
Article:
Mesa, C, Newbury, DF, Nash, M et al. (2020) The effects of reading and language intervention on literacy skills in children in a remote community: An exploratory randomized controlled trial. International Journal of Educational Research, 100. 101535. ISSN: 0883-0355

https://doi.org/10.1016/j.ijer.2020.101535

The Effects of Reading and Language Intervention on Literacy Skills in Children in a Remote Community: An Exploratory Randomized Controlled Trial

Carol Mesa[1]*, Dianne F. Newbury[2]*, Marysia Nash[3], Paula Clarke[4], Rosanne Esposito[5], Louise Elliott[6], Zulema De Barbieri[7], Ma. Angélica Fernández[7], Pia Villanueva[7], Charles Hulme[8], Margaret J. Snowling[1,3]*

1: Department of Experimental Psychology, University of Oxford, Oxford, UK, OX2 6HG;

2: Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK, OX3 0BP;

3: Department of Experimental Psychology and St John's College, University of Oxford, Oxford, UK, OX1 3JP;

4: School of Education, University of Leeds, Leeds, UK, LS2 3AR;

5: Institute of Education, University College London, London, UK, WC1H 0AL;

6: York Trials Unit, Department of Health Sciences, University of York, UK, YO10 5DD

7: Departamento de Fonoaudiología, Universidad de Chile, Santiago, Chile;

8: Department of Education, University of Oxford, Oxford, UK, OX2 6PY;

* - these authors contributed equally to this work

Correspondence concerning this article should be addressed to Prof Margaret Snowling, Department of Experimental Psychology, University of Oxford, Oxford, UK, OX2 6HG, +44 1865 277419. E-mail: maggie.snowling@sjc.ox.ac.uk

Running Head: *A Reading and Language Intervention*

Abstract

This study explored the effects of a 27-week reading and language intervention, for low-income children living in a remote Chilean community, using a randomized controlled trial. At the end of the intervention, children in the intervention group showed improvements compared to the waiting group on pre-literacy, reading, language, and reading comprehension measures (effect sizes d >.25). The gains in pre-literacy skills, word reading and word knowledge were maintained at 9-month delayed follow-up, though the improvements in language and reading comprehension were not. Intervention programs designed to support literacy and language skills in remote communities can be delivered successfully by trained assistants. Our findings suggest that language and literacy programs can be useful for improving attainment in children living in disadvantaged and isolated communities.

Keywords; reading and language intervention, remote community, RCT, development trial

1. Introduction

Language skills form the cornerstone of educational achievement and children who enter school with poor language and reading skills typically struggle in the classroom. We know that structured, multi-component intervention programs can help to ameliorate some of these difficulties when applied in Western education contexts (Whitehurst & Lonigan, 1998). However, there is a dearth of evidence about 'what works' for literacy instruction in communities with differing cultural backgrounds (Bekman, Aksu-Koç, & Erguvanlı-Taylan, 2011; Nag, Chiat, Torgerson, & Snowling, 2014; Opel, Ameer, & Aboud, 2009) where it is widely reported that rural, indigenous populations are "vulnerable" to literacy failure (LAC Reads Capacity Program, 2016). The reasons underlying this are complex and the circumstances are usually unique to a given population. Nonetheless, contextual challenges can be common across communities and include geographic factors (which can limit delivery of materials and accessibility to education), demographic influences (including social and economic disadvantages and differences in expectations), educational settings (which may have to be adapted to allow for vulnerabilities) and social isolation. It is important to understand the impact of these factors upon the "real-world" delivery of intervention programs.

In this study, we explore ways of improving language and literacy skills in remote communities. We assess the generalization of a program built upon evidence-based practices and successful intervention methods from Western cultures (Foorman, Breier, & Fletcher, 2003; McCardle & Chhabra, 2004; National Reading Panel, 2000). The principles of the program build upon the 'Simple View of Reading' (Gough & Tunmer, 1986) and draw on longitudinal data that suggest causal links between language as a foundation for literacy, mediated by phonological awareness and decoding, and a direct effect of oral language skills on reading

comprehension (Hulme, Nash, Gooch, Lervag, & Snowling, 2015; Lervåg, Hulme, & Melby-Lervåg, 2018). Accordingly, we trained decoding skills (letter-sound knowledge and phonological awareness in the context of graded book reading following the strategies of an established UK program (Hatcher, Hulme, & Ellis, 1994)), alongside language comprehension (vocabulary and narrative skills, and reciprocal reading strategies following (Clarke, Snowling, Truelove, & Hulme, 2010)). We hypothesized that the program would lead to gains in language and literacy skills in the disadvantaged and isolated community studied.

1.2 The study population

This study involves the community of the Robinson Crusoe Island, 600km to the West of mainland Chile. This population is geographically and socially isolated and is reported to have a high incidence of language disorders (De Barbieri, Fernandez, Newbury, & Villanueva, 2018). Difficulties are reported across a range of linguistic measures including phonological production and expressive and receptive morphology (De Barbieri et al., 2018).

People on Robinson Crusoe Island live in a single town and, except for a few children who are home-schooled, all children attend the same school in which Spanish is the language of instruction. Since a tsunami in 2010, the school has been housed in shipping containers with limited light and ventilation and poor conditions for teaching. Access to the island (and the internet) is difficult, further restricting provisions for high-quality education. Added to this, travel to the mainland for medical and dental treatment can affect regular school attendance.

1.3 Reading instruction in Chile

Beyond theoretical grounding, macro-environmental factors need to be taken into account when considering the design and implementation of an intervention. These include the home and school language of the community and the prevailing educational policy. Traditionally in Latin

America, reading is taught solely through the school system and is considered a mechanical skill. In many Chilean schools, children are taught to read through rote learning and reading fluency is considered a proxy of comprehension. Given the transparency of the Spanish orthography, sound-letter linkages are not explicitly taught and the teaching of reading focuses upon language structure (grammar) and memorization methods (Ponce, López, & Mayer, 2012). Children are often encouraged to read stories aloud with little focus upon context, inference or meaning. Opportunities for book sharing, dialogic interaction and group work are limited (Strasser, Mendive, Vergara, & Darricades, 2018).

Chilean law requires one adult (teacher or teaching assistant) per 16 children in kindergarten, with a maximum class size of 32 (Strasser et al., 2018). The school on Robinson Crusoe Island follows the Mainland curriculum but, due to space constraints, has mixed-age classrooms. All lessons are given in Chilean Spanish.

In 2006, the Chilean Government announced the "Chile Crece Contigo" program aimed at providing essential early childhood services in deprived areas. Three areas of support were identified: language and literacy development, classroom behavior management and improved coordination between education and health care (Peralta, 2011). Since then, new teacher-training strategies have been introduced; however, school systems practice freedom in curricula activities meaning that these new skill sets are not always put into practice regularly (Yoshikawa et al., 2015).

1.4 Reading intervention studies in Latin America

In a systematic review of intervention studies in Latin America and the Caribbean (LAC), the LAC Reads Capacity Program (2016) found that phoneme awareness, phonics, fluency and language comprehension were associated with reading outcomes in early-grade readers, supporting

the Simple View of Reading (LAC Reads Capacity Program, 2016). They coincidentally report disparities among rural, indigenous and disadvantaged populations and a correlation between socioeconomic status and reading ability in early grades, supporting the importance of targeted intervention strategies (LAC Reads Capacity Program, 2016). The systematic review reported five previous reading RCTs in Chile (Gomez Franco, 2014; Mendive, Weiland, Yoshikawa, & Snow, 2016; Pallante & Kim, 2013; Strasser & Lissi, 2009; Yoshikawa et al., 2015) many of which focused upon teacher-training in mainland Chile (Gomez Franco, 2014; Mendive et al., 2016; Yoshikawa et al., 2015). Positive effects were reported for teachers, who received training in vocabulary, oral production, comprehension, and writing teaching methods, but no significant effects were found for student outcomes (Gomez Franco, 2014; Mendive et al., 2016; Yoshikawa et al., 2015). The authors observed that although teachers were competent and engaged with the program, delivery may be limited by factors that prevent consistent and automatic presentation within the classroom. Accordingly, they introduced a teacher checklist to encourage self-monitoring (Strasser et al., 2018). This increased the frequency with which teachers applied target language tasks (such as shared reading, taught vocabulary use, discussion of books), but did not improve child oral language skills (Strasser et al., 2018).

