---

**Article:**
Clarke, PJ orcid.org/0000-0002-2541-7051, Paul, S-AS, Smith, G et al. (2 more authors) (2017) Reading intervention for poor readers at the transition to secondary school. Scientific Studies of Reading, 21 (5). pp. 408-427. ISSN 1088-8438

https://doi.org/10.1080/10888438.2017.1318393

---

# Reading intervention for poor readers at the transition to secondary school.

## Abstract

This study evaluated two 20-week reading interventions for pupils entering secondary school with reading difficulties. The interventions were delivered by trained teaching assistants (three, 35-minute sessions per week). 287 pupils (aged 11 – 13) from 27 schools were randomly allocated to three groups; Reading Intervention (targeting word recognition and decoding skills), Reading Intervention plus Comprehension, or a waiting list control group. Neither intervention produced statistically significant gains in word reading but the Reading Intervention plus Comprehension intervention produced significant gains in reading comprehension (d =.29) and vocabulary (d = .34). Further evaluations of methods to improve word reading in this population are needed.


Keywords: Reading difficulties; Intervention; Randomised Controlled Trial; Secondary School

Although the majority of children attain fluent reading skills, others experience persistent difficulties and their literacy skills fall well below expectation. Indeed, in the US, 24% of eighth grade students (13 to 14 years) have yet to master basic level reading skills (NAEP, 2015). Such students are destined to underachieve academically and are likely to disengage from texts and other literary activities (Polychroni, Koukoura & Anagnostou, 2006). Here we evaluate the effectiveness of two interventions designed to improve reading in students entering UK secondary schools (equivalent to seventh grade) with low literacy skills.

Despite the importance of the transition between primary and secondary school systems, there is a dearth of evidence relevant how to ameliorate reading difficulties at this stage of schooling (See & Gorard, 2014, for a review). Secondary school teachers are typically not well prepared to meet the needs of students with reading difficulties in their classrooms (MacMahon, 2014) and more worryingly, Teaching Assistants (TAs) (also known as classroom assistants) are often responsible for supporting the needs of pupils with reading difficulties. Furthermore, TAs typically receive limited training and have little time to plan their work in liaison with subject specialists (Webster et al., 2011).

Focusing on Randomised Controlled Trials (RCTs), Paul and Clarke (2016) conducted a systematic review of reading interventions for secondary school students. Eight studies published from 1999 to 2014 met inclusion criteria. Promising approaches included modifying teacher language use (Starling, Munro, Togher, & Arciuli, 2012), combining reading strategies with attribution retraining (Berkeley, Mastropieri, & Scruggs, 2011), a class-based learning strategies curriculum (Cantrell et al. 2010), co-operative group strategy training (Vaughn et al. 2011) and content acquisition focused instruction (Vaughn et al. 2013). In general the effect sizes reported were small and it was noted that most of the interventions used strategy based techniques. Moreover, it is striking is that very little

attention has been paid to the core components of reading. According to the simple view of reading (Gough & Tunmer, 1986), two essential sets of skills are required in order to read for meaning: decoding and language comprehension. Decoding refers to translating print into sound. Language comprehension encompasses vocabulary knowledge, understanding of grammar and knowledge of the pragmatic aspects of language use. None of the interventions reported in Paul and Clarke (2016) explicitly focused on decoding, and only one, Vaughn et al. (2013) directly targeted student's language comprehension (by focusing on key content vocabulary).

Slavin, Lake, Davis and Madden (2009) reviewed 96 interventions for younger 'struggling readers' in grades K-5 which had employed an RCT or matched groups design. Findings relate mainly to methods to support decoding; they highlight that, in contrast to computer administered methods which have little measurable impact on reading, one-to-one tutoring with an emphasis on phonics is effective and has more impact than small group tutorials. Furthermore, classroom based approaches based on co-operative learning and structured phonics are effective.

Focusing on UK studies, Snowling and Hulme (2012) reported that the intervention with the strongest evidence for improving decoding is Reading Intervention (RI; Hatcher, Hulme & Ellis, 1994), which originates from Reading Recovery (Clay, 1985) but is supplemented by phonological awareness activities and phonically based instruction. RI involves, reading easy and instructional level books, letter-sound work, phoneme awareness activities, phonological linkage training, writing sentences and spelling. The approach was initially designed to promote reading in 6- to 7-year-old poor readers in a 20-week program delivered by trained teachers. Those who received the RI programme outperformed all other groups in terms of gains made in reading (accuracy and comprehension) and spelling. More recently, the RI program has been adapted for delivery by TAs and evaluated using an RCT

(Hatcher et al. 2006) in which medium to large effect sizes were obtained for word reading accuracy (d =0.69 and d =0.79). It has also been used with younger children (Bowyer-Crane et al. 2008) and children with Down syndrome (Burgoyne et al. 2012).

Fewer studies have been conducted to improve reading comprehension skills of children with specific difficulties in these processes (Nation, 2005). Small-scale evaluations of inference making interventions and mental imagery training have yielded promising results (Yuill & Oakhill 1988, Yuill & Joscelyne 1988). However, at the time of writing, there is robust evidence from only one RCT of interventions designed for 'poor comprehenders' (Clarke, Snowling, Truelove & Hulme, 2010); following recommendations of the National Institute of Child Health and Human Development's National Reading Panel (2000) this program adopted a multi-componential approach (see Clarke, Truelove, Hulme, & Snowling, 2013 for details) to evaluate three interventions alongside a waiting list control group: a text level program, an oral language program and a combined program.

One of the main strategies used in these comprehension interventions was Reciprocal Teaching (Palinscar & Brown, 1984) and specifically, the four reciprocal teaching strategies (clarification, summarisation, question generation and prediction); in the Text comprehension program these strategies were applied to text, in the oral language program to spoken language, and in the combined program across both domains. In addition, research on the effectiveness of vocabulary training for supporting reading comprehension skills (for a meta-analysis see Elleman et al. 2009) influenced its content. In the oral language program children were taught words following the Multiple Context Learning approach (Beck, McKeown & Kucan, 2002). This approach targets 'tier 2 words' selected to be challenging, interesting, relevant and applicable to a child's life experiences. Through structured discussion, pupils learn the meanings of new words, practice clarification strategies and have the opportunity to form associations with known related words.

The interventions were delivered by trained TAs to pairs and through one to one training, three times a week for 20 weeks. All participants who received training performed significantly better than controls on a measure of reading comprehension (WIAT-II) at immediate post-test. Eleven months later, the text comprehension and combined groups had maintained their gains but the oral language group showed increased gains, on average showing an 8 standard score point increase from pre-test to long term follow up. Effect sizes for the oral intervention were medium at immediate post-test (d=0.69) and large at long term follow up (d=1.24).

The data from Clarke et al (2010) are unusual in showing transfer to non-taught vocabulary: at immediate post-test the oral language group demonstrated significant gains on an expressive test of taught and non-taught words (matched for word type and frequency), and a standardised measure of expressive vocabulary (Wechsler Abbreviated Scale of Intelligence, WASI, Wechsler, 1999). Furthermore, the gains made in expressive vocabulary by the oral language group partially mediated the improvements they made in reading comprehension, suggesting a causal role for vocabulary instruction in improving children's reading comprehension skills. A recent systematic review of the impact of vocabulary instruction on reading comprehension by Wright and Cervetti (2016) found very little evidence that direct teaching of word meanings or teaching word learning strategies can improve generalised comprehension; though they do suggest that future research should investigate the effectiveness of approaches which combine multiple, flexible clarification strategies with vocabulary comprehension monitoring. In Clarke et al. (2010) it is possible that by embedding vocabulary instruction within a multi-componential program which includes Reciprocal Teaching strategies alongside opportunities for comprehension monitoring that the effectiveness of vocabulary training for improving reading comprehension was enhanced.

The present study has two aims. First, to investigate whether reading intervention delivered to secondary school pupils on a 1:1 basis by trained TAs, would be effective. Second, to investigate whether pupils with poor word reading skills would benefit more from intensive intervention in decoding skills or a combined program of decoding and language comprehension support. With suitable adaptation to ensure age appropriateness, the Reading Intervention (RI) program developed by Hatcher et al. (2006) is employed to target decoding skills and a new Reading Intervention and Comprehension program (RI + C) which combines RI with the oral language program from Clarke et al. (2010) to target language comprehension as well as decoding difficulties.

Our specific hypotheses were:

1. The Reading Intervention program would lead to improvements in decoding and reading comprehension.
2. The Reading Intervention plus Comprehension program would also lead to improvements in decoding but these would be smaller than those shown by the Reading Intervention group since the time spent teaching decoding skills was reduced.
3. The Reading Intervention plus Comprehension program would lead to larger improvements in reading comprehension ability than the Reading Intervention program.
4. Given the vocabulary component in the Reading Intervention plus Comprehension program, this program should result in gains in vocabulary knowledge (such gains were not expected for the Reading Intervention program).

## Method

### Design

We conducted a Randomised Controlled Trial (RCT) in which both interventions were delivered by the same Teaching Assistants (TAs; equivalent to a classroom assistant in US terminology) in each school. To examine whether gains made by the intervention groups were greater than those made through standard classroom instruction, we included a waiting-list control group.

Due to lower than expected recruitment at the beginning of the study, the schools were recruited in two cohorts. Thirteen schools were recruited to Cohort 1 and 15 to Cohort 2 (two schools participated in both cohorts). Pupils in each school assigned to three conditions, 1) Reading Intervention 2) Reading Intervention plus Comprehension 3) Waiting list control. Pupils' reading performance was assessed at pre-test (t1), mid-test after 10 weeks of intervention (t2), post-test following 20 weeks of intervention (t3) and at a delayed follow-up approximately eight months after the intervention had finished (t4).

**Participants**

The participants in Cohort 1 began receiving the intervention while in Year 7 (at age 11 – 12), and completed the program during Year 8 (at age 12 – 13). All of the participants in Cohort 2 received their intervention during Year 7 (except for the participants in the Control condition who received theirs in Year 8). The flow of participants through the study in each cohort is summarised in Figures 1 and 2 according to CONSORT guidelines (http://www.consort-statement.org/).

Insert Figure 1 here

Insert Figure 2 here

Schools within roughly 90 minutes journey time of Leeds, UK were approached to take part in the study. Demographic information was collected from the National Pupil

Database. Of pupils who provided consent for their data to be accessed, approximately 30% were eligible for free school meals, 60% had a statement of special educational needs[1] or were recorded as School Action Plus[2], and 20% had English as an Additional Language (https://educationendowmentfoundation.org.uk/evaluation/projects/reach). All pupils were taught in mainstream classrooms following the English National Curriculum (https://www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study).

