Article:
Jones, Manon, Evans-Jones, Gwennant, Huws, Tesni et al. (2026) Cross-language Transfer in Bilingual Children with Reading Difficulties:Effects of a Literacy Intervention on Word Reading. READING AND WRITING. ISSN: 0922-4777

https://doi.org/10.1007/s11145-026-10795-4

Reading and Writing
https://doi.org/10.1007/s11145-026-10795-4

Cross-language transfer in bilingual children with reading difficulties: effects of a literacy intervention on word reading

Manon Jones[1] · Gwennant Evans-Jones[1] · Tesni Huws[1] · Catherine Clark[1] · Charles Hulme[2,3] · Cameron Downing[4]

Received: 6 October 2025 / Accepted: 2 March 2026

Abstract
Approximately 40% of children worldwide learn to read in a language other than their home language or in two languages simultaneously. Although literacy instruction in one language can enhance reading outcomes in another, less is known about how such cross-language effects unfold over time in children with reading difficulties. In a pre-registered, cross-over randomised controlled trial (registered on 23/09/2022; ID: ISRCTN10963572), we evaluated a 15-week Welsh-language literacy intervention in 257 bilingual children with poor reading skills. Initially, 129 children (mean age = 111.44 months, SD = 12.23) received the intervention, while 128 (mean age = 110.62 months, SD = 12.34) served as a waitlist control. Reading efficiency in Welsh and English was assessed at baseline (t1), post-intervention (t2), and follow-up (t3; ~ 3 months later). The intervention produced significant improvements in Welsh word reading at posttest, and children who received the intervention earlier showed advantages in Welsh at follow-up consistent with a timing-dependent effect. Short-term improvements were also observed in English decoding at posttest, providing evidence of cross-language transfer in this low-achieving sample. At follow-up, no between-group differences were observed in English. To our knowledge, this is the first randomised controlled trial to evaluate a literacy intervention delivered in Welsh, and the first to evaluate temporal effects on cross-language transfer in poorer readers. The findings highlight the importance of instructional timing and sequencing in bilingual literacy intervention.

Keywords Bilingual literacy · Cross-language transfer · Reading intervention

Extended author information available on the last page of the article

Published online: 23 March 2026

Introduction

Globally, multilingual education is the norm rather than the exception, with a large proportion of children receiving literacy instruction in a language other than their home language, or in two languages simultaneously (UNESCO, 2016). In bilingual learners, literacy development in one language is often associated with literacy development in another. Numerous studies have shown that foundational skills such as phonological awareness, word decoding, and vocabulary correlate across languages (e.g., Comeau et al., 1999; Durgunoğlu et al., 1993; Verhoeven, 2007; Zhang et al., 2017). This convergence has been interpreted within theoretical frameworks proposing that bilinguals draw on a set of shared underlying linguistic and cognitive resources when learning to read. In particular, the Common Underlying Proficiency (CUP) hypothesis suggests that bilingual literacy development is supported by core metalinguistic and phonological processes that are not language-specific. More recent interactive accounts (Chung et al., 2019; Koda, 2005) emphasise that transfer is not automatic, but rather depends on instructional quality, the developmental level of the learner, and the linguistic relationship between the two languages.

The nature of the writing systems involved plays an important role in shaping transfer. According to psycholinguistic grain size theory (Ziegler & Goswami, 2005), children rely on different units of phonological structure when decoding depending on orthographic transparency. In languages with transparent alphabetic systems such as Spanish, Finnish, or Turkish, consistent grapheme–phoneme mappings support early and efficient decoding, whereas in more opaque systems, such as English, children must rely on larger and more variable grain sizes. Welsh is one such transparent alphabetic orthography, with relatively consistent grapheme–phoneme correspondences, whereas English is markedly more opaque. Children typically learn to decode more efficiently and earlier in Welsh than in English (Caravolas et al., 2013, 2019; Spencer & Hanley, 2003). To the extent that literacy instruction in Welsh strengthens sublexical decoding processes, these processes may, in principle, support later English word reading. However, Welsh and English differ considerably in morphosyntax and vocabulary: Welsh is a VSO Celtic language featuring initial consonant mutation and relatively rich inflectional morphology, whereas English is SVO with reduced morphological marking. Therefore, we might expect transfer to be strongest for sublexical decoding processes rather than for higher-level oral language or morphosyntactic knowledge.