Positive, but small, effects were reported in other Chilean studies using alternative intervention strategies. Pallante and Kim (2013) reported non-significant gains in reading through a teacher training program that targeted phonology, fluency, vocabulary, writing and reading comprehension in kindergarten and first-grade children (SMD=0.16, 95% CI=-0.17-0.48) (Pallante & Kim, 2013). Larrain, Strasser and Lissi (2012) found that word elaboration and explicit vocabulary training through shared reading had positive effects upon vocabulary acquisition. Ponce *et al.* (2012) evaluated a computer-based program that taught reading

strategies to support language comprehension across 21 schools (both municipal and private) in Central Chile (Ponce et al., 2012). They found significant gains (d=0.5) for those children who started the program with the lowest reading ability. Although some rural schools were included in the study, only one population was classified as having low socioeconomic status (Ponce et al., 2012). Rosas *et al.* (2017) also used computer-based methods within a phonological training game (GraphoGame) in low-school achievers in mainland Chile. There was a direct effect of the program upon knowledge of letter-sound linkage but this did not generalize to reading comprehension (Rosas, Escobar, Ramírez, Meneses, & Guajardo, 2017).

In short, while some positive effects have been reported in Chilean reading intervention studies, most have focused upon a single aspect of reading-related skills and do not make a distinction between decoding and reading and language comprehension. Few studies include vulnerable communities and many are reported to have selection biases or flawed experimental designs (LAC Reads Capacity Program, 2016). There is, therefore, a clear need for additional research on the application of evidence-based intervention programs in vulnerable communities. Information regarding the portability of evidence-based interventions is critical in order implement programs at scale in developing countries and, so address the Sustainable Development Goals (SDGs) of the United Nations; 'to employ creative and innovative solutions in order to change the course of education and raise educational standards' (SDG4 https://www.un.org/sustainabledevelopment/education/).

1.5 Aims of the current study

The current study designed and applied a 27-week multi-component intervention program (the "Chilean Language And Reading Alliance – CLARA"), on the Robinson Crusoe Island in

Chile. The mechanisms of delivery, implementation, context and outcomes are described in the logic model shown in Figure 1.

<Figure 1 about here>

CLARA incorporated training in letter-sound knowledge, decoding and phonological awareness (e.g. Catts, Herrera, Nielsen, & Bridges, 2015) in the context of graded book reading (Hatcher et al., 1994) in addition to training in vocabulary (e.g. Catts et al., 2015), narrative and skills (e.g. Griffin, Hemphill, Camp, & Wolf, 2004; Justice et al., 2006) together with a focus on reading strategies that allow text understanding beyond literal interpretation (e.g. Duke, Pearson, Strachan, & Billman, 2011; Hogan, Bridges, Justice, & Cain, 2011; Perfetti, Landi, & Oakhill, 2005) as described by (Clarke et al., 2010). CLARA was delivered within a school setting by four resident tutors. We predicted that children who received the intervention would show stronger reading and language skills than those in the waiting-list arm. We also hypothesized that gains in pre-literacy and basic reading skills would mediate short- and long-term gains in reading comprehension.

2. Materials and Methods

2.1 Participants

The entire child population of the island (132 children), aged 4-14 years, was eligible for the study. Sample size was limited by total population size. Following information sessions, parents of 78 children gave signed consent and were enrolled by the first author. Ten children then relocated to another island leaving a final sample of 68 (Figure 2). Ethical approval was granted by [Removed to preserve anonymity].

We conducted a randomized controlled trial, with a cross-over design, in accordance with the CONSORT guidelines (Schulz, Altman, & Moher, 2010) (Figure 2). Group allocation was

conducted independently by the University of York Trials Unit. Sixty-eight participants completed pre-tests to assess language and reading ability. Participants were allocated to intervention (experimental) (n= 34, 17 male, 17 female) or waiting (control) groups (n =34, 17 male, 17 female) minimizing for age and a composite score of pre-test language measures.

Our previous work using similar methods yielded standardized mean differences between the intervention and control groups with medium to large effect sizes (standardized mean differences in the range .4 to .8). A sample sizes of 68 (34 per arm), and a pre-test post-test correlation of .8 (which is line with expectations) using ANCOVA gives 80% power to detect a group difference with a standardized mean difference = 0.41, two tailed, p = .05 (a medium effect). In summary, on reasonable assumptions, we have adequate power here to detect effects of the size anticipated.

<Figure 2 about here>

2.2 Measures

The primary outcomes of the intervention were captured by factor scores encapsulating multiple measures of reading (word-level decoding) and language (vocabulary) skills. Three factors of basic reading and language skills were formed to capture these primary outcomes; pre-literacy skills (phoneme awareness and letter knowledge), word reading (one-minute word reading, picture word matching reading test and nonword reading), and vocabulary (CELF and WISC vocabulary, receptive vocabulary, and bespoke vocabulary). Secondary outcomes were reading comprehension, listening comprehension and oral narrative skills. All measures were collected pre- (t1) and post- (t3, 27 weeks after the start of intervention) intervention. Baseline measures of nonverbal IQ, recalling sentences and arithmetic ability were also collected. Additional assessments to monitor progress were performed at mid-test (t2, 14 weeks after the

start of the intervention) and 9 months after the completion of intervention for the intervention group (t4); this was week 20 of the intervention for the waiting control group. A full list of measures taken at each time point and their reliabilities (alpha) can be found in Table 1. Further details of tests are given below.

Where possible, assessments employed standardized Chilean or European Spanish measures. Test rubrics were adapted for the target population by Spanish Speakers from Chile. In some instances (vocabulary, reading comprehension), standardized tests were not available and assessments were developed by translation of standardized English tests. Translated items were reviewed by three Spanish-speaking people and assessed for cultural suitability by an Islander. After translation, these tests were trialed in a group of 8 Spanish-speaking children and modifications were made as needed. Ten percent of the audio-recorded measurements (WISC vocabulary, bespoke vocabulary, listening comprehension, reading comprehension and story grammar) were double scored to obtain point-by-point percent of agreement allowing estimation of inter-rater agreement.

Assessments took two hours, over three sessions, and the order of tasks varied across participants. Assessors were blind to group allocation at t1 (pre-test) and t3 (post-test), and trained in test administration and scoring. All data entries were double checked, approximately 50-60% by the first author.

2.2.1 Primary outcome measures

The primary outcome measures were pre-literacy and basic reading skills (decoding, word recognition and vocabulary). All reading tasks were taken from Caravolas et al. 2012, a longitudinal study of literacy development across four languages, including Spanish. In this previous study, all tasks were tested and standardized in 190 native Spanish speakers aged 61-73

months (Caravolas et al., 2012). One vocabulary measure (CELF-4) had been standardized in Spanish. The other two measures (WISC-IV and ROWPVT) were translated from tests developed and standardized on English-speaking children.

Letter knowledge. This task assessed children's knowledge of letter names and sounds. Twenty-nine Spanish letters (i.e., 27 letters and 2 digraphs) were assessed. The number of letters correct was summed to give measures of letter-name and letter-sound knowledge (Caravolas et al., 2012).

Phoneme awareness. This measure assessed the identification, deletion, blending and segmentation of phonemes in spoken words (Caravolas et al., 2012). The test consisted of 30 items at t1 and 60 items for the remaining administration points. The score represents the number of correct responses across the test. Testing was discontinued after four consecutive errors.

Word Reading. Two tasks were used to measure children's word reading. In the Picture Word Matching Reading Test (PWMRT), the child identifies the word that matches a target picture for as many words as possible in three minutes (max=63) (Caravolas et al., 2012). The score represents the number of correct target pictures identified in one minute. In the one–minute reading test, children must read as many words as possible from a list of 140 high-frequency Spanish words from two to five syllables (Caravolas et al., 2012). The score represents the number of correct words read in one minute.

Nonword reading. This task measured decoding and comprised 140 nonwords of between 2 and 5 syllables constructed to resemble Spanish word structures (Caravolas et al., 2012). The score represents the number of nonwords read correctly in one minute.