To identify pupils with reading difficulties we asked each school to identify those with attainment in English below level 4 (the level expected at the end of primary school, see https://www.gov.uk/government/publications/english-programmes-of-study-for-key-stage-1-and-2-until-july-2015). These pupils were then screened using the Single Word Reading Test (SWRT, Foster & National Foundation for Educational Research, 2008), an untimed test measuring reading accuracy. Participants were asked to read aloud as many of the words as they could; no discontinuation rule was used. Our inclusion criterion for the study was a standard score of 91 or below on this test. Initially this criterion was set at 85. However our design required a minimum of three pupils per school and to meet this practical need we included a few pupils, with scores just above 85, who we considered would also likely benefit from support. Randomisation was done at the child level. Children were allocated with equal probability within each school to the three arms of the study by an independent body. The allocation used minimisation for gender, age and T1 Single Word Reading scores.

**Measures**

[1] The educational histories of the participants were not available to the research team.
[2] School Action Plus was in place in the UK up until 2014. It was a system for seeking advice from external support services for students who are not making adequate progress in school.

The selected participants were assessed using a full battery of reading and language measures (t1); this was repeated at t3 and t4. The battery consisted of well validated measures of each construct with higher degrees of correlation between measures of the same as opposed to different constructs (see below). The full battery took approximately one hour to complete, assessments were conducted in school and administered by a trained Research Assistant (RA).

**New Group Reading Test Digital.** A computerised assessment, the New Group Reading Test Digital (NGRT, GL Assessment, 2010) was used at all time points except t4. The test took approximately 45 minutes to complete and was supervised by one of the Teaching Assistants in school. This is an adaptive, multiple choice test, which begins with a sentence completion task and then, depending upon the accuracy of responses, either drops down in difficulty to phonics tasks or becomes harder and assesses passage comprehension. The comprehension questions target a range of skills including inferencing and vocabulary. Given the adaptive nature of the test, pupils complete different subtests: at T1, 223 pupils completed the passage comprehension subtest and 16 pupils completed the phonics subtest. At T3 206 pupils completed the passage comprehension subtest and 6 pupils completed the phonics subtest. At both time points, all pupils completed the sentence completion task.

**Test of Word Reading Efficiency 2.** Reading accuracy was measured using the Test of Word Reading Efficiency 2 (TOWRE-2, Wagner, Torgesen & Rashotte, 2011) which includes two subtests; Sight Word Efficiency and Phonemic Decoding Efficiency (Nonword reading). In each of these participants are required to read aloud as many items as possible in 45 seconds. Errors are noted by the examiner but not corrected.

**Single Word Reading Test.** In addition to being used for screening, the SWRT, (Foster & National Foundation for Educational Research (NFER) 2008) was also used at each time point to assess word reading accuracy. The SWRT is a well-validated measure of single

word reading for 6-16 year-olds. Hulme, Nash, Gooch, Lervag and Snowling (2015) report correlations of .78 with a measure of early word reading and .72 with spelling in 7 year-olds (compared with .23 for sentence repetition). Snowling, Duff, Nash and Hulme (2015) found it to differentiate well between typically developing children and poor readers with language impairment.

**York Assessment of Reading for Comprehension – Secondary.** The intention was for passage reading accuracy and passage comprehension to be measured using two supplementary passages from the York Assessment of Reading for Comprehension (YARC-Secondary, Stothard et al. 2012). However, it became apparent early in the assessment process that these passages were too complex for many of the participants in this study and a decision was made to introduce a discontinuation rule when sixteen errors were made. At pre-test only 28% of the sample were able to complete enough of this test to generate accuracy and comprehension scores, this rose to 46% at post-test. Because this test was too difficult for the majority of participants in the trial, scores from the YARC were not analysed.

**Wechsler Individual Achievement Test 2nd Edition UK for Teachers.** An additional measure of reading comprehension was the Wechsler Individual Achievement Test 2nd Edition UK for Teachers (WIAT II UK-T, Wechsler, 2006) which includes a range of reading material (e.g. sentences, passages, non-fiction, fiction, reviews) which pupils can choose to read either silently or aloud before responding orally to open ended comprehension questions. The questions tap a wide variety of comprehension skills including vocabulary knowledge, inferential processing, and prediction and locating specific details in text.

**Wechsler Abbreviated Scale of Intelligence II – Vocabulary.** The ability to define the meanings of words was assessed using two vocabulary measures. The Vocabulary scale from the Wechsler Abbreviated Scale of Intelligence II (WASI II, Wechsler, 2011) required

participants to provide spoken definitions for increasingly complex and abstract words. Responses were scored using a rating system quantifying the depth of the knowledge demonstrated (0 = no/limited knowledge, 1 = partial knowledge, 2 = detailed knowledge).

**Bespoke Vocabulary.** We also used a bespoke vocabulary test devised to assess pupils' ability to define the words taught in the Reading Intervention plus Comprehension intervention. The bespoke test consisted of sixteen words, eight taught words (e.g. "prevent") and eight comparable non-taught words (e.g. "claim") pairwise matched on Thorndike and Lorge (1944) written frequency. The test assessed tier 2 vocabulary (Beck, McKeown & Kucan, 2002), i.e. useful, and relatively high-frequency words with an age of acquisition that is above the pupil's chronological age. Responses were scored using a 0-2 point rating system, similar to that employed in the WASI II (see above). Assessing non-taught words allowed us to assess the extent to which pupils showed generalization to novel items. Correlations between the two bespoke measures of vocabulary and the WASI standardized vocabulary measure were substantial at both times of measurement.

**Intervention programs.** Intervention materials were taken from Hatcher et al. (2006) and Clarke et al. (2010) and modified by the research team to ensure their appropriateness for secondary school pupils. The interventions were delivered by TAs who received five days of intensive training and ongoing email, telephone and in-school support where necessary. Each intervention program shared the same basic structure and consisted of three 35- minute sessions per week for 20 weeks (35 hours intervention per child). Pupils in the Reading Intervention condition received three Reading Intervention sessions; those in the Reading Intervention plus Comprehension condition received two Reading Intervention sessions and one Comprehension session per week. The structure of the Reading Intervention and Comprehension sessions is shown in Figures 3.

Insert Figure 3 here

The Reading Intervention sessions contained six components: 1) Reading an easy level book 2) reading an instructional level book during which a running record was taken 3) Letter knowledge, word level and phonological work 4) Writing 5) Introduction to a new instructional level book 6) Reading aloud of a new instructional level book. Each school received a box of finely graded books to use in the Reading Intervention sessions. These books were selected in consultation with experts from a local independent book store, and were graded according to the system developed by Hatcher (2000) which takes into account word length, sentence length, and grammatical features of text. Easy level refers to books read with < 95% accuracy; instructional level books are read with 90-94% accuracy. A structured set of materials and activities taken from the 'Sound Linkage' approach (Hatcher, Duff & Hulme, 2014), was used to teach phonological awareness and links with letter knowledge in the Reading Intervention sessions. The Reading Intervention sessions were outlined in a detailed teaching manual; tactile letters, phoneme charts and prompt cards were provided to support the delivery of the components.

Insert Figure 4 here

The Comprehension sessions contained six components: 1) Introduction 2) Vocabulary 3) Listening Comprehension 4) Figurative language 5) Narrative 6) Plenary. Teaching in this program involved working with spoken language, with minimal reading. The sessions were based on two books, The Firework Maker's Daughter (Pullman, 1995) and The Tales of Beedle the Bard (Rowling, 2008), which were selected because of their wide appeal to this age group, range of complex vocabulary and interesting storylines. Following a brief introduction the session began with a 'word of the day' which was taught using a slightly modified version of the Multiple Context Learning (MCL) approach (Beck, McKeown &

Kucan, 2002). Discussion was supported with graphic organisers, definition cards and illustrations, and the new words were introduced via listening to a section of one of the selected books. In component 2, pupils completed a worksheet or game-like activity inspired by Reciprocal Teaching (RT) (Palincsar & Brown, 1984), linked to the section of the book they had listened to. In component 3, pupils explored figurative language including idioms, riddles and jokes related to the story. In component 4, pupils created a story map collaboratively with the TA; using the map they discussed key parts of the narrative structure, including characters, places, timelines and key events. Story retells were also encouraged. The plenary provided an opportunity for the revision of main ideas and strategies, and pupils were prompted to make inferences about what might happen next in the story.

The interventions were designed to be administered by TAs with no pre-requisite knowledge of reading instruction; we provided detailed manuals and session plans with scripted instructions to promote treatment fidelity. The TAs received explanations about the principles underpinning randomised controlled trials; we stressed the importance of maintaining the integrity of the two different programs and avoiding contamination. Due to the scale and funding constraints of the study we were not able gather systematic data on treatment fidelity.

**Waiting list control group.** The pupils in the waiting list control group received 'business-as-usual' instruction. Whilst this is likely to have been largely at the classroom level, we cannot rule out the possibility that they may also have received individualised or group-based remedial teaching. We are not aware of any systematic confounds; however, we were unable to document the activity of this group in detail. There was considerable variation across schools in terms of timetabling; intervention pupils were taken out of different classes at different times, thus, it was not possible to estimate the instruction that the business-as-usual group received.

## Results

The majority of the analyses were conducted in Stata 14.0 (Stata Corp, College Station, Texas, USA). Structural equation models were constructed using Mplus 7.3 (Muthen & Muthen, 2014) using Full Information Maximum Likelihood estimators to allow for missing data and robust (Huber–White) standard errors to allow for the clustering of children within schools.

Participant characteristics and scores on baseline measures are shown for the two cohorts separately, and for the combined sample, in Table 1. It is clear that the sample as a whole had low reading scores (mean standard scores between 74 and 78 across all measures). There were small and inconsistent differences between the two cohorts on the scaled scores for the key outcome measures shown in Table 1, with the only statistically significant difference being on the SWRT (SWRT SS d = -.33 95% CI [-.57, -,09]; TOWRE Word Efficiency SS d = .20 95% CI [-.03 , .45]; NGRT Ability Score d = .22, 95% CI [-.04 - .48]; WIAT II Reading Comprehension SS d = -.21, 95% CI [- .45 , .04]; WASI II Vocabulary SS d = .02 95% CI [-.22 , .26] ). Therefore, for the purposes of further analyses the two cohorts were combined to maximize statistical power.

