Higher-Level Language Strategy-Based Intervention for Poor Comprehenders: A pilot single case experimental design

Katrina Kelso, Anne Whitworth, Suze Leitão
Curtin University

Corresponding author:
Katrina Kelso M Ed in Applied Linguistics (Honours)
Curtin School of Allied Health
Curtin University
GPO Box U1987, Perth
Western Australia, 6845
Email: katrina.kelso@postgrad.curtin.edu.au

Email addresses for all authors
Anne Whitworth PhD anne.whitworth@curtin.edu.au
Suze Leitão PhD S.Leitao@exchange.curtin.edu.au

Acknowledgements
We would like to thank the schools, children and their families, who participated in this study.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

This paper was accepted for publication in Child Language Teacher and Therapy on January 11, 2022
Please cite as
Kelso, K., Whitworth A., & Leitão, S. (2022a). Higher-level language strategy-based intervention for poor comprehenders: A pilot single case experimental design. *Child Language, Teaching and Therapy*. Copyright ©2021 The Author(s). pp.1-15.
https://doi.org/10.1177/02656590211071003

Abstract

In contrast to the large body of research investigating intervention for poor decoding skills, far fewer studies have evaluated interventions for reading comprehension. There is even less research on children with more specific difficulties with reading comprehension, often referred to as "poor comprehenders". Levels of effectiveness have varied for interventions targeting lower- and higher-level language, including inference making, on trained measures, with little transfer to generalised reading comprehension measures in both skilled and less-skilled readers. Outcomes have been more positive for poor comprehenders, however findings have been inconsistent as to which programme components have led to gains in reading comprehension. This pilot study utilised a case series design to explore whether a novel intervention targeting oral inference making and comprehension monitoring was effective in improving the targeted skills and reading comprehension of 11 children, aged 9;2 – 12;3 years, with average-for-age phonological and lower-level language skills but weak inferencing. All participants improved on the primary inference subtest post-intervention and continued to score higher at maintenance than at pre-intervention. Results on the remaining higher-level language tasks were more varied, as were the results for reading comprehension, with fewer participants demonstrating generalisation to these tasks, particularly the nonfiction texts. While the results are preliminary and descriptive, they suggest that improvements can be made in higher-level language in a 10-session intervention, and provide directions for future research.

Keywords: reading comprehension, poor comprehenders, intervention, inference making, comprehension monitoring

I Introduction

A subgroup of poor readers with more specific difficulties with reading comprehension, often referred to as "poor comprehenders" in the literature, have marked deficits in comprehending what they read despite being able to read aloud accurately and fluently at an age-appropriate level. The reported prevalence of poor comprehenders has varied over the years depending on the tests and cut-off criteria, ranging from 10-15% based on comprehension being at least 6 months below age-appropriate text reading accuracy in earlier studies (Yuill and Oakhill, 1991), to 3.3% in more recent studies using strict criteria to identify children with clinically significant reading comprehension difficulties (Hulme and Snowling, 2011). The current consensus is that approximately 7% of middle primary students

can be classified as poor comprehenders (e.g., Clarke et al., 2010; Elwér et al., 2015), with this number increasing across the school years (e.g., Catts et al., 2005).

There is now a sizeable body of research exploring the language and cognitive profiles of poor comprehenders, with many studies identifying intact phonological skills which support age-appropriate word reading accuracy and fluency, but deficits on a range of oral comprehension skills (e.g., Adlof and Catts, 2015; Nation et al., 2010). Two alternative hypotheses as to the source of poor reading comprehension have been proposed (Snowling & Hulme, 2011). The first purports difficulty with lower-level language (vocabulary, morphology and grammar) required for understanding sentences so the basic meaning can be extracted from the text (e.g., Nation, 2005 for a review; Nation et al., 2010). Retrospective longitudinal studies have found these oral language weaknesses are evident before children learn to read (e.g., Catts et al., 2006; Elwér et al., 2015; Nation et al., 2010). A second hypothesis, highlighted by Oakhill and colleagues, proposes difficulty with higher-level language skills required to form a coherent mental model of a text's meaning (inferencing, understanding of text structure and comprehension monitoring - see Cain, 2010 for a review).

1 Comprehension Intervention

Research into classroom reading comprehension instruction is limited, as are studies on effective intervention tailored to the individual child's needs. This is in marked contrast to research into remediation of phonological and decoding difficulties. Reviewing the research on instruction for typically developing children, the National Reading Panel (NRP: NICHD, 2000) found support for the effectiveness of (1) specific vocabulary instruction and (2) eight types of comprehension strategy instruction. The NRP also found that when teachers were taught to use comprehension instruction methods effectively, students' awareness and use of strategies increased, which in turn led to improvements in reading comprehension.

A systematic review of 43 studies with a control group and a pre- to post-test design, evaluated the effectiveness of oral language instruction in cohorts of preschool and school age participants (Rogde et al., 2019). Small immediate and follow-up effects were identified on generalised tests (i.e., tests that did not include items trained in the intervention) of lower-level vocabulary and grammatical knowledge, and moderate effects on narrative and listening comprehension. There were no immediate effects, however, on generalised tests of reading comprehension, and equivocal findings in the few studies that reported follow-up.