Spelling. A spelling-to-dictation task measured children's word-spelling skills (Caravolas et al., 2012). Ten target words were dictated at t1, and 35 at later time-points. The score represents the number of words correctly spelled.

Expressive vocabulary. Children's expressive vocabulary was assessed using two subtests from the CELF-4 (Spanish version) and WISC-IV (English version) respectively. In the CELF-4 vocabulary subtest, children name a set of pictures (Semel, Wiig, Secord, & Langdon, 2006). At t1, twenty nine words were tested. After t1, nine bespoke words were added to this original subtest to avoid ceiling effects. The scoring rubric allowed for dialectal differences and children received 2 points for correct definitions, 1 point for partial definitions, and 0 points for incorrect or unrelated definitions. Testing was discontinued after eight consecutive errors.

In the Wechsler Intelligence Scale for Children (WISC-IV) vocabulary subtest, children define a series of 36 words (Wechsler, 2003). Children received 2 points for correct definitions, 1 point for partial definitions, examples or correct use of the word in a sentence, and 0 points for incorrect or unrelated definitions. Testing was discontinued after six consecutive errors. Inter-rater agreement varied from 89 to 100% across cases.

Receptive vocabulary. Children's receptive vocabulary was assessed using the Receptive One-Word Picture Vocabulary Test (ROWPVT) translated from English to Spanish. In this test, the child points to the picture that best represents the stimuli word (Martin & Brownell, 2012). One hundred pictures are provided and the starting point is determined by the child's age. The score represents the number of correct responses across the test.

Bespoke vocabulary. This experimental measure assessed children's ability to define words explicitly taught during the intervention. Children had to provide a definition for each of 21 target words. Responses were scored 2 points if the child provided a complete definition, 1

point if the child provided a partial response (e.g., incomplete definition, an example, or correct use of the word meaning in a sentence), and 0 points if the child's definition was incorrect or unrelated. Inter-rater agreement was 81-100%.

2.2.2 Secondary outcome measures

The secondary outcomes were measures of reading comprehension,oral language proficiency (narrative) and listening comprehension skills which are predicted to develop once baseline skills are in place (Scarborough, 2001).

Reading comprehension. A Spanish translation of the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2009) was used to measure reading comprehension as no Chilean Spanish measure was available. In this test, the child read six short paragraphs aloud. After each paragraph, the assessor asked eight comprehension questions. Responses were scored 1-point if the child provided a correct response and 0-points if the response was incorrect. Inter-rater agreement ranged from 93-100% across stories.

Listening comprehension. Children's listening comprehension was assessed with the Evaluación del Discurso Narrativo (EDNA; Guzmán, Tirapegui, & Landaeta, 2012). This test was developed and standardized in Chile (Pavez, Coloma, & Maggiolo, 2008). The assessor reads a short story after which the child is asked 10 comprehension questions. Responses were scored 2 points if the child provided a complete response, 1-point if the child provided a partial response or 0-points if the response was incorrect or unrelated. Inter-rater agreement was 90-100%.

Narrative production. Children's narrative language was assessed using Procedimientos para Evaluar Discurso (PREDI). This test was developed and standardized in Chile (Pavez, Coloma, Maggiolo, Martínez, & Romero, 2002). The assessor asks the child to arrange a picture

sequence and tell a story about it. Children's stories were audio recorded, transcribed verbatim, and coded using Systematic Analysis of Language Transcripts (SALT) software (Chapman & Miller, 1984). Using transcriptions, children's narratives were analyzed to generate three measures of narrative; story grammar, Mean Length of Utterance in words (MLU-w), and Subordination Index.

The measure of Story Grammar determined whether the child's narrative included key elements of Story Grammar (character, setting, problem, emotion, attempt, consequence and end). Each story element was scored as 2, 1 or 0. Inter-rater agreement was 71-100%.

Mean Length of Utterance was the average of number of words per C-unit (a main clause or a main clause with all its subordinating clauses). This measure of child language complexity is reliable for Spanish speakers (e.g. Gutierrez-Clellen & Hofstetter, 1994).

The subordination index is a measure of language complexity (total number of clauses divided by total number of utterances) (Heilmann, Miller, Nockerts, & Dunaway, 2010).

2.2.3 Cognitive and control measures

Nonverbal ability (pre-test only). Children's nonverbal ability was assessed using the Raven's Progressive Matrices (test-retest reliability from 0.83 to 0.93 according to age (RAVEN; Raven & Court, 1998).

Recalling Sentences (pre-test only). The Recalling Sentences subtest from CELF-4 (Spanish version) evaluates the ability to repeat spoken sentences of increasing complexity (Semel et al., 2006).

Arithmetic (pre-test only). The Addition and Subtraction subtasks from Test of Basic Arithmetic and Numeracy Skills (TOBANS; Brigstocke, Moll, & Hulme, 2016) assessed children's arithmetic ability.

2.3 Principles of the intervention

CLARA aimed to improve literacy skills by promoting decoding skills and oral language comprehension and by linking these two sets of skills. It was designed to follow a highly structured and systematic format (so that pupils knew what to expect in each session) and to make use of multi-sensory techniques for the teaching of code-related elements and vocabulary. The general aim was to deliver an individualized program tailored to the specific needs of individual children but involving small group work (usually dyads) to encourage interaction, particularly in relation to language learning. The program was manualized with detailed lesson plans; however tutors were trained to ensure that teaching was at an appropriate level, and for reading, in the proximal level of instruction to avoid despondency. Practice and reinforcement was embedded in the approach; teachers planned lessons using checklists, recorded pupil progress using running records for reading and narrative recordings for language and conducted more formal checks on emergent reading skills. Materials were based on existing intervention programs but were adapted to fit within the context of the community, choosing books and activities which were meaningful to participants and therefore engaging, using a variety of texts and genres.

2.4 Intervention program and components

Children received four sessions a week following three formats; (1) a reading component that targeted decoding and basic reading skills, (2) a language component that supported vocabulary and narrative language, and (3) a reciprocal teaching component that promoted the use and understanding of reading strategies. At the start of the program, all children received three 20-minute reading sessions and one 30-minute language session a week. At mid-test (t2, after 14 weeks of intervention), all children with proficient decoding skills were introduced to

the reciprocal teaching component. Decoding proficiency was defined as having letter knowledge 0.4SD and phonological awareness 0.25SD above the group mean and being able to read a book graded greater than 21 on a scale developed by (Hatcher et al., 1994) at >94% accuracy. For these children (n=19), the dose of all 3 components was rebalanced such that they received one 20-minute reading session, one 30-minute language session and two 30-minute reciprocal teaching sessions per week for the remainder of the program (13 further weeks). The children who were not proficient decoders at t2 (n=15) continued receiving three 20-minute reading sessions and one 30-minute language session per week (over 13 further weeks).

All sessions were designed to allow modelling of target skills, guided practice, repeated exposure, rich interactions and opportunities for independent use. Example lesson plans are provided in Appendix 1 and each of the three components are described below. Researchers who were familiar with the local context checked that the materials were suitable for the community and adapted the materials if necessary. All sessions were administered during the school day outside of the classroom. Children in the intervention group received four sessions a week delivered over 27 weeks between October 2017 and October 2018. Children in the waiting control group received lessons as normal during this time. Following 27 weeks of interventions sessions and immediately after post-test, the children in the waiting control group began the program and received four sessions a week delivered over 20 weeks due to funding constraints. Child attendance was monitored throughout the program.

2.4.1 Reading component

One-to-one reading sessions targeted basic reading skills; letter knowledge, decoding, phoneme awareness, sight word reading and reading practice. Before starting, tutors assessed participants' pre-literacy and reading skills to determine the sounds, letters, and level of

phonological awareness that each child needed to work on. Tutors selected activities from a bank of resources to help the child practice linking letters and sounds and completed some simple writing activities to reinforce these. Importantly, they also practiced reading connected text using books at a level appropriate for their reading ability, determined by a reading accuracy of at least 94% on a particular book (Clay, 1985; Hatcher et al., 1994). Teachers made a running record to check that selected books were at an appropriate level and to decide upon a teaching point to target in the next lesson.