Insert Table 1 here

Descriptive statistics for all outcome measures at screening, pre-test and post-test for the Reading Intervention, Reading Intervention plus Comprehension and the Waiting list control groups are shown in Table 2. It is clear that there are appreciable rates of attrition between t1 and t3 (28% of the sample are lost overall at immediate posttest) but these rates appear similar across groups (attrition rates of 27% RI; 26% RI+C; 28% Control) and the two cohorts (27% attrition in cohort 1 and 28% attrition in cohort 2). We also checked for possible biasing effects due to attrition by examining equivalence between groups at baseline; for this purpose we computed a decoding factor score (Towre word, Towre Nonword,

SWRT) and a reading comprehension factor score at t1 (NGRT, WIAT-II). For the sample as a whole there was baseline equivalence on both measures as indicated by a one-way analysis of variance on the factor scores comparing groups (Decoding $F(2, 233) = 0.10$; $p = 0.91$; Reading Comprehension $F(2, 233) = 0.04$; $p = 0.96$). Similarly, for children present at immediate posttest (t3) test there was no sign of imbalance on decoding or reading comprehension skills at baseline (Decoding $F(2, 207) = 0.02$; $p = 0.97$; Reading Comprehension $F(2, 203) = 0.22$; $p = 0.80$). Finally, a chi-square test showed no difference in rates of attrition as a function of group and cohort ($\chi^2 = 0.59$; $df = 4$; $p = .96$). Little's MCAR test confirmed that missing data for the language and reading pre-test and post-test measures used in the SEM models presented later could be considered to be missing completely at random ($\chi^2 = 112.74$; $df = 102$; $p = 0.22$). In addition Little's CDM test confirmed that data missing from the language and reading post-test measures did not vary as function of group ($\chi^2 = 23.07$; $df = 51$; $p = 0.999$). In short, though rates of attrition were substantial in this study, attrition appears to have been random and there is no evidence to suggest that it has introduced significant bias.

To aid interpretation of the descriptive statistics relating to the NGRT (which is an adaptive measure in which pupils complete different subtests), the details relating to the number of pupils completing specific subtests at T1 and T3 are as follows:

- At T1 223 pupils completed the passage comprehension subtest and 16 pupils completed the phonics subtest. All pupils completed the sentence completion task.
- At T3 206 pupils completed the passage comprehension subtest and 6 pupils completed the phonics subtest. All pupils completed the sentence completion task.

Insert Table 2 here

It is clear that the intervention groups were approximately equated on all measures at screening/pre-test as expected given assignment using minimisation for gender, age and T1 SWRT.

Preliminary analyses on each measure shown in Table 2 assessed differences between groups in regression (ANCOVA models). These models were implemented in Mplus using FIML to account for missing data and with robust (Huber-White) standard errors to allow for the non-independence of observations due to clustering of children within schools. In each model there were two dummy codes (Reading Intervention vs Control; Reading Intervention plus Comprehension vs Control) with initial level of performance on the same measure (the autoregressor) as a covariate. To test the assumption of homogeneity of regression slopes across groups the group x covariate interaction terms were included in initial models. In no case did the group x covariate interaction terms approach statistical significance so they were dropped from the models reported. The absence of significant group by covariate interaction terms justifies the use of the simpler models with parallel slopes reported here.

Table 2 shows the marginal mean differences (gains adjusted for baseline imbalance) at post-test between the Reading Intervention vs Control groups, and the Reading Intervention plus Comprehension vs Control groups, together with associated tests of the significance of these group differences. At t3 (immediate post-test), the Reading Intervention plus Comprehension group generally show larger gains than the Reading Intervention group (except on SWRT). The only test on which the Reading Intervention group shows a significantly greater improvement than the Control group is the NGRT. The Reading Intervention plus Comprehension group show significantly greater improvements than the Control group on the NGRT, and the taught bespoke vocabulary and untaught bespoke vocabulary measures.

At t4 (8 month delayed post-test) there are no significant effects of either intervention on any reading measure which indicates that any improvements on the NGRT at t3 are short-lived. However, both intervention groups show higher scores than the Controls on WASI II vocabulary, with numerically larger improvements in the RI+C group. These improvements are hard to explain given that there were no effects on this measure immediately following training for either group, and furthermore the RI group did not receive any vocabulary instruction.

Insert Table 3 here

**Effects of intervention on word reading, reading comprehension and vocabulary knowledge.** Our principal interest was to examine the extent to which the interventions produced improvements on word reading, reading comprehension and vocabulary knowledge. Each of these constructs was measured by multiple tests (Word reading – SWRT; TOWRE II Sight Word Efficiency; TOWRE II Phonemic Decoding Efficiency; Reading comprehension – WIAT II UK-T; NGRT; Vocabulary knowledge - Taught bespoke vocabulary; Untaught bespoke vocabulary; WASI II vocabulary) each of which are subject to measurement error. We therefore constructed five latent variable models: three models at t3 for word reading, reading comprehension and vocabulary knowledge; and two models at t4 for word reading and vocabulary knowledge. The latent variables correct for measurement error and assess the underlying factors that capture the common variance shared by the different measures used to assess each construct.

In the SEM models, effect sizes for the intervention effect were calculated from the y-standardized partial regression coefficients for the dummy-coded Group variables. These partial regression coefficients can be interpreted as equivalent to Cohen's d; they express the difference in group means in z-score units after allowing for any group differences at baseline

(Brown, 2006). It should be noted that a critical assumption for the analyses presented below is that there are equivalent slopes between the pre-test and post-test factor scores across groups. For each analysis preliminary models were run which included the interaction terms between the pre-test score and the two group dummy codes. In no case was any of these interaction terms close to being statistically significant, confirming that the slopes for the intervention groups did not differ significantly from the slope for the control group. Accordingly, the models reported below omit these interaction terms.

The model for word reading at t3 is shown in Figure 5 and at t4 in Figure 6; both models provide good to excellent fits to the data (t3 $\chi^2$ (18) = 30.346, p = 0. 034; RMSEA = 0.050 [90% CI 0.014, 0.079]; CFI = 0.99; TLI = 0.99; t4 $\chi^2$ (16) = 15.27, p = 0.50; RMSEA = 0.00 [90% CI 0.000 -0.063]; CFI = 1.00; TLI = 1.00). At t3 it is notable that word reading shows very high longitudinal stability which attests to the reliability of the latent variables used to assess the construct. The most critical result from this analysis is that the Reading Intervention and Reading Intervention plus Comprehension groups show no improvement compared to the control group with effect sizes that are negligible (d=.06 [95% CI -.050; .171] and d=.10 [95% CI -.023; .222] respectively). At t4, the picture is essentially identical.

The model for reading comprehension at t3 is shown in Figure 7 and provides a good fit to the data ($\chi^2$ (18) = 24.004, p = 0.1549; RMSEA = 0.035 [90% CI 0.000 -0.068]; CFI = 0.99; TLI = 0.99). The reading comprehension factor shows very high longitudinal stability. The most critical result from this analysis is that the Reading Intervention plus Comprehension group show a small but statistically significant improvement in reading comprehension compared to the control group (d=.29 [95% CI 0.037, 0.545] while the Reading Intervention group do not differ from the control group d=.076 [95% CI -0.220, 0.372]. An equivalent model at t4 is not possible as only one measure, WIAT II, was administered.

Finally, models at t3 and t4 for vocabulary knowledge are shown in Figures 8 and 9 respectively; both provide good to excellent fits to the data (t3 $\chi^2$ (17) = 11.432, p = 0.8333; RMSEA = 0.00 [90% CI 0.000 -0.00]; CFI = 1.0; TLI = 1.01; t4 $\chi^2$ (17) = 28.29, p = 0.029; RMSEA = 0.53 [90% CI 0.017 -0.033]; CFI = 0.98; TLI = 0.97). The vocabulary factor shows high longitudinal stability. The most critical result from these analyses is that at t3 the Reading Intervention plus Comprehension group show a small but statistically significant improvement in vocabulary compared to the Control group (d=.34 [95% CI 0.142, 0.540] while the Reading Intervention group do not differ from the Control group d=.10 [95% CI -087; 0.287]; the same pattern is maintained at t4 where the Reading Intervention plus Comprehension group show an improvement in vocabulary compared to the Control group of the equivalent size (d=.33 [95% CI 0.034; 0.63]).

It should be noted that the good fit indices for the models reported here indicates that the measurement models used (the identified factors) are consistent with the data. It is notable however, that the models (with the exception of word reading at t3, Figure 5) do not display factorial invariance (the unstandardized loadings on the factors differ between pre-test and post-test). This means that the composition of the factors assessing word reading, reading comprehension, and vocabulary at t1 differ somewhat from the same factors measured later (with the exception of word reading at t3). This likely reflects the fact that different measures show different degrees of improvement as a result of the interventions. Nevertheless, these models give an estimate of the size of changes in word reading, reading

comprehension and vocabulary produced by our interventions when these constructs are assessed by latent variables with high reliability.

## Discussion

We evaluated the effectiveness of two reading intervention programs, one emphasizing decoding the other decoding and comprehension, for pupils experiencing reading difficulties upon entering secondary school at age 11 to 12 years. We expected that a Reading Intervention program designed to promote decoding and word recognition skills in the context of graded book reading would lead to improvements in word reading for both intervention groups but that these would be smaller in the Reading Intervention plus Comprehension group because the students in this group received fewer sessions of Reading Intervention. Disappointingly neither of the groups showed significant gains in word reading in comparison to the Control group at t3 or t4 on a latent variable measure of word reading.

The findings for reading comprehension are more in line with our predictions. We hypothesized that the Reading Intervention plus Comprehension program would lead to greater improvements in reading comprehension ability than the Reading Intervention program. This was the case; on the reading comprehension latent variable, the Reading Intervention plus Comprehension group showed a small but statistically significant improvement compared to Controls at t3, while the Reading Intervention group did not. However, since we only had one measure of reading comprehension at t4, we cannot be sure that this modest gain was maintained. Finally, in line with our hypothesis, the Reading Intervention plus Comprehension group (who received explicit vocabulary training) made significant improvements relative to the Reading Intervention group and the Control group on a latent variable measure of vocabulary knowledge and this improvement was maintained at follow-up.

Notwithstanding these promising findings, at least for reading comprehension and vocabulary, scrutiny of the improvements on the individual measures of reading and vocabulary complicates the results. First, although we expected a dosage effect with regard for reading accuracy, so that the Reading Intervention plus Comprehension group would show less improvement, the trend was in fact in the opposite direction – on all measures (with the exception the untimed single word reading test), the word reading gains were greater in the Reading Intervention plus Comprehension group. It seems possible that the comprehension elements of the program facilitated the development of reading fluency (and hence higher scores on timed tests) but since none of the gains were statistically significant, this idea remains speculative.