The higher-level skill of inference generation is prominent in models of reading comprehension (see McNamara and Magliano, 2009 for a review). There is empirical support

for techniques to develop the higher-level language skills of inferencing, comprehension monitoring and text structure knowledge in children from preschool to grade 3, with and without oral and/or reading difficulties (Hogan et al., 2011). A synthesis of nine studies (Hall, 2016) and a meta-analysis of 25 studies (Elleman, 2017), examined the impact of inference instruction on reading comprehension in primary and secondary school students. Significant effect sizes were found on experimental measures of inferential reading comprehension in each of the studies involving struggling readers in Hall's (2016) synthesis, and for some studies on standardised measures. Inference instruction was effective in increasing both skilled and less skilled readers' general and inferential comprehension (Elleman, 2017). Less-skilled readers not only improved substantially in forming accurate inferences, but also in literal comprehension of text; however, few studies used standardised or generalised reading comprehension outcome measures. Two recent inference instruction studies, one with Dutch participants aged 8-12 years in mainstream school (Bos et al., 2016) and the other in the USA targeting 12-15 year old struggling readers (Barth and Elleman, 2017), reported improvements on both trained items and on generalised reading comprehension measures.

2 Intervention for Poor Comprehenders

To date, few studies have evaluated interventions for poor comprehenders (see Snowling and Hulme, 2011, for a summary). Oakhill and colleagues investigated the effect of training on drawing inferences, comprehension monitoring, use of text organisers and mental imagery on text comprehension (Yuill and Oakhill, 1991). The first three approaches were effective in improving comprehension outcomes for poor comprehenders aged 7-8 years compared with controls, while mental imagery was effective with participants aged 9-10 years. Inference training showed the greatest effect on both study outcome measures and a standardised reading comprehension test. This finding was replicated by McGee and Johnson (2003) using a standardised measure in 6-10 year old poor comprehenders.

Clarke et al. (2010) evaluated the effectiveness of three training programmes for poor comprehenders aged 8-9 years in a randomised control trial. The programmes, each delivering 30 hours of intervention per child over 20 weeks, were driven by the two hypotheses as to the source of the reading comprehension difficulties in poor comprehenders (Snowling & Hulme, 2011). One programme focused on training the oral lower- and higher-level language skills of vocabulary, listening comprehension, figurative language and spoken narrative. The outcomes were compared to those of two other programmes, one of which focused on developing text comprehension using written texts, comprising training in

metacognitive strategies, reading comprehension, inferencing from text and written narrative, while the third integrated components of both programmes (Clarke et al., 2010). All groups made significant gains, compared with a waiting control group, on a reading comprehension test immediately after training. Gains continued to be significant for all three intervention groups on follow-up testing 11 months later, with those receiving the oral language training showing the most improvement (Clarke et al., 2010). As the programmes contained multiple components, however, it was not clear which components led to the gains in reading comprehension (Snowling and Hulme, 2011). Within the constraints of clinical and educational practice, it is important to select the most effective and efficient intervention approach. This study, therefore, aimed to explore the effect of a pilot programme, specifically designed to target higher-level language skills, on reading comprehension in a case series with 11 participants who presented with a profile of higher-level language difficulties (as reported in Kelso et al., 2021b).

The aims of this study were to

i. explore whether a pilot programme utilising a novel intervention designed to target higher-level language skills was effective in improving oral inference making and comprehension monitoring skills, and
ii. investigate generalisation to tests of reading comprehension.

II Method

The participants recruited to this study were part of a larger research programme investigating the identification, profiling, and subsequent intervention with individual poor comprehenders. Ethical approval was granted by Curtin University Human Research Ethics Committee (HRE2016-0438-01) and the Government of Western Australia Department of Education. Written consent was obtained from participating school principals, teachers, students, and their parents/guardians.

1 Participants

Initial testing using a phonological awareness and a listening comprehension task was carried out with 218 children in School Years 3–6, aged 7;8 –12;1 years. The children attended one of two local primary schools serving predominantly middle and upper-middle class catchment areas in inner metropolitan Perth, Western Australia. Of these, 24 children were confirmed as poor comprehenders using measures of nonword reading (i.e., the ability to decode nonsense words), reading comprehension and nonverbal IQ (Kelso et al., 2020). These children scored in the average range on the nonword reading accuracy and nonverbal

IQ tasks but were classified as poor comprehenders on a reading comprehension task, the York Assessment of Reading Comprehension Primary – Australian Edition – Form A (Snowling et al., 2012), using cut-off scores as described in Kelso et al. (2020).

Seventeen of the 24 children (eight boys and nine girls, aged 8;8 – 11;9 years) subsequently completed a full assessment battery. This was comprised mainly of standardised tests for the purpose of profiling the oral and written language input skills within the Sublexical, Lexicon and Comprehension Processes component systems of Perfetti and Stafura's (2014) Reading Systems Framework (RSF), together with assessments of verbal memory. Of these 17 participants, the profile for 15 indicated difficulty with inferencing skills within the higher-level language Comprehension Processes component of the RSF, but no difficulty on the majority of vocabulary and grammar lower-level language tasks within the Lexicon and Sentence levels of the RSF, along with intact Sublexical component skills. The remaining two participants also had intact Sublexical component skills but had difficulty with Lexicon component vocabulary tasks, as well as at each of the three levels of the Comprehension Processes component of the RSF (Kelso et al., 2021b).