2.4.2 Language component

Language sessions were delivered to groups of two students and aimed to promote vocabulary and narrative skills through story discussion and retelling (Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Fricke et al., 2017). Lessons plans were scripted, but tutors were encouraged to use comments, questions, feedback, and support to maintain appropriate responsiveness. Sessions were based around a selected book in which five tier-2 words (Beck & McKeown, 2007) appeared across topics and content-areas. The same book was presented across three consecutive sessions (Beck & McKeown, 2007; Beck, McKeown, Sinatra, & Loxterman, 1991). To encourage narrative skills, children participated in rich discussions about the book and were prompted to retell the story. Children were asked literal and inferential questions and engaged in conversation about story elements. These sessions were frequently audio-recorded encouraging children to self-monitor missing elements and/or words to connect their ideas. Tutors transcribed audio-recordings of children's narratives and used these to decide on appropriate teaching points for the next lesson.

2.4.3 Reciprocal teaching component

Competent decoders took part in reciprocal teaching sessions (after Palinscar & Brown, 1984), designed to support the development of reading comprehension through the promotion of independent metacognitive strategies (Clarke et al., 2010). These sessions took place in groups of two students and were child-led with support from the tutor. Children practiced the use of four taught strategies (predicting, summarizing, questioning and clarifying) while reading books designed for this component (Global Education Systems Connectors Reciprocal Reading books). As recommended by (Reutzel, Smith, & Fawson, 2005), the strategies were introduced together rather than sequentially.

2.5 Tutor training

All intervention sessions were delivered by three native Chilean Speech Language Therapists; these professionals had a strong background in child language but training included essential background to ensure knowledge of how language is a foundation for learning to read within the Framework of the Simple View of Reading. Following this, and a review of evidence-based approaches to literacy instruction, more specific training was given on how to implement the program, use the manual and to assess and monitor progress before the intervention began. Training lasted 25 hours and consisted of presentations by the first author and sessions explaining the program's components and procedures, with delivery modelled. The tutors performed mock sessions with feedback and created activities for the reading component and lessons for the language component. Activities were created within the framework given and approved by the project leads prior to use.

During the delivery, the tutors received bi-weekly telephone calls from the first author who offered support and discussion as the intervention proceeded. The island was visited ten

times over the course of the program by the senior researchers allowing face-to-face contact and feedback.

2.6 Intervention fidelity

Ten percent of sessions were video-recorded and coded by two trained researchers to assess fidelity. Implementation of program indicators ranged from 87-100% for the reading component, 73-94% for the language component, and 83-95% for reciprocal teaching. Twenty percent of the videos were double-coded by the first author to estimate inter-rater agreement (89-95%). Attendance records were used to determine the amount of intervention received per child. On average, in the intervention group, children attended 82% of available sessions (SD=12%, Range 49% to 100%). Non-attendance was due to visits to the Mainland, absence from school and clashes with school timetabling.

3. Theoretical framework

The theoretical framework guiding program development was the Simple View of Reading (Gough & Tunmer, 1986). This model proposes that the ability to understand written text relies upon decoding, which facilitates links between text and sound, and language comprehension, which is critical for reading with meaning. Importantly, decoding is a crucial step in this process. Weaknesses in decoding skills are an obstacle to reading comprehension (Castles, Rastle, & Nation, 2018). Conversely, strong decoding skills allow the effective application of language comprehension within a multiplicative model, enabling advanced levels of text processing, such as making inferences (Cain & Oakhill, 1999). This model has proven a strong framework for effective teaching and intervention programs in developed countries for both typically developing and language-impaired children (reviewed by Stahl & Fairbanks, 1986). The evidence supporting this framework has been obtained in a range of alphabetic

languages (Norwegian (e.g. Høien-Tengesdal & Høien, 2012), Swedish (e.g. Høien-Tengesdal, 2010), French (e.g. Megherbi, Seigneuric, & Ehrlich, 2006), Spanish-speakers in the USA (e.g. Joshi, Tao, Aaron, & Quiroz, 2012), Italian (e.g. Tobia & Bonifacci, 2015) and Finnish (e.g. Torppa et al., 2016)), as well as in languages that use non-alphabetic writing systems (Chinese (e.g. Hulme, Zhou, Tong, Lervåg, & Burgoyne, 2019; Joshi et al., 2012))

At the macro-environment level, reading development requires motivation and literacy outcomes can be influenced by factors such as teacher and parent expectations, demographic factors (including the home environment, Sénéchal & LeFevre, 2001) and dialect differences (Aaron, Joshi, Gooden, & Bentum, 2008; Francis, Kulesz, & Benoit, 2018; Nag, Snowling, & Asfaha, 2016). Of particular relevance to the current study, is that children with language learning impairments may struggle with basic decoding even in transparent languages. In vulnerable populations, baseline skills may differ from expected, demographic differences are inflated, education may not be prioritized and engagement may require adaptations. Moreover, intervention programs that target social, emotional and environmental factors have been shown to have small but positive effects upon reading attainment and vocabulary in vulnerable populations (Corcoran, Cheung, Kim, & Xie, 2018; Leyva & Skorb, 2017). Together these finding underscore the inter-dependence of environmental and educational factors and support the idea that a multi-componential and holistic approach is required when delivering reading intervention programs to vulnerable populations.

4. Results

4.1 Population Demographics

Demographic information was collected from parents of participants through a questionnaire. Information is presented here to contextualize the cohort (but was not used in the

analyses). On average, participant children spent six weeks of the year on the Mainland, usually during the school holidays. Twenty eight percent of participants (nineteen children) had special educational needs which warranted an individual learning plan at school. Twenty six percent of caregivers reported that their child had a reading difficulty (eighteen children) and five percent reported a language difficulty (three children). These individuals were evenly spread across the intervention and waiting control groups (the differences in parental report for reading and language problems is typical, the latter being less easy to notice in a familiar family setting). Thirty percent of participants (twenty children) had at least one parent who was born on the island. All but one child lived with their Mother or Father. All but one household included at least one working adult and an average of 1.8 children under the age of 18 (range 1-3). Ninety five percent of participants (sixty five children) had at least one parent who completed high school. Thirty eight percent of participants (twenty six children) had at least one parent who attended college and seven percent (five children) had at least one parent with a higher education qualification. Ninety three percent of caregivers (sixty three individuals) said that they thought being able to read was an important life skill.

4.2 Analysis plan

The main objective of the study was to evaluate gains made by the intervention group at the end of the program, compared with those of the waiting-list control group in the same school. We therefore planned, for each measure, to use analysis of covariance to assess outcome at t3 controlling for baseline performance at t1 and additional covariates. We included tests of the interaction with covariates as predictors, proceeding to drop the interaction terms if not significant. In addition to analyzing gains for separate measures, we planned to use factor scores to derive reliable constructs of pre-literacy, reading and vocabulary and to subject these to the

same analyses. If gains in reading comprehension were significant, we also planned to test the mediators of gains using mediation analysis. Finally, we planned to use ANOVAs to assess whether any gains were sustained.

4.3 Baseline reading and language skills

The CLARA study involved 68 children (34 male, 34 female), aged between 4;6 and 14;8 years at baseline assessment (t1) (mean 8;7, median 8;4, SD 2;7). Despite the age range and the transparency of the Spanish language, many children showed inconsistent letter knowledge and poor phonological awareness. None of the 68 children tested scored full marks on the test of letter knowledge, which assessed knowledge of names, sounds for both capital and small letters. The maximum mark achieved was 115 of 129 (89%) which was obtained by three children (aged 7;9, 9;3 and 10;5). It was noted that children often added vowel sounds onto the letter forms and frequently confused the sounds for the capital and small letters. Twenty of the 68 children tested (29%, aged 4;6 to 7;8) knew only half of the letter-sound correspondences and twenty (29%, aged 4;6 to 10;0) could not read any of the non-words tested indicating poor decoding ability. Phonological awareness was assessed at the level of phoneme, syllable and rhyme. Children below the sixth grade often struggled with phoneme deletion and blending tasks. Of the 68 children tested, fifteen scored zero (aged 4;6 to 6;4) and 27 (40%, aged 4;6 to 11;10) scored less than 50% indicating poor phonological awareness even among the older children. Eighteen children (26%, aged 4;6 to 7;4) could not read any words at baseline and most kindergarten children could not recognize their own names. For those that could read paragraphs (47 children aged 5;4 to 14;8), fluency was generally good with an average reading rate of 81 words per minute (SD=36.5) and an average reading accuracy of 96% (SD=3.9%, range 79.7%-99.9%) across the YARC passages at t1.