Second, it is clear that the improvement in reading comprehension for the Reading Intervention plus Comprehension group is largely carried by the NGRT measure, since effects on the WIAT II were negligible. We express caution here; the NGRT is an adaptive test which involves multiple choice questions where guessing may play a role. Moreover, on the more commonly used WIAT II test, the gains were actually negative at follow-up. Further research is needed with more diverse measures to assess whether reliable improvements in comprehension can be obtained with the methods used here.

Finally, the Reading Intervention plus Comprehension group made gains on both Taught and Non-Taught bespoke vocabulary knowledge at t3 relative to the Reading Intervention and control groups. The improvements shown on the Taught words represent one of near transfer only whereas the improvement on Non-Taught words provides some evidence that the pupils were able to apply strategies to new words. This finding is consistent with Clarke et al. (2010) but contrasts with the majority of previous research which has failed to find generalisation beyond taught words (Kamil & Chou, 2009). In the present study and in Clarke et al. (2010) a Multiple Context Learning approach (Beck, McKeown and Kucan,

2002) to vocabulary training was integrated with Reciprocal Teaching (Palinscar and Brown, 1984) and importantly repeated opportunities to practice the 'clarification' strategy. This combination of complimentary approaches may be an optimal way to reinforce strategy use and should be investigated further. In addition future research needs to consider the extent to which gains made on expressive vocabulary tests may actually instead reflect changes in the pragmatic language skills needed for successful test responses. These teaching methods may not be improving vocabulary knowledge per se but rather enabling pupils to produce richer, more detailed oral definitions of existing knowledge.

At t4, gains on the bespoke vocabulary measures are no longer apparent which is likely to be because the control group is no longer untreated at this stage, they will have also been taught the target words and so differences are not to be expected. What is more perplexing is that, on the standardized vocabulary measure, WASI II, both intervention groups show gains compared to controls, even though vocabulary was not specifically targeted in the Reading Intervention sessions. It is possible that these gains are attributable to some implicit learning of words contained in the graded texts introduced during the Reading Intervention sessions, though we have no way of testing this idea.

It is clear from the above, that the present study had a number of limitations. Together the findings highlight the fact that reading tests vary in the skills they assess, a point made cogently by Keenan, Betjemann and Olson (2008); while our use of latent factors allowed us to evaluate the impact of an intervention on the shared variance between tests deemed to assess a given construct, interpretation of the findings is complicated by inconsistent gains across different measures. Furthermore, as with any null result, it is not possible to explain why neither intervention produced gains in word reading. We speculate that the long-standing nature of these pupils' difficulties and its effects on self-efficacy and motivation to read may be relevant but we have no evidence to support this conjecture.

A further point to consider is the effectiveness of the training in Reading Intervention. The training program was closely modelled on the method used successfully in previous research with TAs in primary schools; however it was our impression that, given their experience was primarily with secondary school pupils, the TAs in the current study were less well-equipped to benefit from the training. In the light of limited information on the fidelity of implementation , our hypothesis remains speculative. Nonetheless. with hindsight, more needed to be done to tailor the training to those working in secondary schools who may have little or no experience of teaching phonics or delivering specialist individualised interventions.

Nevertheless, the finding of gains in reading comprehension and in vocabulary in the Reading Intervention plus Comprehension group provides some modest support, for the claim that oral language training can be effective in boosting pupils' ability to extract meaning from text (see also Clarke et al. 2010; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013). Furthermore, it would seem to make sense that such an approach should be beneficial for pupils with very weak word recognition and decoding skills as the activities are based around pupils listening to and discussing text rather than reading passages of text aloud. More generally, however, given the complexity of reading, and the need for different skills to act in concert with each other, it would be possible to have greater confidence in the findings if all measures had showed similar effects.

In conclusion, this RCT evaluated the effectiveness of two programs designed to improve reading accuracy and reading comprehension skills in poor readers on entering secondary school at age 11 - 12. It is clear that many of the children recruited to this study had severe and longstanding reading difficulties. Neither of our interventions produced appreciable improvements in reading accuracy in this group, while the Reading Intervention plus Comprehension group showed small improvements in reading comprehension ability

and in vocabulary knowledge. It is clear from our study that a significant number of children are entering secondary school with serious reading problems. Given the high longitudinal stability of reading, coupled with the complexity and diversity of the secondary school curriculum, it is difficult to provide effective remediation at this late stage. The current study underlines two important facts; it is important to ensure that reading interventions are delivered early in the school system before difficulties become entrenched and, more research is urgently needed to develop programs for use in secondary schools that can effectively support children with severe and persistent difficulties.

## References

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). Bringing words to life: Robust vocabulary instruction. New York: Guilford Press.

Berkeley, S., Mastropieri, M. A., & Scruggs, T.E. (2011). Reading Comprehension Strategy Instruction and Attribution Retraining for Secondary Students With Learning and Other Mild Disabilities. Journal of Learning Disabilities, 44 (1), 18 – 32.

Bowyer-Crane, C., Snowling, M.J., Duff, F., Carroll, J.M., Fieldsend, E., Miles, J., … Hulme, C. (2008). Improving early language and literacy skills: differential effects of an oral language versus a phonology with reading intervention. Journal of Child Psychology and Psychiatry, 49(4), 422–432.

Brown, G (2004). The Efficacy of Question-Answering Instruction for Improving Year 5 Reading Comprehension. Unpublished PhD dissertation. Sydney, Australia: University of Western Sydney.

Burgoyne, K., Duff, F.J., Clarke, P.J., Buckley, S, Snowling, M.J., Hulme, C. (2012). Efficacy of a reading and language intervention for children with Down syndrome: a randomized controlled trial. Journal of Child Psychology and Psychiatry, 53(10), 1044 – 1053.

Cantrell, S.C., Almase, J.F., Carter, J.C., Rintamaa, M., & Madden, A. (2010). The Impact of a Strategy-Based Intervention on the Comprehension and Strategy Use of Struggling Adolescent Readers. Journal of Educational Psychology, 102 (2), 257 – 280.

Clarke, P., Snowling, M., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties: A Randomized Controlled Trial. Psychological Science, 21 (8), 1106-1116.

Clarke, P. J., Truelove, E., Hulme, C., & Snowling, M. J. (2013). Developing reading comprehension. Chichester: Wiley-Blackwell.

Clay, M. (1985). The early detection of reading difficulties (3rd ed.). Tadworth, Surrey: Heinemann.

Elleman, A., Lindo, E., Morphy, P., & Compton, D. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. Journal of Educational Effectiveness, 2, 1 – 44.

Foster, H. and the National Foundation for Educational Research (2008) Single Word Reading Test 6–16. London: GL Assessment.

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.

GL Assessment (2010) New Group Reading Test (NGRT). London: GL Assessment.

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

Hatcher, P. (2000). Predictors of Reading Recovery book levels. Journal of Research in Reading, 23(1), 67-77.

Hatcher, P. J., Duff, F. J., & Hulme, C. (2014). Sound linkage: An integrated programme for overcoming reading difficulties. Chichester: Wiley-Blackwell.

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, 41–57.

Hatcher, P.J., Hulme, C., Miles, J.N.V., Carroll, J.M., Hatcher, J., Gibbs, S., Smith, G.,

Bowyer-Crane, C. & Snowling, M. (2006). Efficacy of Small Group Reading Intervention for Beginning Readers with Reading-Delay: A Randomized Controlled Trial. Journal of Child Psychology and Psychiatry, 47(8), 820-827.

Hulme, C., Nash, H.M., Gooch, D.C., Lervåg, A., & Snowling, M.J. (2015) The Foundations of Literacy Development in Children at Family Risk of Dyslexia. Psychological Science, 26(12), 1877–1886. doi: 10.1177/0956797615603702

Keenan, J. M., Betjemann, R. S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. Scientific Studies of Reading, 12, 281 - 300.

MacMahon, B. (2014). Making the invisible visible: disciplinary literacy in secondary school classrooms. Irish Educational Studies, 33(1), 21 – 36.

NAEP. (2015). *The nation's report card.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

National Institute of Child Health and Human Development (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. Washington, DC: U.S. Government Printing Office.

Palincsar A.S., & Brown, A.L. (1984) Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. Cognition and Instruction, 1(2), 117-175

Paul, S-A.S., & Clarke, P.J. (2016). A systematic review of reading interventions for secondary school students. International Journal of Educational Research, 79, 116 – 127.

Polychroni, F., Koukoura, K., & Anagnostou, l. (2006). Academic self-concept, reading attitudes and approaches to learning of children with dyslexia: Do they differ from their peers? European Journal of Special Needs Education, 21(4), 415-430.

Pullman, P. (1995). *The Firework Maker's Daughter.* United Kingdom: Doubleday.

Rowling, J.K. (2008). The Tales of Beedle the Bard. London: Bloomsbury Publishing PLC.

See, B.H. & Gorard, S.A. (2014). Improving Literacy in the Transition Period: A Review of the Existing Evidence on What Works. British Journal of Education, Society and Behavioural Science, 4 (6), 739-754.

Slavin, R.E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. Reading Research Quarterly, 43, 290-322.

Slavin, R.E., Lake, C., Davis, S., & Madden, N. (2009). What Works for Struggling Readers? An adapted review from Slavin, R.E., Lake, C., Davis, S., & Madden, N. (2009). Effective Programs for Struggling Readers: A Best Evidence Synthesis. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education. Retrieved from http://www.bestevidence.org.uk/assets/what_works_for_struggling_readers.pdf

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*, 57, 1360-1369 doi:10.1111/jcpp.12497

Snowling, M.J., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. International Journal of Language & Communication Disorders, 47(1), 27–34.

Stothard, S.E, Snowling, M.J, Clarke, P.J, Barnby, P., & Hulme, C. (2012). York Assessment of Reading & Comprehension: Secondary Edition. London: GL Assessment.

Starling, J., Munro, N., Togher, L., & Arciuli, J. (2012). Training Secondary School Teachers in Instructional Language Modification Techniques to Support Adolescents With Language Impairment: A Randomized Controlled Trial. Language, Speech, and Hearing Services in Schools, 43, 474 - 495.

Thorndike, E.L. and Lorge, I. (1944). The Teacher's Word Book of 30,000 Words. New York: Teachers College, Columbia University.