Table 1

Demographic and background information

P	Gender	School Year	Age years;months	Language Background
1	Female	4	9;9	Bilingual
2	Female	6	10;11	Bilingual
3	Male	4	9;2	Monolingual
4	Male	5	10;2	Bilingual
5	Male	5	10;6	Monolingual
6	Female	6	11;9	Monolingual
7	Male	7	12;3	Monolingual
8	Male	5	10;3	Monolingual
9	Male	5	10;10	Monolingual
10	Male	5	10;5	Monolingual
11	Female	6	10;11	Monolingual

Notes. P = participant

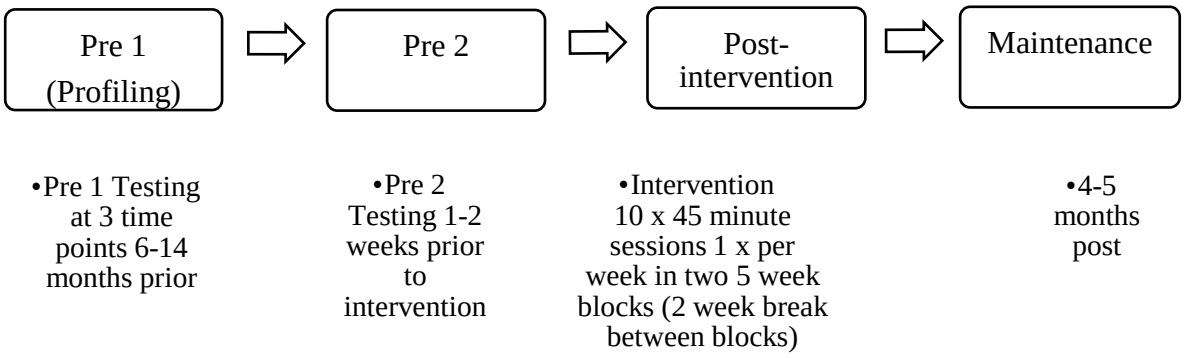
The 15 children identified with difficulty inferencing were invited to participate in a targeted 10-session intervention. Eleven children (aged 9;2 – 12;3 years) and their parents/ caregivers agreed to participate. Demographic information is set out in Table 1. Three participants were bilingual (P1, P2, P4), with P1 and P2 having been exposed to English since birth and P4 for five years.

2 Procedure

The intervention consisted of 10 individual 45 minute sessions, one session per week, in two blocks of five sessions with a two week school holiday break between the blocks. The intervention was carried out by the first author in a room at the child's school or in clinic rooms. A case series design was implemented, replicated across each participant, and follows the SCRIBE reporting guidelines (Tate et al., 2016).

Figure 1

Testing Schedule



All measures, apart from the comprehension monitoring task, were administered at initial testing (Pre 1), pre-intervention (Pre 2) to facilitate some monitoring of any pre-intervention spontaneous change, post-intervention and maintenance. Data collection for initial testing (Pre 1) took place on three occasions, with phase 1 oral tasks carried out an average of 14 months prior to the intervention, confirmation reading testing 9-12 months prior (see Kelso et al., 2020), and detailed profiling testing completed an average 6 months prior to the commencement of the intervention (see Kelso et al., 2021b). Pre 2 testing was carried out 1-2 weeks prior to the intervention, post-intervention testing approximately four months after Pre 2, and maintenance testing four-five months later (see Figure 1). Alternate versions of the reading tests were used at different time points; however, these were not available for the oral language or comprehension monitoring tasks.

3 Measures

a Oral Language Tasks

Clinical Evaluation of Language Fundamentals-4 (CELF-4 Australian Edition)

Understanding Spoken Paragraphs. (Semel et al., 2006). This subtest measures ability to listen to three short paragraphs and answer five factual and inferential questions for each. Percentage of inferential questions (main idea, inference, prediction) correct is reported. The subtest has adequate internal consistency reliability (r = *.70*), test-retest reliability (r = .76), and appropriate construct validity.

Test of Problem Solving-3 Elementary (TOPS-3E). (Bowers et al., 2005). The TOPS-3E assesses critical thinking abilities, with test questions focusing on language-based thinking skills including clarifying, analysing, generating solutions, evaluating, and affective thinking. The total test has appropriate test-retest reliability (r = .84) and inter-rate agreement of 89%. Test-retest reliability for individual subtests reported in this study is: Making Inferences (r = .79), Problem Solving (r = .75) and Predicting (r = .62).

b Comprehension Monitoring Tasks

Two tasks (Jane Oakhill, 2016, personal communication) were used to assess comprehension monitoring based on age, one for participants aged 8-9 years (in School Year 4) and one for older participants. The task for younger participants consists of seven short texts, five of which include one piece of inconsistent information, separated in the text. Older participants are presented five stories, two of which make sense and three containing two or three inconsistent pairs of sentences, separated in the text, resulting in seven inconsistencies altogether. Following a demonstration, participants were asked to read each text silently, decide if the text made sense and, if not, underline the pairs of inconsistent information. Percentage of inconsistencies identified at each time point is reported.

c Reading Comprehension Tests

York Assessment of Reading for Comprehension (YARC – Australian Edition). The Primary (Snowling et al., 2012) or Secondary (Stothard et al., 2012) version was administered depending on participants' School Year. Each version (Form A and B) of the tests has appropriate reliability and content validity. Comprehension scores are reported as a standardised reading outcome measure, along with percentage of questions requiring an inference correct.

PROBE-2. (Parkin, and Parkin, 2011). PROBE-2 is a criterion-referenced assessment of reading comprehension for children from 7 years through to adults, drawing on 20 sets of

graded texts (for decoding-age levels 5;0 – 15;6 years). Each set, comprised of a fiction and nonfiction text, covers an age span of 12 months, with consecutive sets overlapping by 6 months. The pre-intervention texts for the youngest participant (decoding-age 8;6-9;6 years) have 8 comprehension questions (7 inferential) and those for decoding-age 9 years and above have 10 questions (9 inferential). The criteria for determining a reading age are achieving a minimum of 96% decoding accuracy and 70% comprehension, with only the comprehension percentage being reported in this study. Participants' texts were selected at each test-time based on their chronological age being approximately the middle of the 12 month reading-age range.