Crucially, the extent to which such transfer occurs appears to depend on instruction. For example, Kim and Piper (2019) found that explicit instruction targeting foundational skills in Kiswahili and English produced bi-directional transfer, while typical classroom instruction did not. Kim et al. (2025) further demonstrated that structured literacy instruction in Setswana led to stronger longitudinal transfer to English word reading, even when instruction was delivered only in the L1. These findings suggest that instructional quality shapes the degree to which underlying phonological and orthographic knowledge can generalise across languages, and raise the question of whether similar effects occur in children with weaker literacy skills, whose phonological and lexical representations may be less robust and therefore less likely to transfer spontaneously.

The current study

The present study examines this question in bilingual Welsh-English children identified as having reading difficulties. This population is of particular interest because cross-language transfer mechanisms may be fragile in struggling readers, and there are few randomised controlled trials (RCTs) testing causal transfer effects in this group. Wales also offers a distinctive educational context: children in Welsh-medium schools are typically taught to read in Welsh first, with English literacy introduced later. However, children in these schools vary in their home language backgrounds. Some speak Welsh predominantly at home, whereas others speak English at home and learn Welsh primarily through schooling. Because this study was not powered to detect subgroup differences, we do not present separate inferential analyses for these bilingual profiles in the main results. Instead, we report exploratory descriptive and model-based subgroup analyses in the supplementary materials.

In this randomised controlled trial, we evaluated a 15-week (~30+hours) structured literacy intervention—combined explicit phonics, phonological awareness activities, vocabulary instruction, and shared/guided reading—delivered exclusively in Welsh. In North Wales, where this study was conducted, many schools adopt an immersion model: children first learn to read exclusively in Welsh, with systematic English literacy instruction introduced around Year 3. The activities in our intervention intensively trained word-level and sublexical skills, which are hypothesised to underpin cross-language transfer (Cummins, 1979; Koda, 2005).

Assessments of Welsh and English word reading were conducted at baseline (t1), posttest (t2), and follow-up (~3 months later, t3). The intervention group received the programme first and after they had completed it, the waitlist control group received the same intervention, while the original intervention group returned to regular classroom instruction. Final assessments (t3) were conducted after both groups had completed the intervention.

The waitlist design allowed us to test causal effects of the intervention at posttest (t2), when only one group had received the intervention. At follow up (t3), both groups had received the intervention but at different times. Consequently, group differences at t3 reflected instructional timing (earlier or later treatment).

We therefore tested whether the intervention improved Welsh word reading, whether these gains generalised to English word reading, and whether children who received the intervention earlier showed evidence of continued advantage at follow-up after returning to standard classroom instruction. Our hypotheses were that (a) the intervention would produce immediate improvements in Welsh and English decoding relative to a waitlist control group, (b) if receiving the intervention earlier meant that children took better advantage of subsequent classroom instruction, benefits in Welsh would be evident at follow-up within the original intervention group (c.f., Latimier et al., 2021), and (c) that transfer effects (to English) would be weaker and potentially less stable in children with reading difficulties than has been observed in unselected bilingual samples when examined at posttest. The design allowed us to explore whether earlier instruction in Welsh would also enable enhanced long-term transfer, reflected in increased English reading outcomes.

Methods

Design & participants

A randomised controlled trial (RCT) was conducted in 20 Welsh-medium state primary schools in North Wales, where 43.65% of schools meet the Welsh Government's definition of Welsh-medium provision (i.e., ≥ 80% of activities conducted in Welsh). The trial targeted children in Years 3–6 and was pre-registered (https://doi.org/10.1186/ISRCTN10963572). Ethical approval was granted by the Research Ethics Committee of [omitted for review]. A previous trial (Authors, 2024) reported an effect size of $d = 0.32$ for reading accuracy. Assuming a smaller effect size for cross-language generalisation ($d \approx 0.25$), a power analysis of the ANCOVA model indicated that a sample size of $N = 142$ (71 per arm) would be required to achieve 80% power ($p < .05$, two-tailed) with a pretest–posttest correlation ($r = .85$; as reported in a previous trial; Authors, 2024). We aimed to recruit 20 schools with 12 poor readers each, yielding approximately 120 children per arm. Children were randomised within classrooms to either a 15-week literacy intervention or a business-as-usual waitlist control. Assessments were conducted at pre-test (t1), posttest (t2), and again three months later (t3).

Headteachers provided consent for school participation, and informed consent was obtained from parents. Schools identified the 12 children in Years 3–5 with the weakest literacy skills using national literacy test scores, resulting in a total sample of $N = 258$ participants. The intervention arm included $N = 129$ children (mean age = 111.44 months, SD = 12.23); of these, $n = 65$ spoke Welsh at home, and $n = 64$ spoke English at home. The control arm included $N = 128$ children (mean age = 110.62 months, SD = 12.34); $n = 61$ spoke Welsh at home, and $n = 67$ spoke English at home. On the TOWRE English word reading efficiency subtest (normed mean = 100, SD = 15), the sample had a mean standard score of $M = 80.58$ (SD = 16.33). 56% of the sample scored ≤ 85 (≥ 1 SD below the normative mean) and 28% scored ≤ 70 (≥ 2 SDs below the normative mean), indicating that a majority of participants were in the low-achievement range. Random allocation was carried out within classes by an independent evaluator.