In contrast to reading-related assessments, the vocabulary skills of the children were generally good. The majority of children above the fourth grade (aged 9 years plus) scored well in the CELF vocabulary (picture-naming) task. On average, across this higher age range, children knew the names of 51 of the 58 pictures tested. However, testers reported that although children had wide vocabularies and enjoyed telling stories, their narratives were often poorly constructed and did not always make sense.

4.4 Intervention gains

Gains between pre-test (t1) and post-test (t3) measures were compared between intervention and waiting control groups for individual measures using analysis of covariance models in Stata (Version 15.1; StataCorp. 2017). In each model, outcomes were predicted from the same measure at pre-test (the covariate) and a group dummy code; minimization variables (age, t1 CELF vocabulary, and t1 recalling sentences) were used as additional covariates. Effect sizes were calculated using Cohen's *d*; the difference in gains (t1 to t3) between groups divided by the pooled pre-test standard deviation (see Morris, 2008). The equality of regression slopes was assessed in all models, by including a group by covariate interaction term. This interaction was only significant for the pre-literacy factor; for all other measures the interaction terms were trivial in size and omitted from the models. Descriptive statistics for individual measures, effect sizes of observed gains (t1 to t3) and measures of consistency are reported in Table 1.

<Table 1 about here>

To simplify analyses and increase power, pre- and post-test factor scores were constructed using exploratory factor analyses. Three factors were formed; pre-literacy skills (phoneme awareness and letter knowledge), word reading (one-minute word reading, picture word matching reading test and nonword reading), and vocabulary (CELF and WISC

vocabulary, receptive vocabulary, and bespoke vocabulary). In all cases, clear 1-factor solutions were provided with substantial and relatively uniform factor loadings. Pre-test factor scores were generated using regression scoring, whereas post-test factor scores employed the Bartlestt method (Skrondal & Laake, 2001). The intervention group showed greater gains on all three factor scores (pre-literacy, word reading and vocabulary) than the waiting control group. For the pre-literacy factor score there was a negative interaction such that children with the weakest skills at pre-test made larger gains. There was a substantial advantage for the intervention group evaluated at the mean of the covariate (i.e. for children with average or below average initial scores; standardized mean difference between groups =0.83).

The intervention group also showed significantly greater gains than the waiting control group on vocabulary factor scores (formed from CELF and WISC vocabulary, receptive vocabulary, and bespoke vocabulary) (difference between groups = 0.34 [95% CI 0.17, 0.51], t=4.03 p<0.001; standardized mean difference between groups =0.32) with an equivalent pattern for word reading factor scores (formed from one-minute word reading, picture word matching reading test and nonword reading) (difference between groups = 0.27 [95% CI 0.08, 0.47], t=2.46, p=0.017; standardized mean difference between groups =0.28).

For the individual measures not included in the factor scores, the intervention group showed gains in reading comprehension (difference between groups = 5.88 [95% CI 3.25, 8.53], t=4.46 p<0.001; standardized mean difference between groups=0.498), MLUw (difference between groups = 1.40 [95% CI, 0.71, 2.10], t=4.05 p<0.001; standardized mean difference between groups = 0.74), and subordination index (difference between groups = 0.35 [95% CI, 0.25 0.45], t =7.07 p<0.001; standardized mean difference between groups = 1.39). No

significant gains were seen for listening comprehension, story grammar and spelling (standardized mean difference between groups = 0.09; -0.15 and 0.04 respectively).

4.5 Mediation Analyses

Mediation analysis was employed to explore the effects of intervention on word reading and vocabulary factors and reading comprehension at t3, after controlling for autoregressive effects of each of these variables at t1. Analyses were performed in Mplus (Muthén & Muthén, 2012) using maximum likelihood estimation, with bias corrected bootstrapped confidence intervals; missing data were handled with Full Information Maximum Likelihood estimation (FIML).

The mediation model provided a good fit to the data (Figure 3). Both word reading and vocabulary factors at t3 independently predicted reading comprehension. Furthermore, the indirect path from intervention to reading comprehension via word reading at t3 was significant (ß=0.042 [95% CI 0.006, 0.094]) as was the indirect path from intervention to reading comprehension via vocabulary at t3 (ß=0.062 [95% CI 0.022, 0.107]). There was also a direct effect of intervention on reading comprehension in the final model (ß=0.117, p=0.001) (Figure 3).

<*Figure 3 about here*>

Finally, ANOVAs were used to examine trajectories across t1, t3 and t4 for individual measures. Continued progress to post-test was observed for the one-minute word reading test (Figure 4). For measures of letter knowledge, phoneme awareness, nonword reading, CELF vocabulary and bespoke vocabulary, gains were maintained although no further progression was made following the end of the intervention (Figure 4). Measures of narrative (MLUw and

subordination index) and reading comprehension fell back to the starting level after the intervention ceased.

<Figure 4 about here>

5. Conclusions

In this study, we developed a language and reading program for Spanish-speaking children from a vulnerable population and evaluated its effectiveness in an exploratory randomized trial. Despite the challenging context in which the program was delivered, it produced substantial improvements in language and literacy, and, in line with the predictions of the Simple View of Reading, these developments partially mediated significant gains in reading comprehension. Our findings provide 'proof in principle' that evidence-based reading and language interventions can be implemented in vulnerable communities where there are low levels of attainment.

At baseline assessment, many children had poor letter-sound knowledge and none of the children were able to provide letter sounds or names for every symbol. These observations are in line with the conclusions made by Strasser *et al.* (2018) regarding the application of theoretical models within the Chilean classroom under a busy curriculum. Although emerging policies mean that teachers are trained to understand that decoding is an important skill in reading development, the "free" nature of the Chilean curriculum may mean that such basic skills are over-looked within post-kindergarten classrooms. Alternatively, in line with previous studies, which show that decoding may be disrupted in children with reading difficulties (Rosas, Escobar, Ramírez, Meneses, & Guajardo, 2017), these findings may reflect the high level of language difficulties documented within this particular population. Since other studies indicate that home environment (Nag, Vagh, Dulay, & Snowling, 2019; Strasser & Lissi, 2009) and socioeconomic

factors (Lervåg, Dolean, Tincas, & Melby-Lervåg, 2019) can affect the trajectory of reading growth, these findings may be further explained by the isolated and rural nature of this population. Whatever the route of mediation, given the multiplicative nature of decoding and listening comprehension within the Simple View of Reading, our baseline investigations reinforce the need to include pre-literacy skills within educational programs even in transparent languages and especially in vulnerable populations with a high incidence of disorder.

Encouragingly, despite the low baseline level, the CLARA program produced substantial improvements in the primary outcomes of reading and language (pre-literacy skills and word reading) and oral language (vocabulary), as well as in secondary outcomes (narrative skills and reading comprehension), with effect sizes of a magnitude considered to be of "educational value" (Cohen's d >0.25) for reading, vocabulary and narrative language skills (What Works Clearing House, 2017). For the most part, these improvements were maintained. Gains in reading comprehension were partially mediated by progress in word reading and word knowledge (as measured by vocabulary) consistent with the Simple View of Reading (Hoover & Gough, 1990). Together the findings add to a body of evidence that multi-componential interventions can improve language and literacy skills of children at risk for low educational attainments (Snowling & Hulme, 2011). Specifically, we demonstrate in line with our logic model (Figure 1), that, given sensitive adaptation and appropriate training and support of local professionals, intervention programs developed in Western culture can be effectively applied in other settings, including vulnerable communities (Bekman et al., 2011).

Our findings support the importance of decoding and oral language skills as a foundation for reading comprehension; those children with the lowest level at entry made the biggest gains as these foundations developed. The direct effect of intervention remained significant after the

mediators of word reading and vocabulary were controlled. These findings are in line with a computer-based intervention in Central Chile where gains were only significant for those children with the lowest reading ability at entry (Ponce et al., 2012).