4 Intervention Structure

The structure of the intervention is set out in Table 2. All students received the intervention as per the available protocol. The first five sessions revolved around three aims: (a) highlighting awareness of the role of inference-making in text comprehension, (b) introducing inference types using the Question Answer Relationship (QAR) framework (Raphael et al., 2006), and (c) teaching strategies for use throughout the Reading Cycle. A procedure based on the Transactional Strategies Instruction approach (in Klingner et al., 2015), utilising a think-aloud "I do/We do/You do" (Fisher and Frey, 2013) process, was used to introduce strategies. Questions were asked at the start and end of each session to check recall of what had been taught, and a sheet summarising what had been covered was provided after these five sessions. The second five sessions aimed to practise application of the strategies taught using longer fiction and non-fiction texts. A copy of all handouts was given to each participant in the last session. The intervention programme session plans and summary sheets are available to download free from: https://www.dropbox.com/sh/vn2was2q3yp2kld/AACK9FDWn4rDihnxG0DHItlua?dl=0

III Results

To address the first aim, we present the oral inferential comprehension and comprehension monitoring outcomes, followed by the reading comprehension measures to explore the second aim of generalisation. Data for each task at each time point is reported, along with the change in score, Pre 1 to Pre 2, Pre 2 to post-intervention, and post-intervention to maintenance. Observed and clinically significant changes, such as standard score increases of $\geq 1SD$, or crossing of a clinical boundary, are also reported.

Table 2

Intervention session outline.

Session	Target/Strategy	References/ Resources
1	Importance of inference making Text vs Knowledge Based inferences Being a Reading Detective Visualising	Bos et al. (2016); Core QARs - Raphael et al. (2006); Shanahan et al. (2010); Visualizing and Verbalizing Stories Book 1 (VV1) Level 3 short texts (Bell 2007)
2	Visualising Predicting	VV1 Level 3 short texts
3	Predicting, Visualising Comprehension Monitoring – Click vs Clunk	Klingner et al. (2015); VV1 Level 3 and 4 short texts
4	Expanded QAR sources of information Comprehension Monitoring Fix-Up strategies for Clunks Reading Cycle	Raphael et al. (2006); VV1 Level 4 short text; Lubliner (2005); Klingner et al. (2015)
5	QAR practise Summarising	Raphael et al. (2006); VV1 Level 4 short text; Lubliner (2005); Klingner et al. (2015)
6 - 10	Practise making inferences and applying strategies taught in longer texts using "Boxing Up" QAR Graphic Organiser	Key Into Inference fiction and nonfiction texts (Parkin et al. (2002) Comprehend It texts (Tuffin and Henderson, 1993) Pardo et al. (2011)

1 Oral Inferential Comprehension

On the Making Inferences subtest of the TOPS-3E, seven participants had an increased Standard Score (SS) of between 1 and 10 points Pre 1 to Pre 2 (see Table 3), with five crossing a clinical boundary into the low average range (SS 86-90: shaded). All 11 participants improved their score Pre 2 to post-intervention. Four (P1, P2, P4, P9) gained 15 SS points or greater (a shift of $\geq 1SD$), and the scores for two of these (P4, P9) also moved across a clinical boundary into the average range (SS >85: shaded). A further four participants made an 8-11 SS point gain (P7, P8, P10, P11), while the SS for P6, P7, P10 and P11 moved into the average range. At maintenance, seven participants achieved the same post-intervention SS or higher. Two participants (P6, P10) who had made a small gain Pre 2 to post-intervention, gained $\geq 1SD$ at maintenance, with P6 crossing a further clinical boundary. While four participants achieved a lower SS at maintenance compared to post-intervention, all scores were higher than at Pre 2 and in the average range.

Table 3

Oral inferential comprehension measures.

Participant School Year		P1 4	P2 6	P3 4	P4 5	P5 5	P6 6	P7 7	P8 5	P9 5	P10 5	P11 6
Test	**Time**											
TOPS Infer [a]	Pre 1	82	80	83	80	82	88	81	82	88	81	97
	Pre 2	90	90	90	82	88	85	82	90	82	80	85
	Post	109	107	95	100	93	89	92	100	103	88	96
	Main	97	114	97	97	96	115	86	100	107	110	93
Pre 1–Pre 2 △		+8	+10	+7	+2	+6	-3	+1	+8	-6	-1	-12
Pre 2-Post △		+19	+17	+5	+18	+5	+4	+10	+10	+21	+8	+11
Post-Main △		-12	+7	+2	-3	+3	+26	-6	=	+4	+22	-3
TOPS Predict [a]	Pre 1	85	75	90	85	80	85	100	76	85	100	85
	Pre 2	80	75	95	80	80	80	95	85	85	106	103
	Post	85	90	110	103	80	100	95	85	80	106	100
	Main	90	105	95	106	90	90	95	90	100	110	90
Pre 1–Pre 2 △		-5	=	+5	-5	=	-5	-5	+9	=	+6	+18
Pre 2-Post △		+5	+15	+15	+23	=	+20	=	=	-5	=	-3
Post-Main △		+5	+15	-15	+3	+10	-10	=	+5	+20	+4	-10
TOPS PS [a]	Pre 1	88	98	95	102	98	95	91	104	86	95	100
	Pre 2	100	85	102	86	98	88	84	98	91	95	102
	Post	90	100	114	102	102	91	91	98	105	95	106
	Main	113	105	114	111	106	100	103	108	100	105	116
Pre 1–Pre 2 △		+12	-13	+7	-16	=	-7	-7	-6	+5	=	+2
Pre 2-Post △		-10	+15	+12	+16	+4	+3	+7	=	+14	=	+4
Post-Main △		+23	+5	=	+9	+4	+9	+12	+10	-5	+10	+10
CELF-4 USP [b]	Pre 1	45	45	56	33	0	56	67	33	33	56	67
	Pre 2	56	67	78	67	67	89	67	66	22	67	78
	Post	89	89	89	67	56	78	100	89	67	89	78
	Main	100	78	78	89	89	67	100	78	56	78	89