Vaughn, S., Klingner, J.K., Swanson, E.A., Boardman, A.G., Roberts, G., Mohammed, S.S., & Stillman-Spisak, S.J. (2011). Efficacy of Collaborative Strategic Reading with Middle School Students. American Educational Research Journal, 48(4), 938 - 964.

Vaughn, S., Swanson, E.A., Roberts, G., Wanzek, J., Stillman-Spisak, S.J., Solis, M., & Simmons, D. (2013). Improving Reading Comprehension and Social Studies Knowledge in Middle School. Reading Research Quarterly, 48(1), 77 – 93.

Wagner, R., Torgesen J.K., & Rashotte, C. (2011) Test of Word Reading Efficiency 2. Austin TX: Pro-ED.

Webster, R. Blatchford, P., Bassett, P., Brown, P., Martin, C., & Russell, A. (2011). The wider pedagogical role of teaching assistants. School Leadership & Management, 31(1), 3 - 20.

Wechsler, D. (2005). Wechsler Individual Achievement test – Second UK Edition (WIAT II – UK). London: Harcourt Assessment.

Wechsler, D. (2006) Wechsler Individual Achievement Test 2nd Edition UK for Teachers (WIAT II UK-T). London: Pearson Assessment.

Wechsler, D. (2011) Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI II) London: Pearson Assessment.

Wright, T.S. Cervetti, G.N. (2016). A Systematic Review of the Research on Vocabulary Instruction That Impacts Text Comprehension. Reading Research Quarterly. Advance online publication. doi:10.1002/rrq.163

Yuill, N., & Joscelyne, T. (1988). Effect of organisational cues and strategies on good and

poor comprehenders' story understanding. Journal of Educational Psychology, 80, 152–158.

Yuill, N., & Oakhill, J. (1988). Effects of inference training on poor reading comprehension. Applied Cognitive Psychology, 2(1), 33–45.

Table 1 Participant characteristics, mean standard scores (standard deviation), for the sample and each cohort at T1, T2, T3, and T4

| | Cohort 1 | | Cohort 2 | | Sample | |
|---|---|---|---|---|---|---|
| | | n | | n | | n |
| Gender (male: female) | | | | | | |
| T1 | 60:49 | 109 | 101:68 | 169 | 161:117 | 278 |
| T2 | 46:41 | 87 | 80:50 | 130 | 125:91 | 217 |
| T3 | 44:41 | 85 | 77:50 | 127 | 121:91 | 212 |
| T4 | 36:39 | 75 | 65:46 | 111 | 101:85 | 186 |
| Mean Age (years: months) | | | | | | |
| T1 | 12:3 | 109 | 11:8 | 169 | 11:10 | 278 |
| T2 | 12:5 | 86 | 11:10 | 130 | 12:1 | 216 |
| T3 | 13:0 | 85 | 12:3 | 127 | 12:6 | 212 |
| T4 | 13:6 | 76 | 13:0 | 111 | 13:3 | 187 |
| SWRT | | | | | | |
| T1 | 76.09 (6.453) | 109 | 78.49 (7.813) | 169 | 77.55 (7.391) | 278 |
| T3 | 78.06 (7.951) | 85 | 79.20 (9.562) | 127 | 78.75 (8.949) | 212 |
| T4 | 79.69 (9.414) | 75 | 80.11 (9.93) | 36 | 79.31 (9.684) | 111 |
| TOWRE II (WRE)* | | | | | | |
| T1 | 76.33 (10.302) | 107 | 74.11 (10.876) | 169 | 74.97 (10.694) | 276 |
| T3 | 77.34 (9.999) | 85 | 77.43 (11.440) | 127 | 77.40 (10.861) | 212 |
| T4 | 80.29 (10.102) | 74 | 79.14 (11.444) | 28 | 79.60 (10.909) | 186 |
| NGRT | | | | | | |
| T1 | 77.66 (10.209) | 100 | 79.93 (8.176) | 138 | 76.66 (9.105) | 238 |
| T2 | 78.45 (8.748) | 87 | 78.35 (8.543) | 129 | 78.39 (8.606) | 216 |
| T3 | 77.67 (8.356) | 85 | 79.06 (9.787) | 127 | 78.50 (9.244) | 212 |
| WIAT II (Reading Comp) | | | | | | |
| T1 | 71.93 (15.555) | 106 | 75.05 (14.848) | 169 | 73.85 (15.172) | 275 |
| T3 | 71.96 (15.820) | 83 | 73.26 (15.600) | 125 | 72.74 (15.663) | 208 |
| T4 | 73.33 (13.153) | 75 | 75.86 (15.341) | 111 | 74.84 (14.515) | 186 |
| WASI II Vocabulary (SS) | | | | | | |
| T1 | 4.49 (2.331) | 106 | 4.44 (2.035) | 169 | 4.46 (2.150) | 275 |
| T3 | 3.82 (2.077) | 85 | 4.52 (2.054) | 126 | 4.24 (2.086) | 211 |
| T4 | 4.37 (1.951) | 75 | 4.65 (2.231) | 111 | 4.54 (2.12) | 186 |

*TOWRE II (WRE) = Word Reading Efficiency ** WIAT II (Reading Comp) = Reading Comprehension *** WASI II Vocabulary (SS) = Scaled Score

Table 2 Mean raw scores (and standard deviations) for the outcome measures for all groups at each time point, cohorts combined.

| | Reliability[1] | | | | | | | | RI vs. Control | | RI+C vs. Control | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | RI | n | RI+C | n | Control | n | mmd | p | mmd | p |
| SWRT | .9 | T1 | 32.76 (8.112) | 95 | 32.43 (8.317) | 94 | 32.19 (8.269) | 89 | | | | |
| | | T3 | 36.90 (8.665) | 73 | 35.60 (8.500) | 72 | 35.30 (9.883) | 67 | 1.25 | 0.46 | 0.41 | 0.59 |
| | | T4 | 38.37 (8.680) | 68 | 37.48 (10.517) | 65 | 38.15 (9.721) | 54 | 0.94 | 0.24 | -1.07 | 0.50 |
| NGRT* | .90 | T1 | 230.91 (57.205) | 82 | 239.25 (58.905) | 80 | 230.75 (66.463) | 77 | | | | |
| | | T3 | 258.30 (50.526) | 73 | 269.82 (42.456) | 72 | 246.99 (56.925) | 67 | 13.26 | 0.03 | 17.22 | 0.003 |
| TOWRE II Sight Word | .9 | T1 | 53.15 (12.479) | 94 | 52.98 (13.275) | 94 | 52.13 (13.477) | 89 | | | | |
| | | T3 | 56.34 (13.485) | 73 | 57.32 (12.348) | 72 | 56.46 (13.382) | 67 | -0.64 | 0.49 | 0.73 | 0.43 |
| | | T4 | 59.37 (12.787) | 67 | 61.35 (12.125) | 65 | 60.85 (12.705) | 54 | -0.55 | 0.73 | 0.96 | 0.61 |
| TOWRE II Phonemic Decoding | .9 | T1 | 21.68 (9.766) | 94 | 22.39 (10.130) | 94 | 21.32 (10.424) | 88 | | | | |
| | | T3 | 24.59 (9.487) | 73 | 25.76 (10.184) | 72 | 23.42 (10.414) | 67 | 1.26 | 0.12 | 1.81 | 0.64 |
| | | T4 | 26.46 (9.767) | 67 | 29.08 (10.959) | 65 | 27.26 (10.751) | 54 | -1.24 | 0.56 | 1.18 | 0.71 |
| WIAT II Reading comprehension** | .95 | T1 | 74.35 (13.528) | 92 | 73.07 (15.342) | 94 | 74.15 (16.666) | 89 | | | | |
| | | T3 | 70.10 (14.842) | 73 | 74.60 (15.665) | 70 | 73.71 (16.382) | 65 | -3.35 | 0.14 | 1.97 | 0.36 |
| | | T4 | 73.26 (13.923) | 68 | 76.45 (15.071) | 64 | 74.91 (14.630) | 54 | -1.08 | 0.72 | -0.12 | 0.97 |
| WASI II Vocabulary | .86 - .94 | T1 | 17.22 (5.572) | 92 | 16.69 (5.666) | 94 | 16.58 (6.049) | 89 | | | | |
| | | T3 | 16.96 (5.853) | 73 | 17.66 (5.000) | 71 | 17.19 (5.927) | 67 | 0.16 | 0.83 | 0.58 | 0.18 |
| | | T4 | 18.43 (5.156) | 67 | 20.11 (5.458) | 65 | 18.48 (6.191) | 54 | 1.85 | 0.039 | 2.41 | 0.02 |
| Taught words | .81 | T1 | 2.50 (2.756) | 92 | 2.49 (2.801) | 94 | 2.69 (2.794) | 89 | | | | |
| | | T3 | 3.10 (3.359) | 73 | 4.39 (3.491) | 72 | 3.54 (3.492) | 67 | 0.15 | 0.65 | 1.32 | 0.004 |
| | | T4 | 3.97 (3.669) | 68 | 5.03 (4.023) | 65 | 4.80 (3.983) | 54 | -0.32 | 0.62 | 0.93 | 0.41 |
| Non -Taught words | .77 | T1 | 2.43 (2.628) | 92 | 2.17 (2.133) | 94 | 2.44 (2.472) | 89 | | | | |
| | | T3 | 3.01 (2.951) | 73 | 3.69 (2.766) | 72 | 3.15 (3.220) | 67 | 0.15 | 0.72 | 0.82 | 0.03 |
| | | T4 | 3.60 (3.115) | 68 | 3.80 (3.397) | 65 | 4.13 (3.686) | 54 | -0.11 | 0.89 | 0.44 | 0.64 |