The CLARA program differs from previous Chilean intervention studies which have focused upon teacher training or targeted specific skills (Gomez Franco, 2014; Larraín, Strasser, & Lissi, 2012; Pallante & Kim, 2013; Ponce et al., 2012; Rosas et al., 2017; Strasser & Lissi, 2009; Strasser et al., 2018; Yoshikawa et al., 2015). The current program used resident tutors who were qualified speech and language therapists circumventing the need for the re-skilling of existing teachers or the incorporation of the program into the school day curriculum (Figure 1). Although this specific approach would be difficult to implement on a larger scale, it demonstrates the importance of consistent and regular intervention, identified as a problem in previous Chilean studies (Strasser & Lissi, 2009; Yoshikawa et al., 2015); our findings suggest it would be worth evaluating this program as an approach for use with small groups of children.

CLARA employed a multi-component approach combining training in letter-sound knowledge and phonological awareness (e.g. Catts et al., 2015), graded- and shared-book reading (e.g. Hatcher et al., 1994), narrative skills (e.g. Griffin et al., 2004; Justice et al., 2006), and progressive reading strategies for children who have developed competence with basic-reading skills (e.g. Duke et al., 2011; Hogan et al., 2011; Perfetti et al., 2005). While some of these components overlapped with previous Chilean interventions; Pallante and Kim (2015) targeted decoding, phonology, vocabulary and reading comprehension which were included in both the language and reading components of CLARA, Larrain, Strasser and Lissi (2012) focused upon shared reading sessions, which formed part of the language and reciprocal teaching components, Rosas *et al.* (2017) concentrated on decoding skills which were incorporated into the reading

sessions and Ponce *et al.* (2012) taught reading strategies which formed part of our language and reciprocal teaching components, the present study covered a greater number of skills across a wider age range. In general, any gains described in these former studies were related to direct outcomes such as decoding and vocabulary rather than reading comprehension. The generalizations observed in the current study illustrate the importance of considering theoretical frameworks and baseline skills in the design of intervention studies.

All materials used in CLARA were developed by the researchers and Chilean tutors and endorsed by teachers and community members. Materials that were not seen as engaging by the children were removed from use (see Figure 1). This aspect is likely to have contributed to the positive outcomes of CLARA and underline the need for the adaptation and contextualization of materials developed in one culture for use in another (Nag et al., 2014). In particular, we found that children preferred books over printed or electronic materials and favored content that was seen as relevant within island-life (e.g. nature-related texts). Although this required redundancy and flexibility within the program, it allowed a more personalized approach to the intervention as suggested by Aaron et al. (2008) and Francis *et al.* (2018). Furthermore, it permitted the use of a range of texts, which has previously been shown to provide optimal learning experience (van Bergen et al., 2018). Another important feature of the program is that the fidelity of delivery was measured throughout. Self-monitoring tools and mentorship were provided by the research team and may have promoted consistency, as suggested by Strasser and Lissi (2009) and Strasser *et al.* (2018).

Although significant gains were found within CLARA, improvements in expressive language (particularly narrative skills) and reading comprehension waned after the intervention ended. This lack of maintenance likely reflects the withdrawal of instruction, suggesting that

long-term support is needed to facilitate continued progression and generalization. As such, our findings suggest a dichotomy in terms of "learning" and "application"; maintenance of gains in pre-literacy, reading and vocabulary skills suggest that the intervention group had retained the taught information and basic skills were in place. However, gains in expressive narrative and reading comprehension were not maintained, suggesting that the gained knowledge was not effectively applied. One possibility is that it is hard to generalize vocabulary gains to wider language use or to reading comprehension (Wright & Cervetti, 2017). It is also possible that the summer break (more than 10 weeks between t3 and t4) contributed to diminished progression (Allington et al., 2010). An alternative interpretation is that gains were not maintained within the classroom curriculum, underlining the need for ongoing reinforcement of learnt skills in the classroom, as voiced by Strasser et al. (2018). Speculatively, a reduction in active one-to-one encouragement from the tutors may lead to regression if generalization strategies are not fully embedded into classroom practice, and may be magnified where the propensity for literacy problems is high.

This study has limitations that should be noted. The sample size was constrained by the size of population and family's willingness to sign up. Our study has relatively low power and it is clear that further replication studies are required. Second, the program was delivered to children spanning a wide age range (4 to 14 years); while not ideal methodologically, because of the individualized protocols, this did not pose difficulties of implementation. Finally, the costs were considerable: trained professionals delivered the program on a daily basis, a situation that is unusual in remote settings. The study also confronted issues during implementation including access to the island, frequent changes of school staff and management and varying degrees of engagement. These factors are part of the reality of working with a vulnerable community and

need to be considered within any future intervention designs. Nonetheless, the positive gains associated with this exploratory trial offer 'proof in principle' and indicate that it would be worthwhile to adapt the program into a more sustainable package and to evaluate its effectiveness in a larger better powered trial.

Notwithstanding these problems, this study found support for the Simple View of Reading as a basis for the design of multi-component intervention programs in vulnerable populations and developing countries. The study makes a distinctive contribution to the understanding of effective interventions for Spanish-speakers and generates implications for how theoretical frameworks can be applied in vulnerable communities. Findings add to evidence that programs with a focus on pre-literacy skills, oral language and reading strategies can produce educationally significant improvements. If sustained, the program has the potential to address the fourth of the Sustainable Development Goals and close the academic gap frequently observed between children in developing countries and those in more advantaged settings.

Acknowledgments

This research was funded by ESRC:CONICYT Newton Research Partnership: "A language and reading intervention program for Chile, piloted in the Robinson Crusoe Island" (ES/N01913X/1, PI Dianne Newbury). Ethical approval was granted by Oxford University Research Ethics Committee (R42391/RE001).

The views presented in this work do not represent those of the federal government, nor do they endorse any products or findings presented herein.

We are grateful to the Municipality, school, teachers, children and families who participated in the study, Loreto Aliaga-Salas for help with intervention and project materials, and Denise Cripps for help with logistics. We thank GL Assessment for permission to translate the York Assessment of Reading for Comprehension (YARC) into Spanish. We are grateful to Global Education Systems for giving permission to translate the Connectors Reciprocal Reading books into Spanish.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Aaron, P., Joshi, R. M., Gooden, R., & 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. *Journal of Learning disabilities, 41*(1), 67-84. doi:10.1177/0022219407310838

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., . . . Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*(5), 411-427. doi:10.1080/02702711.2010.505165

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251-271. doi:10.1086/511706

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading research quarterly, 26*(3), 251-276. doi:10.2307/747763

Bekman, S., Aksu-Koç, A., & Erguvanlı-Taylan, E. (2011). Effectiveness of an intervention program for six-year-olds: a summer-school model. *European Early Childhood Education Research Journal, 19*(4), 409-431. doi:10.1080/1350293X.2011.623508

Brigstocke, S., Moll, K., & Hulme, C. (2016). *Test of Basic Arithmetic & Numeracy Skills*. Oxford, UK: Oxford University Press.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*(5-6), 489-503. doi:10.1023/A:1008084120205

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., . . . Mikulajová, M. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological science, 23*(6), 678-686. doi:10.1177/0956797611434536

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5-51. doi:10.1177/1529100618772271

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing, 28*(9), 1407-1425. doi:10.1007/s11145-015-9576-x

Chapman, R., & Miller, J. (1984). *SALT: Systematic analysis of language transcripts*. Wisconsin, USA: University of Wisconsin Press.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological science, 21*(8), 1106-1116. doi:10.1177/0956797610375449

Clay, M. M. (1985). *The early detection of reading difficulties*. Portsmouth, NH: Heinemann Educational Books Inc.

Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review, 25*, 56-72. doi:10.1016/j.edurev.2017.12.001

De Barbieri, Z., Fernandez, M. A., Newbury, D. F., & Villanueva, P. (2018). Family aggregation of language impairment in an isolated Chilean population from Robinson Crusoe Island. *Int J Lang Commun Disord, 53*(3), 643-655. doi:10.1111/1460-6984.12377

Duke, N. K., Pearson, P. D., Strachan, S. L., & Billman, A. K. (2011). Essential elements of fostering and teaching reading comprehension. *What research has to say about reading instruction, 4*, 286-314. doi:10.1598/0829.03

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental neuropsychology, 24*(2-3), 613-639. doi:10.1080/87565641.2003.9651913

Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the Simple View of Reading to Account for Variation Within Readers and Across Texts: The Complete View of Reading (CVR i). *Remedial and special education, 39*(5), 274-288. doi:10.1177/0741932518772904

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280-290. doi:10.1111/jcpp.12010

Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., . . . Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: a randomized controlled trial. *Journal of Child Psychology and Psychiatry, 58*(10), 1141-1151. doi:10.1111/jcpp.12737

Gomez Franco, L. E. (2014). *Exploring teachers' read-aloud practices as predictors of children's language skills: the case of low-income Chilean preschool classrooms.* Boston College. Lynch School of Education,

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education, 7*(1), 6-10. doi:10.1177/074193258600700104

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First language, 24*(2), 123-147. doi:10.1177/0142723704042369

Gutierrez-Clellen, V. F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech, Language, and Hearing Research, 37*(3), 645-654. doi:10.1044/jshr.3703.645

Guzmán, M. M. P., Tirapegui, C. J. C., & Landaeta, M. M. (2012). *El Desarrollo narrativo en niños: una propuesta práctica para la evaluación y la intervención en niños con trastorno del lenguaje*. Barcelona, Spain: Ars Médica.

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child development, 65*(1), 41-57. doi:10.2307/1131364

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*(2), 154-166. doi:10.1044/1058-0360(2009/08-0024)

Hogan, T., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children, 44*(3), 1-20. doi:10.17161/fec.v44i3.6688

Høien-Tengesdal, I., & Høien, T. (2012). The reading efficiency model: An extension of the componential model of reading. *Journal of Learning disabilities, 45*(5), 467-479. doi:10.1177/0022219411432688

Høien-Tengesdal, I. (2010). Is the simple view of reading too simple? *Scandinavian Journal of Educational Research, 54*(5), 451-469. doi:10.1080/00313831.2010.508914

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160. doi:10.1007/BF00401799

Hulme, C., Nash, H., Gooch, D., Lervag, A., & Snowling, M. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychol Sci, 26*(12), 1877-1886. doi:10.1177/0956797615603702

Hulme, C., Zhou, L., Tong, X., Lervåg, A., & Burgoyne, K. (2019). Learning to read in Chinese: Evidence for reciprocal relationships between word reading and oral language skills. *Developmental science, 22*(1), e12745. doi:doi: 10.1111/desc.12745

Joshi, R. M., Tao, S., Aaron, P., & Quiroz, B. (2012). Cognitive component of componential model of reading applied to different orthographies. *Journal of Learning disabilities, 45*(5), 480-486. doi:10.1177/0022219411432690

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology, 15*(2), 177-191. doi:10.1044/1058-0360(2006/017)

LAC Reads Capacity Program. (2016). *Early Grade Reading in Latin America and the Caribbean: A Systematic Review*. Retrieved from

Larraín, A., Strasser, K., & Lissi, M.-R. (2012). Lectura compartida de cuentos y aprendizaje de vocabulario en edad preescolar: un estudio de eficacia. *Estudios de Psicología, 33*(3), 379-383. doi:10.1174/021093912803758165

Lervåg, A., Dolean, D., Tincas, I., & Melby-Lervåg, M. (2019). Socioeconomic Background, Nonverbal IQ, and School Absence Affects the Development of Vocabulary and Reading Comprehension in Children Living in Severe Poverty. *Developmental science*, e12858. doi:10.1111/desc.12858

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child development, 89*(5), 1821-1838. doi:10.1111/cdev.12861

Leyva, D., & Skorb, L. (2017). Food for thought: Family food routines and literacy in Latino kindergarteners. *Journal of Applied Developmental Psychology, 52*, 80-90. doi:10.1016/j.appdev.2017.07.001

Martin, N. A., & Brownell, R. (2012). *Receptive One-Word Picture Vocabulary Test-4 (ROWPVT-4)*. Novato, CA: Academic Therapy Publications.

McCardle, P., & Chhabra, V. (2004). *The voice of evidence in reading research*. Baltimore, MD, US: Paul H Brookes Publishing Co.

Megherbi, H., Seigneuric, A., & Ehrlich, M.-F. (2006). Reading comprehension in French 1st and 2nd grade children: Contribution of decoding and language comprehension. *European Journal of Psychology of Education, 21*(2), 135. doi:10.1007/BF03173573

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology, 108*(1), 130. doi:10.1037/edu0000047

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational research methods, 11*(2), 364-386. doi:10.1177/1094428106291059

Muthén, L., & Muthén, B. (2012). *Mplus - The comprehensive modelling program for applied researchers: User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Nag, S., Chiat, S., Torgerson, C., & Snowling, M. J. (2014). *Literacy, foundation learning and assessment in developing countries*. Retrieved from UK: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/305150/Literacy-foundation-learning-assessment.pdf

Nag, S., Snowling, M. J., & Asfaha, Y. M. (2016). Classroom literacy practices in low-and middle-income countries: an interpretative synthesis of ethnographic studies. *Oxford Review of Education, 42*(1), 36-54. doi:10.1080/03054985.2015.1135115

Nag, S., Vagh, S. B., Dulay, K. M., & Snowling, M. J. (2019). Home language, school language and children's literacy attainments: A systematic review of evidence from low-and middle-income countries. *Review of Education, 7*(1), 91-150. doi:10.1002/rev3.3130

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved from USA: https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Opel, A., Ameer, S. S., & Aboud, F. E. (2009). The effect of preschool dialogic reading on vocabulary among rural Bangladeshi children. *International Journal of Educational Research, 48*(1), 12-20. doi:doi.org/10.1016/j.ijer.2009.02.008

Pallante, D. H., & Kim, Y.-S. (2013). The effect of a multicomponent literacy instruction model on literacy growth for kindergartners and first-grade students in Chile. *International Journal of Psychology, 48*(5), 747-761. doi:10.1080/00207594.2012.719628

Pavez, M. M., Coloma, C., & Maggiolo, M. (2008). *El desarrollo narrativo en niños: EDNA*. Barcelona: Ars Medica.

Pavez, M. M., Coloma, C., Maggiolo, M., Martínez, L., & Romero, L. (2002). *Procedimientos para evaluar discurso. PREDI*. Santiago de Chile: Ediciones Universidad Católica de Chile.

Peralta, M. V. (2011). Early childhood education and public care policies in Chile: A historical perspective to analyze the present. *International Journal of Child Care and Education Policy, 5*(1), 17.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Malden, MA: Blackwell Publishing Ltd.

Ponce, H. R., López, M. J., & Mayer, R. E. (2012). Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies. *Computers & Education, 59*(4), 1170-1183. doi:10.1016/j.compedu.2012.05.013

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford pyschologists Press.

Reutzel, D. R., Smith, J. A., & Fawson, P. C. (2005). An evaluation of two approaches for teaching reading comprehension strategies in the primary years using science information texts. *Early Childhood Research Quarterly, 20*(3), 276-305. doi:10.1016/j.ecresq.2005.07.002

Rosas, R., Escobar, J.-P., Ramírez, M.-P., Meneses, A., & Guajardo, A. (2017). Impact of a computer-based intervention in Chilean children at risk of manifesting reading difficulties/Impacto de una intervención basada en ordenador en niños chilenos con

riesgo de manifestar dificultades lectoras. *Infancia y Aprendizaje, 40*(1), 158-188. doi:10.1080/02103702.2016.1263451

Scarborough, H. (2001). Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook of Early Literacy Research, Volume 1* (pp. 23-38). New York: Guildford Press.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine, 8*(1), 18. doi:10.1186/1741-7015-8-18

Semel, E. M., Wiig, E. H., Secord, W., & Langdon, H. W. (2006). *CELF 4: Clinical Evaluation of Language Fundamentals 4: Spanish Edition*. San Antonio, TX: PsychCorp.