Note. PC = Poor Comprehender; TOPS = Test of Problem Solving; Infer = Making Inferences subtest; Predict = Predicting subtest; PS = Problem Solving subtest; CELF = Clinical Evaluation of Language Fundamentals; USP = Understanding Spoken Paragraphs; Main = maintenance testing time point; △ = change in Standard Score points
a = Standard Score (Mean =100 ± 15); b = percentage of inference questions correct out of 9
Shaded cells indicate: crossed clinical boundary into average range

Table 4

Percentage of inconsistencies identified on comprehension monitoring measure.

Participant		**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**
School Year		**4**	**6**	**4**	**5**	**5**	**6**	**7**	**5**	**5**	**5**	**6**
Test	**Time**											
Monitoring	Pre	20	14	20	14	0	0	14	0	0	0	0
	Post	80	43	80	43	0	0	43	0	14	14	14
	Main	80	57	80	29	0	29	43	0	29	86	29

Note. PC = Poor Comprehender; Main = maintenance testing time point

On the Predicting subtest, all participants obtained a similar SS Pre 1 to Pre 2, apart from P11 whose score had increased 18 SS points and crossed a clinical boundary (see Table 3). Four participants made a shift ≥1*SD* Pre 2 to post-intervention, and three of these (P2, P4, P6) moved across a clinical boundary into the average range (SS >85: shaded). The remaining seven made little or no change. At maintenance, seven participants obtained a higher SS than post-intervention, with two making a gain ≥1*SD* (P2, P9), and P9's score moving into the average range. This also occurred for P1, P5 and P8, with SS increases of 5-10 points. Of the remaining four participants, three scored 10-15 SS points below their post intervention score; two being P3 and P6 who had gained ≥1*SD* Pre 2 to post intervention.

On the Problem Solving subtest, four participants had an increased SS Pre 1 to Pre 2, while the remaining seven obtained the same or lower SS (see Table 3). Two participants (P2, P4) gained ≥1*SD* Pre 2 to post-intervention, two improved 10-14 SS points (P3, P9), six made a small or no gain, and P1 scored 10 SS points lower. Two participants (P2, P7)crossed a clinical boundary to score in the average range (SS >85: shaded). Post-intervention to maintenance, P1's SS increased ≥1*SD*, four participants gained 10-12 SS points (P7, P8, P10, P11), and the remaining six had little or no change in score.

On the CELF-4 Understanding Spoken Paragraph subtest, nine participants answered a greater number of inference questions correctly Pre 1 to Pre 2, with six being able to answer two or more questions correctly (see Table 3). Seven participants increased their percentage of questions correct Pre 2 to post intervention, six of these answering two or more questions correctly, two answered the same percentage correct (P4, P11), and two answered one question less (P5, P6). Four participants made further gains post-intervention to maintenance (P1, P4, P5, P11), with the remaining participants achieving a similar percentage correct.

2 Comprehension Monitoring

Five participants (P1, P2, P3, P4, P7) were able to identify one inconsistency at Pre 2, while post-intervention eight could identify one or more inconsistency (see Table 4). At maintenance, all participants except P5 and P9 identified two or more inconsistencies, and eight were able to identify the same number or greater than at post-intervention.

Table 5

YARC Comprehension standard scores and percentage of inference questions answered correctly.

Participant		**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**
School Year		**4**	**6**	**4**	**5**	**5**	**6**	**7**	**5**	**5**	**5**	**6**
Test	**Time**											
Reading Comp [a]	Pre 1 [c]	85	85	85	85	71	80	71	88	86	94	92
	Pre 2 [c]	94	80	100	96	89	105	89	91	78	94	101
	Post [d]	114	111	103	103	84	108	88	91	96	101	95
	Main [c]	93	102	107	107	85	118	94	106	98	104	106
Pre 1-Pre 2 △		+9	-5	+15	+11	+18	+25	+18	+3	-8	=	+9
Pre 2-Post △		+20	+31	+3	+7	-5	+3	-	=	+18	+7	-6
Post-Main △		-21	-9	+4	+4	+1	+10	+6	+15	+2	+3	+11
Inference Questions [b]	Pre 1 [c]	11	34	11	25	25	34	20	58	58	58	50
	Pre 2 [c]	42	27	42	50	33	73	50	42	25	50	60
	Post [d]	64	67	45	67	36	78	31	27	67	67	56
	Main [c]	42	67	58	87	33	87	57	67	60	67	73
Pre 1-Pre 2 △		+31	-7	+31	+25	+8	+39	+30	-16	-33	-8	+10
Pre 2-Post △		+22	+40	+3	+17	+3	+5	-19	-15	+42	+17	-4
Post-Main △		-22	=	+13	+20	-3	+9	+26	+40	-7	=	+17

Note. YARC = York Assessment of Reading Comprehension; Comp = Comprehension; Main = maintenance testing time point; △ = change in Standard Score points

a = Standard Score Mean =100 ± 15; b = percentage of inference questions correct; c = YARC Form A; d = YARC Form B

Shaded cells indicate: crossed clinical boundary into average range

3 Reading Comprehension

a YARC–Australian

In the intervening 9-12 months since initially being identified with poor reading comprehension on the YARC, eight participants' SS had improved, with four now obtaining a SS >95 (P3, P4, P6, P11), and only P2 and P9 obtaining a SS <85. However, as participants presented with ongoing difficulties answering inference questions on both the YARC and PROBE-2, all continued to be considered poor comprehenders for the purposes of this study.