1. Cronbach α
*Reading Ability Score **Standard Score

| | TI TOWRE II Sight | T1 TOWRE II Phon. | T1 SWRT | T1 WIAT II | T1 NGRT | T1 Taught words | T1 Non-Taught words | T1 WASI II | T3 TOWRE II Sight | T3 TOWRE II Phon. | T3 SWRT | T3 WIAT II | T3 NGRT | T3 Taught words | T3 Non-Taught words | T3 WASI II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TI TOWRE II Sight | 1.00 | | | | | | | | | | | | | | | |
| T1 TOWRE II Phon. | 0.68 | 1.00 | | | | | | | | | | | | | | |
| T1 SWRT | 0.79 | 0.61 | 1.00 | | | | | | | | | | | | | |
| T1 WIAT II | 0.50 | 0.37 | 0.54 | 1.00 | | | | | | | | | | | | |
| T1 NGRT | 0.59 | 0.36 | 0.64 | 0.54 | 1.00 | | | | | | | | | | | |
| T1 Taught words | 0.14 | 0.06 | 0.19 | 0.40 | 0.35 | 1.00 | | | | | | | | | | |
| T1 Non-Taught words | 0.15 | -0.01 | 0.22 | 0.34 | 0.35 | 0.65 | 1.00 | | | | | | | | | |
| T1 WASI II | 0.33 | 0.18 | 0.34 | 0.52 | 0.49 | 0.61 | 0.57 | 1.00 | | | | | | | | |
| T3 TOWRE II Sight | 0.86 | 0.64 | 0.73 | 0.43 | 0.59 | 0.16 | 0.16 | 0.29 | 1.00 | | | | | | | |
| T3 TOWRE II Phon. | 0.67 | 0.79 | 0.66 | 0.36 | 0.44 | 0.10 | 0.03 | 0.20 | 0.72 | 1.00 | | | | | | |
| T3 SWRT | 0.75 | 0.63 | 0.84 | 0.50 | 0.63 | 0.16 | 0.20 | 0.34 | 0.72 | 0.66 | 1.00 | | | | | |
| T3 WIAT II | 0.42 | 0.31 | 0.47 | 0.54 | 0.61 | 0.41 | 0.40 | 0.49 | 0.45 | 0.32 | 0.48 | 1.00 | | | | |
| T3 NGRT | 0.55 | 0.33 | 0.57 | 0.48 | 0.64 | 0.38 | 0.30 | 0.42 | 0.59 | 0.45 | 0.65 | 0.50 | 1.00 | | | |
| T3 Taught words | 0.22 | 0.14 | 0.22 | 0.42 | 0.45 | 0.62 | 0.57 | 0.54 | 0.29 | 0.19 | 0.24 | 0.51 | 0.50 | 1.00 | | |
| T3 Non-Taught words | 0.16 | 0.11 | 0.19 | 0.41 | 0.40 | 0.62 | 0.53 | 0.55 | 0.20 | 0.17 | 0.21 | 0.50 | 0.42 | 0.76 | 1.00 | |
| T3 WASI II Vocab | 0.20 | 0.14 | 0.24 | 0.41 | 0.46 | 0.62 | 0.51 | 0.64 | 0.27 | 0.18 | 0.25 | 0.54 | 0.47 | 0.67 | 0.66 | 1.00 |

Table 3 Correlations between the outcome measures at time points 1 and 3 for the whole sample

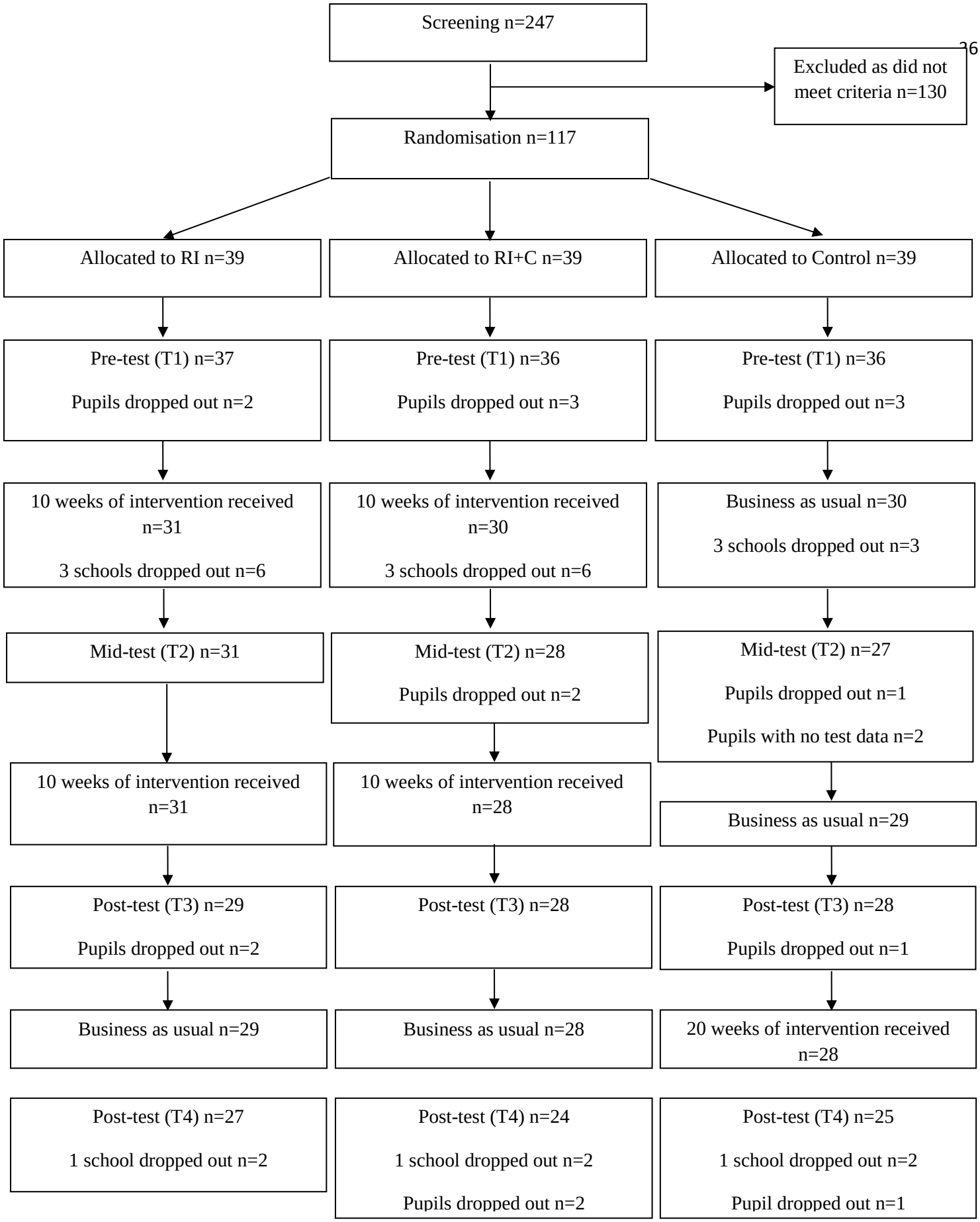


Figure 1 Consort Diagram Cohort 1

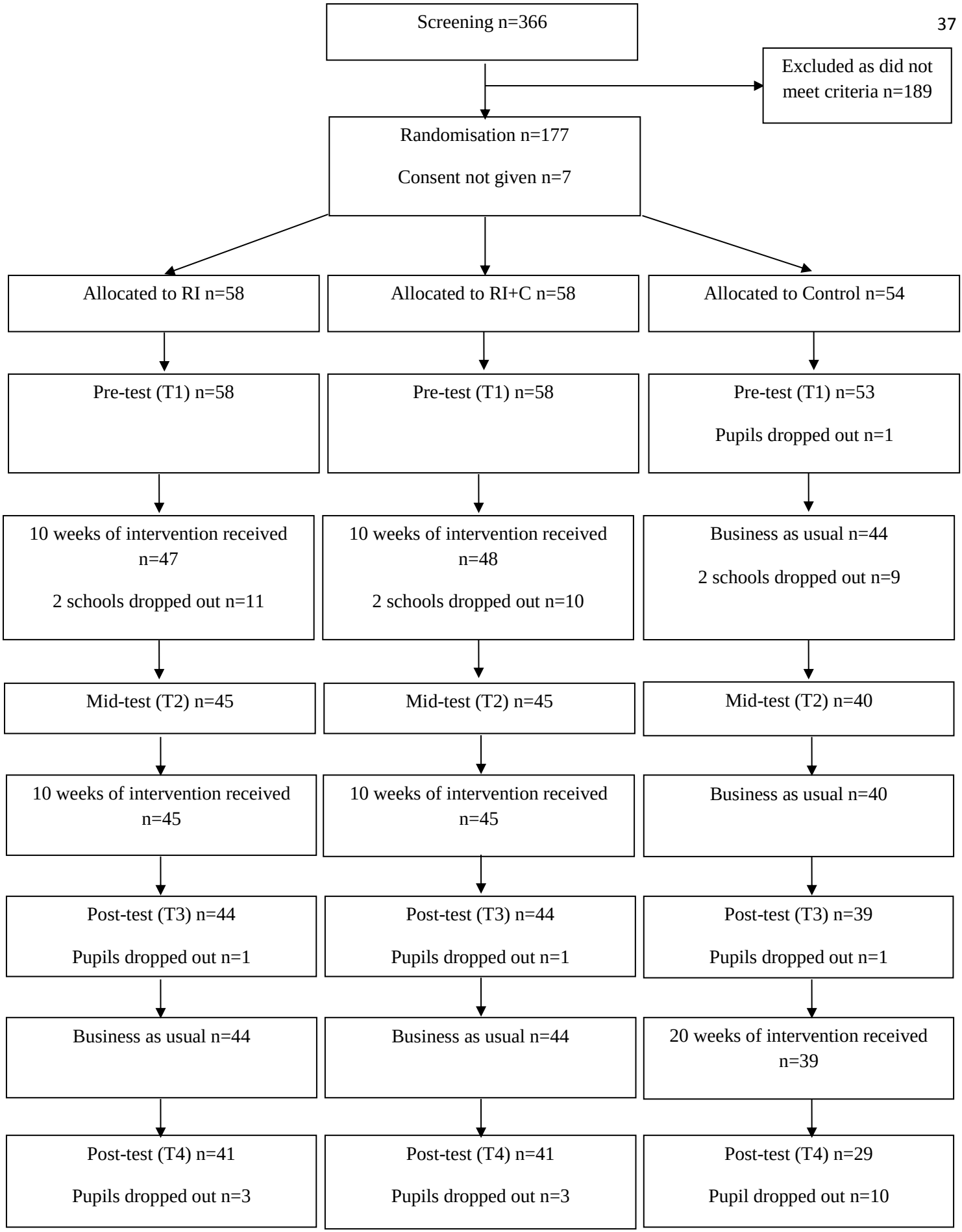


Figure 2 Consort Diagram Cohort 2

| Time (mins) | Activity |
|---|---|
| 9 | Reading ‘easy’ book(s)<br>Running record of book at instructional level (introduced at the end of the previous session) |
| 17 | Letter and word identification<br>Phonological activities (including those from Sound Linkage)<br>Cut-up and reorder sentence from previous session<br>Writing a sentence |
| 9 | Introduction to new book at the ‘instructional’ level<br>Attempt to read new book |

