski, M., Spencer, T. D., Foster, M. E., and Brough, M. P. (2022). The effects of a multitiered system of language support on kindergarten oral and written language: a large-scale randomized controlled trial. *Lang. Speech Hearing Serv. Sch.* 53, 44–68. doi: 10.1044/2021_LSHSS-20-00162

Petersen, D. B., and Stoddard, A. (2018). Psychometric requirements of oral and written language progress monitoring assessments. *Persp. ASHA Special Int. Groups* 3, 180–197. doi: 10.1044/persp3.SIG1.180

Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., et al. (2020). How the science of reading informs 21st-century education. *Read. Res. Q.* 55, S267–S282. doi: 10.1002/rrq.352

Piasta, S. B., Justice, L. M., O'Connell, A. A., Mauck, S. A., Weber-Mayrer, M., Schachter, R. E., et al. (2017). Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *J. Res. Educ. Effect.* 10, 354–378. doi: 10.1080/19345747.2016.1270378

Piper, B., Destefano, J., Kinyanjui, E. M., and Ong'ele, S. (2018). Scaling up successfully: lessons from Kenya's Tusome national literacy program. *J. Educ. Change* 19, 293–321. doi: 10.1007/s10833-018-9325-4

Pullen, P. C., van Dijk, W., Gonsalves, V. E., Lane, H. B., and Ashworth, K. E. (2018). "RTI and MTSS: response to intervention and multi-tiered systems of support: how do they differ and how are they the same, If at All?," in *Handbook of Response to Intervention and Multi-Tiered Systems of Support*, eds. P. C. Pullen, and M. Kennedy (London: Routledge), 5–10.

Ramacciotti, M. C., Sousa, H., Silveira, H. G., Hulme, C., Snowling, M. J., Newbury, D. F., et al. (2023). Scaling up early language intervention in educational settings: first steps matter. *Oxf. Rev. Educ.* 49, 29–47. doi: 10.1080/03054985.2022.2088488

Ratima, M., Smith, J., Macfarlane, A., and Macfarlane, S. (2020). *The Hikairo Schema for Primary: Culturally Responsive Teaching and Learning*. NZCER.

Relyea, J. E., Rich, P., Kim, J. S., and Gilbert, J. B. (2023). The COVID-19 impact on reading achievement growth of Grade 3–5 students in a U.S. urban school district: variation across student characteristics and instructional modalities. *Read. Writing* 36, 317–346. doi: 10.1007/s11145-022-10387-y

Ritchie, S. J., and Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* 24, 1301–1308. doi: 10.1177/0956797612466268

Schulte, A. C., Stevens, J. J., Elliott, S. N., Tindal, G., and Nese, J. F. (2016). Achievement gaps for students with disabilities: stable, widening, or narrowing on a state-wide reading comprehension test? *J. Educ. Psychol.* 108:925. doi: 10.1037/edu0000107

Scott, A., Gillon, G., McNeill, B., and Kopach, A. (2022). The evolution of an innovative online task to monitor children's oral narrative development. *Front. Psychol.* 13:903124. doi: 10.3389/fpsyg.2022.903124

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition* 55, 151–218. doi: 10.1016/0010-0277(94)00645-2

Solari, E. J., Terry, N. P., Gaab, N., Hogan, T. P., Nelson, N. J., Pentimonti, J. M., et al. (2020). Translational science: a road map for the science of reading. *Read. Res. Q.* 55, S347–S360. doi: 10.1002/rrq.357

Sparks, R. L., Patton, J., and Murdoch, A. (2014). Early reading success and its relationship to reading achievement and reading volume: replication of '10 years later'. *Read. Writing* 27, 189–211. doi: 10.1007/s11145-013-9439-2

Stanovich, K. E. (2009). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *J. Educ.* 189, 23–55. doi: 10.1177/0022057409189001-204

Stone, A., Walter, M., and Peacock, H. (2017). Educational outcomes for aboriginal school students in Tasmania: Is the achievement gap closing? *Aust. Int. J. Rural Educ.* 27, 90–110. doi: 10.47381/aijre.v27i3.148

Su, M., Peyre, H., Song, S., McBride, C., Tardif, T., Li, H., et al. (2017). The influence of early linguistic skills and family factors on literacy acquisition in Chinese children: follow-up from age 3 to age 11. *Learn. Inst.* 49, 54–63. doi: 10.1016/j.learninstruc.2016.12.003

Tunmer, W. E., Chapman, J. W., Greaney, K. T., Prochnow, J. E., and Arrow, A. W. (2013). Why the New Zealand national literacy strategy has failed and what can be done about it: evidence from the progress in international reading literacy study (PIRLS) 2011 and reading recovery monitoring reports. *Aust. J. Learn. Diff.* 18, 139–180. doi: 10.1080/19404158.2013.842134

Tunmer, W. E., Chapman, J. W., and Prochnow, J. E. (2006). Literate cultural capital at school entry predicts later reading. *N. Z. J. Educ. Stu.* 41:183.

UNESCO (2019). *Meeting Commitments: Are Counties on Track to Meet SDG 4 Global Education Monitoring Report*. Available online at: https://uis.unesco.org/sites/default/files/documents/meeting-commitments-are-countries-on-track-achieve-sdg4.pdf

United Nations (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. Available online at: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

Vescio, V., Ross, D., and Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teach. Teacher Educ.* 24, 80–91. doi: 10.1016/j.tate.2007.01.004

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., Capin, P., et al. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educ. Psychol. Rev.* 28, 551–576. doi: 10.1007/s10648-015-9321-7

Webber, M., and Macfarlane, A. (2020). Mana tangata: the five optimal cultural conditions for Māori student success. *J. Am. Ind. Educ.* 59, 26–49. doi: 10.1353/jaie.2020.a798554

WHO (2020). *Statement From the 'Civil Society' Track of the 3rd Global Infodemic Management Conference. WHO*. Available online at: https://www.who.int/news/item/10-12-2020-statement-from-the-civil-society-track-of-the-3rd-global-infodemic-management-conference

Zubrick, S. R., Taylor, C. L., and Christensen, D. (2015). Patterns and predictors of language and literacy abilities 4-10 years in the longitudinal study of Australian children. *PLoS ONE* 10:e0135612. doi: 10.1371/journal.pone.0135612