Table 5 shows the YARC reading comprehension SS and change in SS at each time point. The score for four participants (P3, P5, P6, P7) increased ≥1*SD* Pre1 to Pre 2. Their scores also crossed a clinical boundary into the average range (SS >85: shaded), as did the scores for P1 and P4. Three participants made a shift ≥1*SD* Pre 2 to post-intervention (P1, P2, P9), and both P2 and P9 moved across a clinical boundary. The remaining participants had little or no change in SS post-intervention. At maintenance, all participants apart from P5, whose score had regressed across a clinical boundary, obtained a SS >90 and eight a SS >95. P8 scored ≥1*SD* higher and P6 crossed a clinical boundary into the above average range (SS >115), while P1's SS regressed to the same as at pre-intervention.

Five of the six participants whose SS had crossed a clinical boundary Pre 1 to Pre 2 (P1, P3, P4, P6, P7) demonstrated a noticeable increase in the percentage of inference questions answered correctly. The percentage of questions correct for the remaining six participants was either similar or less at Pre 2 than at Pre 1 (see Table 5). Five participants answered noticeably more inference questions correctly post-intervention (P1, P2, P4, P9, P10), including the three who made a SS shift ≥1*SD* at post-intervention, whereas P7 and P8 answered less questions correctly. At maintenance, five participants (P3, P4, P7, P8, P11) improved in the percentage of questions answered correctly while, apart from P1, the remaining participants showed little or no change.

b PROBE-2

Table 6 shows the percentage of comprehension questions answered correctly on the fiction and nonfiction text at each time point, and whether participants achieved the minimum comprehension criterion of 70% of questions correct. At Pre 1, completed on average 6 months prior to Pre 2, two participants (P6, P11) reached the minimum criterion on the fiction text for their age, and no participant reached the criterion on the nonfiction text. At Pre 2, P11 was the only participant to reach the criterion on the fiction text, and again no participant reached the criterion on the nonfiction text. Post-intervention, five participants

Table 6

PROBE-2 percentage of comprehension questions correct and pass/fail criterion at each time point.

Participant		**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**
School Year		**4**	**6**	**4**	**5**	**5**	**6**	**7**	**5**	**5**	**5**	**6**
Test	**Time**											
PROBE-F	Pre 1	25	30	50	30	10	70	40	50	30	50	70
Comp	Pre 2	40	60	0	50	30	40	30	20	60	50	80
	Post	50	50	60	60	60	50	70	80	80	70	80
	Main	70	80	60	80	70	70	70	50	50	70	70
Pre 1-Pre 2 [a]		×	×	×	×	×	×	×	×	×	×	✓
Pre 2-Post [a]		×	×	×	×	×	×	✓	✓	✓	✓	✓
Post-Main [a]		✓	✓	×	✓	✓	✓	✓	×	×	✓	✓
PROBE-NF	Pre 1	62.5	20	38	60	0	30	20	30	20	40	50
Comp	Pre 2	50	40	12.5	60	30	40	40	30	30	40	20
	Post	50	70	60	80	50	60	80	40	50	50	40
	Main	80	50	70	70	50	70	60	60	20	50	50
Pre 1-Pre 2 [a]		×	×	×	×	×	×	×	×	×	×	×
Pre 2-Post [a]		×	✓	×	✓	×	×	✓	×	×	×	×
Post-Main [a]		✓	×	✓	✓	×	✓	×	×	×	×	×

Note. PROBE-F = PROBE-2 fiction text; PROBE-NF – PROBE-2 nonfiction text; Comp = Comprehension; Main = maintenance testing time point

a = Meets 70% comprehension questions correct criterion

reached the criterion on the fiction text (P7, P8, P9, P10, P11) and three (P2, P4, P7) on the nonfiction text, with only P7 reaching the comprehension criterion on both. At maintenance, five further participants reached the criterion on the fiction text, while P8 and P9 no longer achieved 70% of questions correct and P3 maintained the same 60% correct obtained post-intervention Again, fewer participants reached the comprehension criterion at maintenance on the nonfiction text (P1, P3, P4, P6), with only P4 achieving this at both post-intervention and maintenance.

Each text had one literal question. Examination of responses to this question at each time point showed that three participants (P1, P4, P9) answered the literal question incorrectly on the fiction text and five (P2, P5, P8, P9 P10) on the nonfiction text at Pre 1. At Pre 2, seven participants answered the question incorrectly on the fiction text and six on the nonfiction text, while at maintenance no participant answered the literal question correctly and three (P3, P5, P10) answered incorrectly on the nonfiction text.

IV Discussion

While there is evidence to support the effectiveness of interventions for poor comprehenders, most programmes target skills at multiple levels, making it difficult to identify the active ingredients. In this study, following the seminal work of Oakhill and colleagues (Yuill and Oakhill, 1991), we chose to focus on poor comprehenders who presented with difficulties with inferencing but adequate vocabulary, and design and evaluate a pilot programme that directly targeted inferencing. This study, therefore, aimed to explore the effectiveness of a novel intervention, specifically targeting higher-level language skills, in improving oral inference making and comprehension monitoring, and examine any evidence of generalisation to reading comprehension.