Figure 3 Reading Intervention session plan

| Time (mins) | Activity |
|---|---|
| 2.5 | Introduction |
| 10 | Vocabulary - ‘Word of the day’ |
| 10 | Listening comprehension - Using Reciprocal Teaching strategies |
| 5 | Figurative language |
| 5 | Spoken narrative - ‘Story mapping’ |
| 2.5 | Plenary |

Figure 4 Comprehension session plan

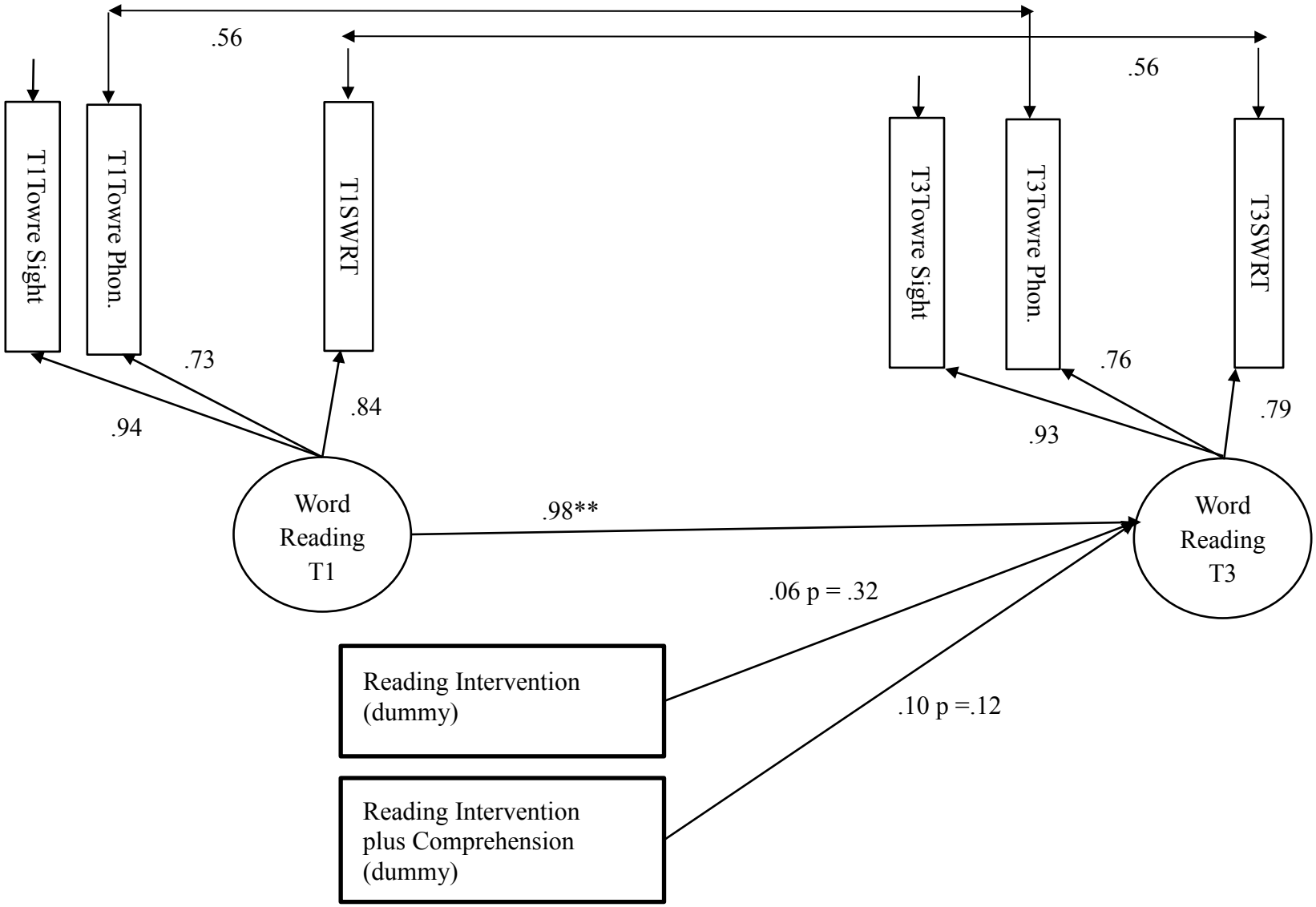


Figure 5 Model showing the effects of the intervention on word reading at immediate post-test (t3). Standardized coefficients shown (except for dummy variables where y-standardized values are shown). Robust (Huber-White cluster estimators) standard errors do not differ appreciably from simple standard errors. The twin headed arrows represent the covariance between the error variances in the model.

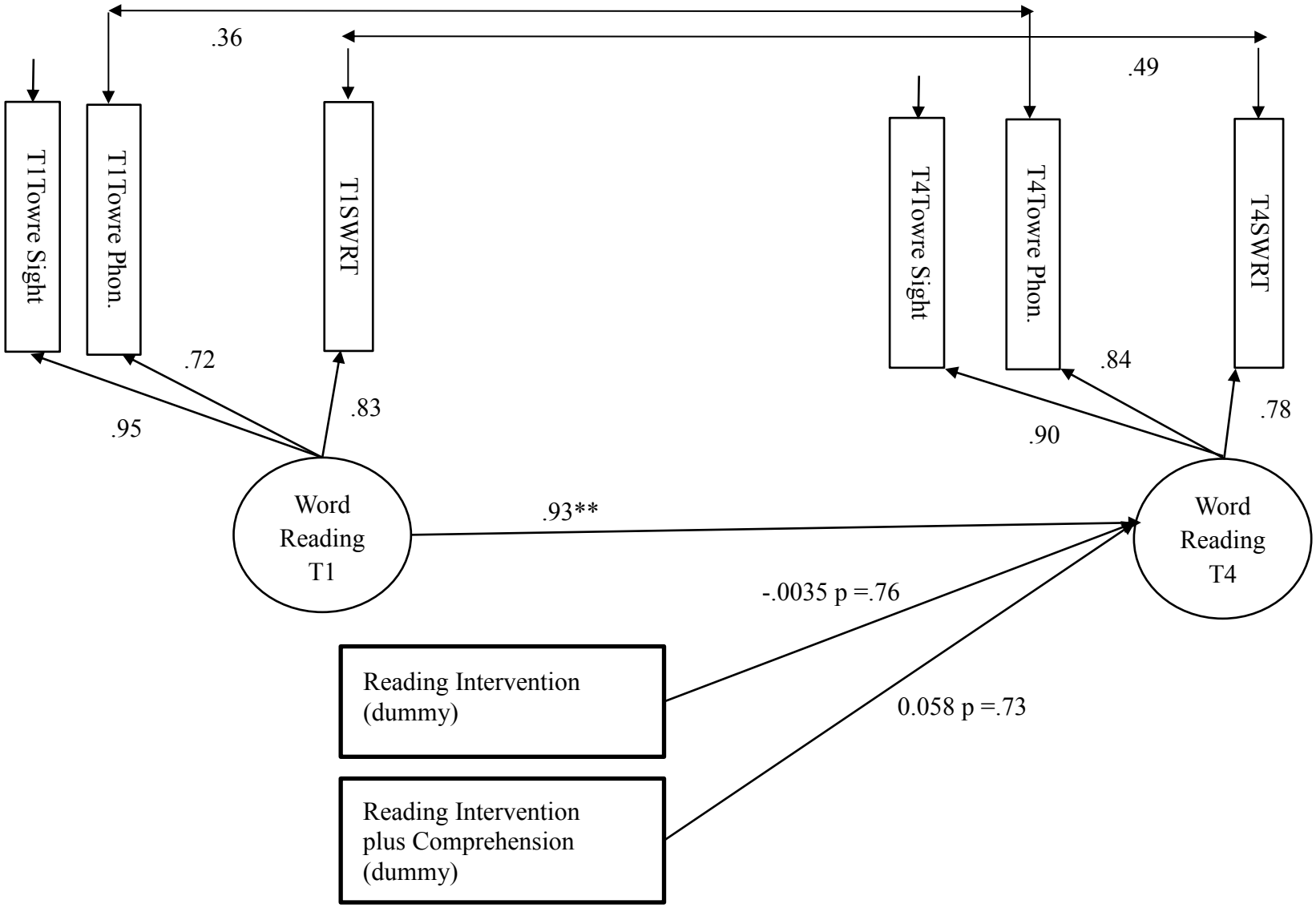


Figure 6 Model showing the effects of the intervention on word reading at delayed post-test (t4). Standardized coefficients shown (except for dummy variables where y-standardized values are shown). Robust (Huber-White cluster estimators) standard errors do not differ appreciably from simple standard errors. The twin headed arrows represent the covariance between the error variances in the model.