Sénéchal, M., & LeFevre, J.-A. (2001). Storybook reading and parent teaching: Links to language and literacy development. In *The role of family literacy environments in promoting young children's emerging literacy skills.* (pp. 39-52). San Francisco, CA, US: Jossey-Bass.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*(4), 563-575. doi:10.1007/BF02296196

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81*(1), 1-23. doi:10.1111/j.2044-8279.2010.02014.x

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2009). *YARC York Assessment of Reading for Comprehension Passage Reading*. London, UK: GL Publishers.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of educational research, 56*(1), 72-110. doi:10.3102/00346543056001072

Strasser, K., & Lissi, M. R. (2009). Home and Instruction Effects on Emergent Literacy in a Sample of Chilean Kindergarten Children. *Scientific Studies of Reading, 13*(2), 175-204. doi:10.1080/10888430902769525

Strasser, K., Mendive, S., Vergara, D., & Darricades, M. (2018). Efficacy of a self-monitoring tool for improving the quality of the language environment in the preschool classroom. *Early Education and Development, 29*(1), 104-124. doi:10.1080/10409289.2017.1287992

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing, 28*(7), 939-957. doi:10.1007/s11145-015-9556-1

Torppa, M., Georgiou, G. K., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly (1982-), 62*(2), 179-206. doi:10.13110/merrpalmquar1982.62.2.0179

van Bergen, E., Snowling, M. J., de Zeeuw, E. L., van Beijsterveldt, C. E., Dolan, C. V., & Boomsma, D. I. (2018). Why do children read more? The influence of reading ability on voluntary reading practices. *Journal of Child Psychology and Psychiatry, 59*(11), 1205-1214. doi:10.1111/jcpp.12910

Wechsler, D. (2003). *Wechsler intelligence scale for children–Fourth Edition (WISC-IV)*. San Antonio, TX: The Psychological Corporation.

What Works Clearing House. (2017). *Standards handbook (Version 4.0)*. Retrieved from Princeton, NJ: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child development, 69*(3), 848-872. doi:10.1111/j.1467-8624.1998.tb06247.x

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading research quarterly, 52*(2), 203-226. doi:10.1002/rrq.163

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., . . . D'Sa, N. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental psychology, 51*(3), 309. doi:10.1037/a0038785

Table 1. Descriptive statistics for the intervention and waiting control groups

Measure	Time Point[a]	Waiting control group		Intervention group		Effect Size[b]	Reliability[c]
		M	*SD*	M	*SD*		
Letter knowledge	Time 1	72.9	38.9	73.9	36.5	0.61††	0.99
(ELDEL; Caravolas et al., 2012)	Time 2	77.1	37.4	94.6	28.3		0.99
	Time 3	91.7	34.4	116	22.2		0.99
	Time 4	118.6	13.1	118.8	8.6		0.94
Phoneme awareness	Time 1	16.5	11.9	17.7	11.8	0.85†††	0.96
(ELDEL; Caravolas et al., 2012)	Time 2	39.1	19.5	43.4	18.4		0.95
	Time 3	41.1	19.0	52.6	10.5		0.94
	Time 4	55.5	8.2	54.8	9.2		0.88
Picture Word Matching Reading Test	Time 1	24.0	20.5	27.3	21.0	0.08	0.66
(ELDEL; Caravolas et al., 2012)	Time 2	28.5	22.3	33.5	21.1		0.86
	Time 3	34.0	21.1	39.1	20.1		0.99
One-minute word reading test	Time 1	51.1	40.3	50.4	40.4	0.33†	0.72
(ELDEL; Caravolas et al., 2012)	Time 2	56.1	43.8	63.6	40.8		0.99
	Time 3	61.8	44.5	74.1	37.7		0.99
	Time 4	73.9	36.2	82.5	35.7		0.99
Nonword Reading	Time 1	28.7	23.5	31.1	24.2	0.20†	0.81
(ELDEL; Caravolas et al., 2012)	Time 2	33.7	25.9	39.8	24.4		0.79
	Time 3	37.2	23.7	44.5	21.3		0.99
	Time 4	43.5	19.7	50.9	21.3		0.97
Spelling	Time 1	7.7	5.6	8.9	5.6	0.18	0.93
(ELDEL; Caravolas et al., 2012)	Time 3	15.3	10.0	17.6	9.8		0.91
	Time 4	17.9	9.1	19.8	8.6		0.92
CELF vocabulary	Time 1	42.6	10.3	41.0	12.9	0.52††	0.79
(Semel et al., 2006)	Time 3	54.3	10.1	58.9	11.7		0.88
	Time 4	59.2	8.1	57.4	11.4		0.86
WISC vocabulary	Time 1	16.6	8.9	16.7	8.35	0.23†	0.79
(translated fromWechsler, 2003)	Time 3	17.5	10.0	19.6	9.7		0.86
	Time 4	21.9	10.6	22.3	11.6		0.73
Receptive vocabulary (ROWPVT; translated	Time 1	80.0	14.8	80.3	16.1	0.007	0.76
from Martin & Brownell, 2012)	Time 3	86.3	10.5	86.7	10.1		0.84
Bespoke Vocabulary	Time 1	9.9	5.4	11.4	5.2	0.57††	0.82
	Time 3	13.0	6.9	17.4	8.3		0.87
	Time 4	18.0	6.7	17.4	8.5		0.84
Listening Comprehension	Time 1	10.6	2.9	11.2	3.4	-0.06	0.70
(EDNA; Pavez et al., 2008)	Time 3	12.2	2.3	12.7	2.0		0.43
	Time 4	14.5	2.6	14.2	1.8		0.45
MLU-w	Time 1	6.7	1.8	7.2	1.9	0.54††	-
(collected from PREDI; Pavez et al., 2002)	Time 3	6.8	1.1	8.4	1.7		-
	Time 4	7.5	1.5	7.2	1.5		-
Subordination Index	Time 1	1.2	0.2	1.3	0.2	1.20†††	-
(collected from PREDI; Pavez et al., 2002)	Time 3	1.1	0.2	1.5	0.3		-
	Time 4	1.3	0.2	1.6	0.2		-
Story Grammar	Time 1	6.5	2.5	6.2	2.3	-0.04	-
(collected from PREDI; Pavez et al., 2002)	Time 3	7.2	2.3	6.7	2.5		-
	Time 4	6.9	1.9	6.9	2.1		-
Reading Comprehension	Time 1	13.4	12.2	13.8	11.6	0.51††	0.86
(YARC; translated from Snowling et al., 2009)	Time 3	16.2	12.3	22.8	12.9		0.83
	Time 4	15.1	11.5	15.2	12.4		0.84
Non-verbal IQ (RAVEN; Raven & Court, 1998)	Time 1	24.2	7.9	24.9	8.1		-
Recalling Sentences (Semel et al., 2006)	Time 1	58.7	21.4	57.0	19.0	0.18	0.91
Arithmetic (TOBANS; Brigstocke et al., 2016)	Time 1	46	19.7	40.6	22.9		-

Where possible, standardized scores are reported. Where standardized tests were not available (see methods section 2.2), raw scores are reported instead.

[a] Data from t4 should be interpreted with caution as the assessors were no longer blind to group allocation and the waiting control group had received seven fewer weeks than the intervention group due to funding constraints and time restrictions.

[b] Effect sizes were calculated using Cohen's *d*; the difference in gains (t1 to t3) between groups divided by the pooled pre-test standard deviation (see Morris, 2008). Cohen's d values are † = small, †† = medium, and ††† = large.

[c] Internal consistency was estimated using Cronbach's alpha.

Figure Legends

Figure 1. Logic Model for the Exploratory Trial.

Figure 2. Consort Diagram Showing Participant Flow.

Diagram follows the updated guidelines for reporting Randomized Control Trials (Schulz et al., 2010).

Figure 3. Mediation Model Showing the Impact of the Intervention on Reading Comprehension.

Indirect effects of intervention on reading comprehension via word reading and vocabulary are significant. The model shows a good fit to the data: $\chi 2 = 7.34$, $p=0.39$, RMSEA=0.027 [0.00, 0.158], CFI=0.999, SRMR=0.021.

Figure 4. Maintained group trajectories.

(a) Phoneme awareness, (b) Letter Knowledge, (c) One-minute word reading test, (d) Nonword reading, (e) CELF vocabulary, (f) Bespoke vocabulary

Mean scores at pretest (t1), posttest (t3) and delayed posttest (t4) for the intervention and waiting control groups. Error bars represent 95% Confidence Intervals. Data from t4 should be interpreted with caution as the assessors were no longer blind to group allocation and the waiting control group had received seven fewer weeks than the intervention group due to funding constraints and time restrictions.