1 Oral Inferential Comprehension

Overall, the results showed that oral inference making improved post-intervention for most participants. This was reflected by improved scores for all 11 participants on the TOPS-3E Making Inferences subtest, with shifts across clinical boundaries for six post-intervention. This was also reflected in the increase in correct responses to inference questions by seven participants on the CELF-4 Understanding Spoken Paragraphs. These gains were maintained or improved upon for most participants at maintenance. The change in SS for the TOPS-3E Predicting subtest was less clear, with four participants making clinically significant gains post-intervention, while for four others this was seen at maintenance. Less clinically significant change was evident on the Problem Solving subtest, likely due to stronger scores

pre-intervention for many participants compared with the other inference subtests, although all participants had made positive gains pre-intervention to maintenance.

2 Comprehension Monitoring

Following intervention, improvement in comprehension monitoring was evident for eight of the 11 participants, and at maintenance for a further participant (P6) who had been unable to identify any inconsistencies previously.

3 Reading Comprehension

The second aim was to investigate generalisation to reading comprehension measures. In summary, there was less change in participants' performance on either the YARC standardised or PROBE-2 criterion-referenced measures of reading comprehension than on the oral inference making and comprehension monitoring measures. In the case of the YARC, this is likely due to nine of the 11 participants' reading comprehension SS improving in the intervening 9-12 months between Pre 1 and Pre 2 to then fall in the average range (SS >85) pre-intervention. Only two participants' scores (P2, P9) crossed a clinical boundary Pre 2 to post-intervention and maintained this 4-5 months later. These were the only two participants who scored below average on the YARC pre-intervention, suggesting that the improvement can generalise to reading comprehension tests for those with clinically significant reading comprehension difficulties, as classified by Hulme and Snowling (2011).

All participants continued to fail to reach the minimum comprehension criterion of 70% questions correct on the PROBE-2 at Pre 2 on both the fiction and nonfiction texts, apart from P11 on the fiction text. Following intervention, the strongest improvements were on the fiction texts, with non-fiction texts remaining difficult for most participants. On the fiction text relevant to their age, five participants reached the minimum comprehension criterion of 70% of questions correct post-intervention, while eight attained the criterion at maintenance, although this did not include P8 and P9 who had reached the criterion post-intervention. On the non-fiction text, only three participants reached the minimum criterion post-intervention and four at maintenance. Why the participants had greater difficulty on the nonfiction texts is unclear from the results, but factors such as knowledge-base, level of interest and the types of inference required to comprehend a factual text need to be considered.

Analysis of the percentage of inference questions answered correctly on the YARC revealed that, in line with their clinically significant improvement in SS, P2 and P9 had made the greatest gain post-intervention. At maintenance, P8 whose SS had increased $\geq 1SD$,

obtained the greatest increase, while P3, P4, P7 and P11, whose SS had remained reasonably stable across Pre 2, post-intervention and maintenance, also made good gains. This analysis at the question level suggests the intervention has the potential to improve children's ability to make inferences when reading. Interestingly, ability to answer literal questions correctly improved both post-intervention and at maintenance, consistent with findings of the meta-analysis carried out by Elleman (2017). When studied in conjunction with the comprehension monitoring results, these findings also suggest that examining a child's performance on different types of comprehension question may be beneficial in order to identify poor comprehenders, rather than just considering the overall reading comprehension score.

V Future directions

This was an exploratory study, and more research is required to replicate the findings on a larger scale, with more participants, including those poor comprehenders from older age groups. Implementation in small groups could be investigated to increase the intervention's applicability in educational settings, such as in classrooms delivered by teachers. Future replications should also strengthen the design through the inclusion of repeated measures. The reduced transfer to generalised measures of reading comprehension highlights the need for future research to investigate a longer period of intervention, inclusion of a second phase with more focus on applying the skills to written texts, and/or the inclusion of consolidation activities between sessions.

The issue of lack of sensitivity to identifying weaknesses and change on standardised tests is apparent on the YARC, however, and suggests that analysis of responses to different question types may be more clinically or educationally beneficial than the overall reading comprehension score, with the PROBE-2 being more sensitive here. Further, future research should conduct intervention studies concurrently with initial diagnosis of poor comprehenders, as the YARC SS of most participants had shifted to within the average range pre-intervention, even though they continued to demonstrate marked weaknesses making inferences. Finally, further rigour would be added by moving to a waitlist control design, introducing blinded pre-post assessment, and gathering further background information, such as developmental and medical history, to identify other factors that may impact on performance.

VI Conclusions and clinical implications

While the results of this pilot intervention study are preliminary, they provide the educator and clinician with early evidence that targeting oral inference making and

comprehension monitoring can be effective after only 10 sessions of individual intervention, and for children with varying severity of reading comprehension difficulty. The study identified that for some individual poor comprehenders, who present with adequate vocabulary and grammar but poor higher-level language skills, this targeted approach can be effective. It also identified that generalisation, albeit reduced, can be obtained to standardised measures of reading comprehension for these individuals.

Full details of the intervention, including session plans can be freely downloaded from:
https://www.dropbox.com/sh/vn2was2q3yp2kld/AACK9FDWn4rDihnxG0DHItlua?dl=0

References

Adlof SM and Catts HW (2015) Morphosyntax in poor comprehenders. *Reading and Writing* 28(7): 1051-1070. https://doi.org/10.1007/s11145-015-9562-3

Barth AE and Elleman A (2017) Evaluating the impact of a multistrategy inference intervention for middle-grade struggling readers. *Language, Speech, and Hearing Services in Schools* 48(1): 31-41. https://doi.org/10.1044/2016_LSHSS-16-0041

Bell N (2007) *Visualizing and Verbalizing Stories Book 1*. CA: Gander Publishing.