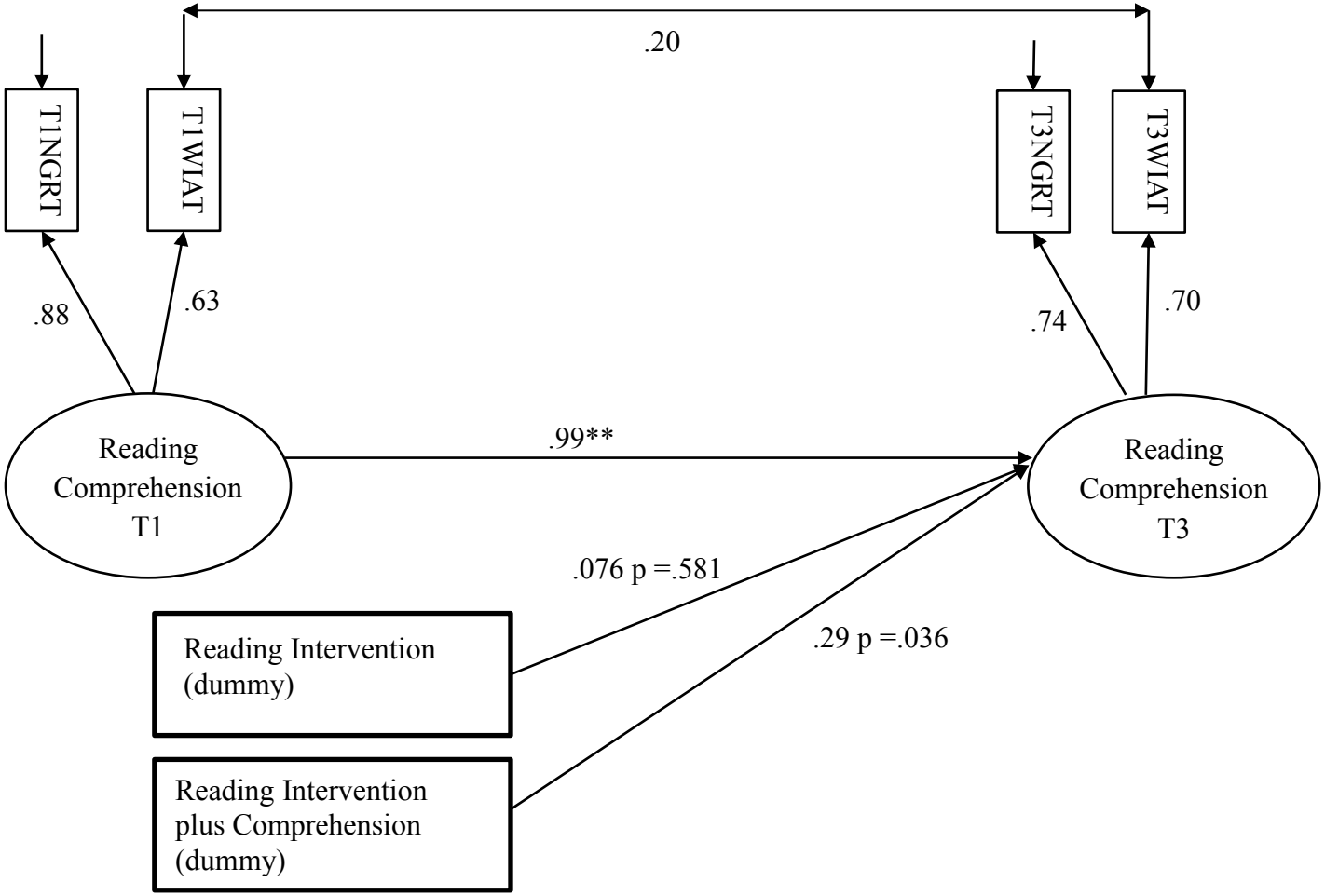


Figure 7 Model showing the effects of the intervention on reading comprehension at immediate post-test (t3). Standardized coefficients shown (except for dummy variables where y-standardized values are shown). Robust (Huber-White cluster estimators) standard errors do not differ appreciably from simple standard errors. The twin headed arrow represents the covariance between error variances in the model.

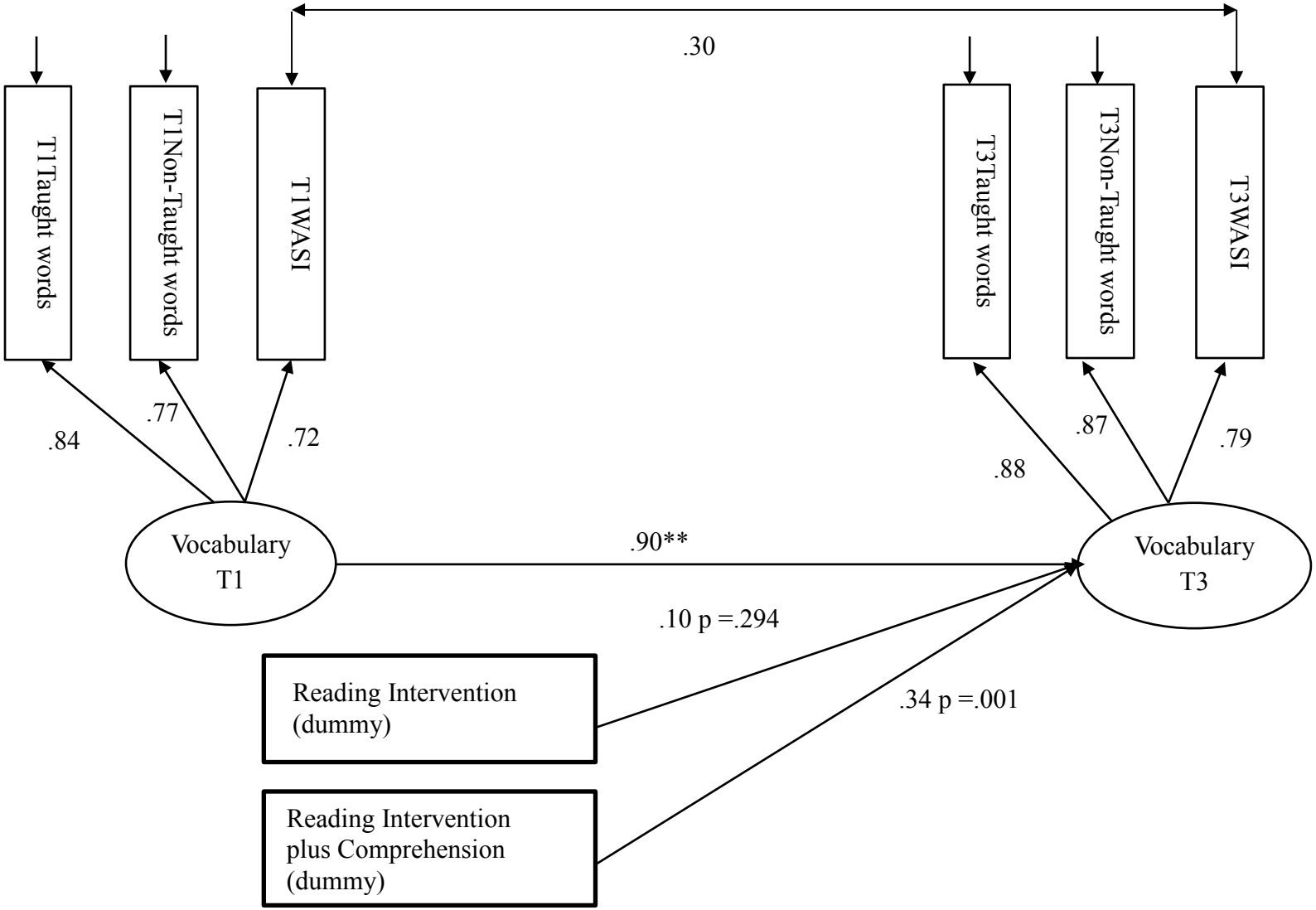


Figure 8 Model showing the effects of the intervention on vocabulary knowledge at immediate post-test (t3). Standardized coefficients shown (except for dummy variables where y-standardized values are shown). Robust (Huber-White cluster estimators) standard errors do not differ appreciably from simple standard errors. The twin headed arrow represents the covariance between error variances in the model.

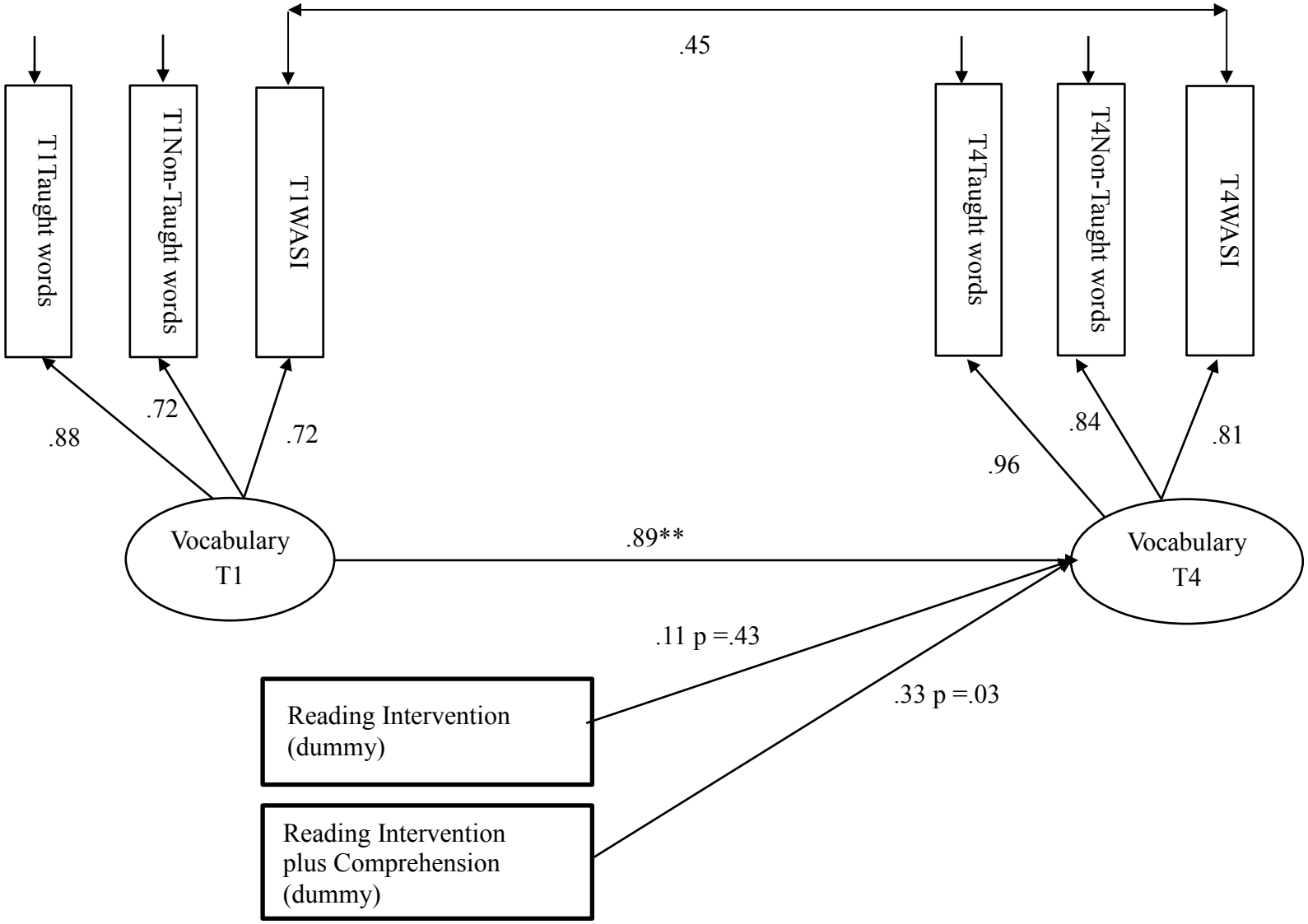


Figure 9 Model showing the effects of the intervention on vocabulary knowledge at delayed post-test (t4). Standardized coefficients shown (except for dummy variables where y-standardized values are shown). Robust (Huber-White cluster estimators) standard errors do not differ appreciably from simple standard errors. The twin headed arrow represents the covariance between error variances in the model.