Bos LT, De Koning BB, Wassenburg SI and van der Schoot M (2016) Training inference making skills using a situation model approach improves reading comprehension. *Frontiers in Psychology* 7(116): 1-13. https://doi.org/10.3389/fpsyg.2016.00116

Bowers L, Huisingh R and LoGiudice C (2005) *Test of Problem Solving 3 Elementary*. Austin: PRO-ED.

Cain K (2010) *Reading development and difficulties*. Chichester, UK:BPS Blackwell.

Catts HW, Adlof SM and Weismer SE (2006) Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research* 49(2): 278-293. https://doi.org/10.1044/1092-4388(2006/023)

Catts HW, Hogan TP and Adlof SM (2005) Developmental changes in reading and reading disabilities. In: Catts HW and Kamhi AG (eds) *The connections between language and reading disabilities*. New Jersey: Lawrence Erlbaum, pp.25-40.

Clarke PJ, Snowling MJ, Truelove E and Hulme C (2010) Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science* 21(8): 1106-1116. https://doi.org/10.1177/0956797610375449

Elleman AM (2017) Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review.

Journal of Educational Psychology 109(6): 761-781. https://doi.org/10.1037/edu0000180

Elwér Å, Gustafson S, Byrne B, Olson RK, Keenan JM and Samuelsson S (2015) A retrospective longitudinal study of cognitive and language skills in poor reading comprehension. *Scandinavian Journal of Psychology* 56: 157– 166. https://doi.org/10.1111/sjop.12188

Fisher D and Frey N (2013) *Better learning through structured teaching: A framework for the gradual release of responsibility* (2nd ed). Alexandria, VA: ASCD.

Hall CS (2016) Inference instruction for struggling readers: A synthesis of intervention research. *Educational Psychology Review* 28: 1-22. https://doi.org/10.1007/s10648-014-9295-x

Hogan T, Bridges MS, Justice LM, and Cain K (2011) Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children* 44(3): 1-20. https://digitalcommons.unl.edu/specedfacpub/79/

Hulme C and Snowling MJ (2011) Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science* 20(3): 139-142. https://doi.org/10.1177/0963721411408673

Kelso K, Whitworth A and Leitao S (2021). Profiles of oral and reading comprehension in poor comprehenders. *Reading & Writing Quarterly* Epub ahead of print 29 Oct 2021:1-18. https://doi.org/10.1080/10573569.2021.1982432

Kelso K, Whitworth A, Parsons R and Leitão S (2020) Hidden reading difficulties: Identifying children who are poor comprehenders. *Learning Disability Quarterly* Epub ahead of print 1 October 2020: 1-12. https://doi.org/10.1177/0731948720961766

Klingner JK, Vaughn S and Boardman AG (2015) *Teaching reading comprehension to students with learning difficulties* (2nd ed). New York: Guilford Press.

Lubliner S (2005) *Getting Into Words*. Baltimore: Brookes.

McGee A and Johnson H (2003) The effect of inference training on skilled and less skilled comprehenders. *Educational Psychology* 23(1): 49-59. https://doi.org/10.1080/01443410303220

McNamara DS and Magliano J (2009) Toward a comprehensive model of comprehension. In: Ross B (ed) *Psychology of learning and motivation* 51. New York: Elsevier, pp.297-384. https://doi.org/10.1016/S0079-7421(09)51009-2

Nation K (2005) Children's reading comprehension difficulties. In: Snowling MJ and Hulme C (eds) *The science of reading: A handbook.* Oxford: Blackwell, pp.248 – 265.

Nation K, Cocksey J, Taylor J S and Bishop DV (2010) A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry* 51(9): 1031-1039. https://doi.org/10.1111/j.1469-7610.2010.02254.x

National Reading Panel (US) (2000) *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health. Washington, DC: US Government Printing Office. Retrieved from www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf

Pardo LS, Raphael TE and Au KH (2011*) QAR Comprehension Lessons Grades 4-5*. New York: Scholastic.

Parkin C and Parkin C (2011) *PROBE2 reading comprehension assessment*. Wellington, NZ: Triune.

Parkin C, Parkin C and Pool B (2002) *Key into Inference*. Lower Hutt, NZ: Triune.

Perfetti C and Stafura J (2014) Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading* 18(1): 22-37. https://doi.org/10.1080/10888438.2013.827687

Raphael T, Highfield K and Au KH (2006) *QAR Now: question answer relationships*. New York: Scholastic.

Rogde K, Hagen ÅM, Melby-Lervåg M and Lervag A (2019) The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews* 15(4): 1-37. https://doi.org/10.1002/cl2.1059

Semel E, Wiig EH and Secord WA (2006) *Clinical Evaluation of Language Fundamentals* (4th Ed) - *Australian Standardised Edition*. Marrickville, NSW: Harcourt.

Shanahan T, Callison K, Carriere C, et al. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from whatworks.ed.gov/publications/practiceguides.

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

Snowling MJ, Stothard SE, Clarke P et al. (2012) *York Assessment of Reading for Comprehension-Primary, Australian Edition (YARC-P Australian).* London: GL Assessment.

Stothard SE, Hulme C, Clarke P, Barmby P and Snowling MJ (2012) *York Assessment of Reading for Comprehension-Secondary, Australian Edition (YARC-S Australian).* London: GL Assessment.

Tate RL, Perdices M, Rosenkoetter U, et al. (2016) The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. *Physical Therapy*, 96(7), e1–e10. https://doi.org/10.2522/ptj.2016.96.7.e1

Tuffin B and Henderson D (1993) *Comprehend It!* Balcatta: R.I.C. Publications.

Yuill N and Oakhill J (1991) *Children's problems in text comprehension: An experimental investigation*. Cambridge: Cambridge University Press.