Assessments were conducted at pre-test (t1), posttest (t2), and again three months later (t3). Because the waitlist control group received the intervention after t2, analyses at t3 compare groups that have both received the intervention but differ in the timing of receipt and in their exposure to classroom instruction following the intervention. Accordingly, follow-up analyses are interpreted as timing-based contrasts rather than sustained treatment–control comparisons. Assessments were administered online via Gorilla (https://gorilla.sc/; Anwyl-Irvine et al., 2020). Participant flow is presented in the CONSORT diagram (Fig. 1; Schulz et al., 2010).

Intervention programme

RILL is a highly structured and scripted 15-week (20-h) literacy intervention. It is based closely on previously validated programmes developed by Hulme and colleagues, particularly the Reading with Vocabulary Intervention (REVI; Duff et al., 2008) and the Reading Intervention (Bowyer-Crane et al., 2008; Hatcher et al., 1994,

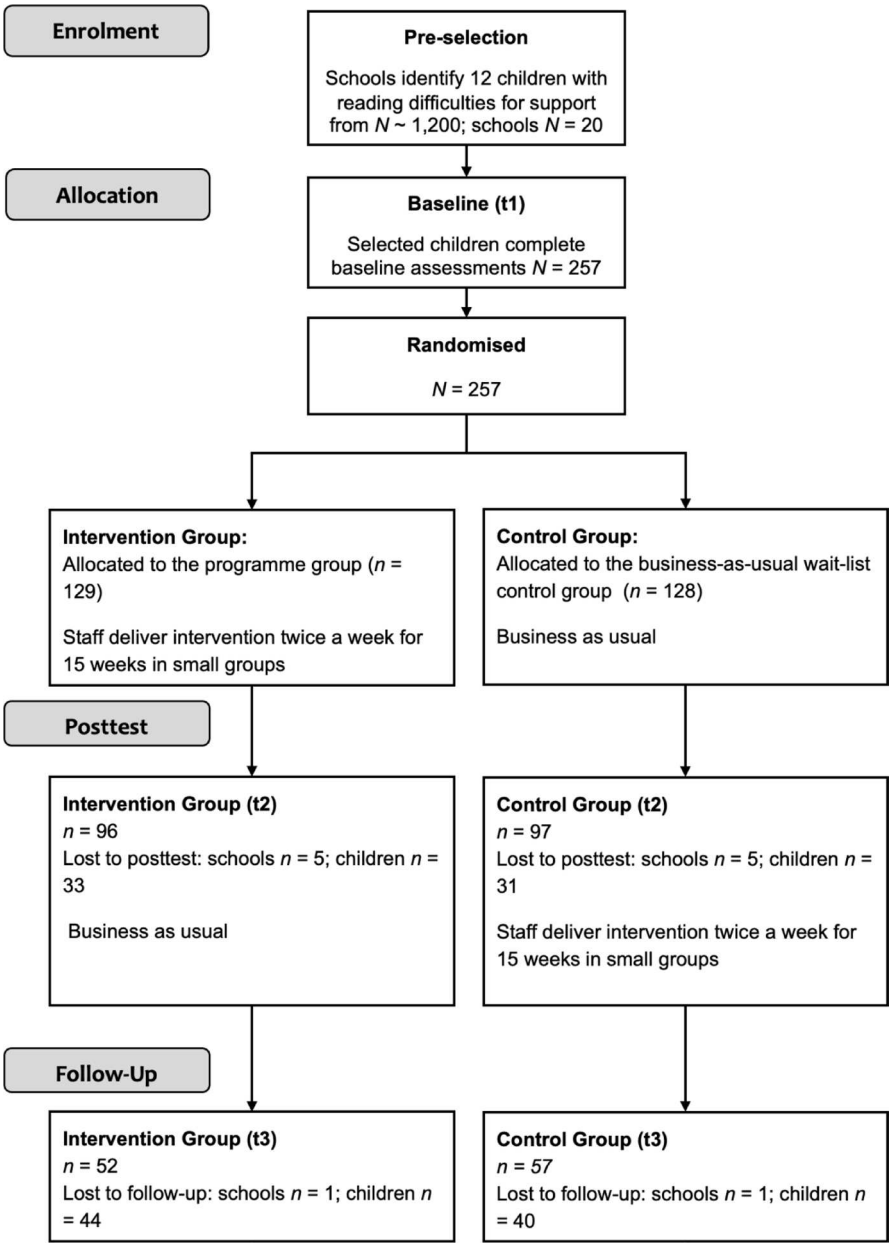


Fig. 1 CONSORT diagram showing the flow of participants through the trial

2004) and previous iterations of RILL (Downing et al., 2025). These programmes emphasise explicit instruction in vocabulary and phonological awareness, combined with tightly structured reading instruction in which word- and text-level materials are carefully matched to the learner's current reading level. Lessons are brief but frequent, consistent with the National Literacy Strategy (2000) and theoretical per-

spectives on the importance of distributed learning for memory consolidation (e.g., Latimier et al., 2021).

RILL targets foundational literacy skills—particularly vocabulary and visual word recognition—appropriate for a wide range of abilities in children aged 7–9 years. Language-rich interventions that focus on vocabulary development and decoding skills form the foundation for strategic and automatic reading comprehension (e.g., Gough & Tunmer, 1986). RILL was designed as a distilled literacy programme targeting empirically validated reading-related skills, drawing on proven instructional techniques and adapting them for online delivery (c.f., Downing et al., 2025).

This version of RILL was developed and piloted in the Welsh language over a period of six months in collaboration with practitioners from North Wales schools. The Welsh programme contains the same components as the English programme, but it is sensitive to the idiosyncrasies of Welsh, such as complex past tense rules, plurals and morphosyntactic mutation (Ball & Müller, 1992; Ball & James, 1984). RILL was delivered in Welsh over 15 weeks. Small groups (3–4 children) received two 45-min lessons per week via Microsoft OneNote, led by trained teaching assistants (TAs). Training comprised a full day on the science of reading development, the rationale for the RILL programme, and implementation of the programme. Ad hoc support was provided throughout the trial. Each lesson followed a structured format and examples are depicted visually in Appendix 1:

1. *Vocabulary Instruction (*5 min*)*: Two Tier 2 words[1] were explicitly taught (Beck et al., 2012) via an animated video of the word meaning and use of the word in two different sentences. Thus, word meaning was conveyed in visual, oral and written form, and in a larger semantic (sentential) context.
2. *Passage Reading (*10 min*)*: Passages incorporating taught vocabulary (see previous section) were read aloud. The instructor asked 2–3 comprehension questions, which probed literal, inferential and evaluative comprehension of the passage. Instructors also asked questions on vocabulary meaning in context as well as the author's purpose and perspective. Children were also encouraged to make connections with their own experiences.
3. *Phonological and Phonics Skills (*5–10 min*)*: The *Word Games* component began with a pre-recorded video that modelled mouth movements to support speech production and articulation. This was followed by two interactive phonological awareness games designed to develop sound decoding and phonemic awareness skills. The games included a range of activities targeting sound repetition, rhyming, blending, segmenting, and phoneme deletion. To ensure consistent delivery of target sounds, high-quality audio clips were used throughout, and a help option was available for each item. After which, children completed a phonics activity which included an introduction to the target grapheme-phoneme correspondence, followed by a short blending activity.
4. *Spelling (*5–10 min*)*: Spelling involved first training letter-sound knowledge (if needed) and then focusing on specific vowel-spelling patterns and consonant

[1] Tier 2 words are high utility words that appear often in mature language. They include words such as *contrast*, *fortunate*, *analyse*, etc. (Beck et al., 2001).

digraphs. Where possible, we incorporated items previously used in the preceding vocabulary and phonemic awareness section to provide opportunities for exposure to the same word in multiple contexts (c.f., Graham et al., 2018).

5. *Narrative Skills (*10 min*)*: Children developed narrative elements (e.g., adjectives, verbs, sequencing) cumulatively, culminating in story writing (Clarke et al., 2010).
6. *Recap (*5 min*)*: Lessons concluded with a review of the vocabulary words.

The intervention was delivered by TAs with training (see above) and support from the authors.

Procedure

Following written caregiver consent, children underwent baseline testing which took approximately 1 h to complete. Teaching Assistants (TAs) were trained to administer identical tests at pre-test, posttest and follow-up (t1, t2, t3). All tests were scored independently and 15% of data were re-scored to assess inter-rater reliability (ICCs > 0.88).

Teaching assistants (TAs) who undertook the RILL training then implemented the intervention in their schools from early 2023 onwards. The assessments at all three time points were administered by trained Research Assistants (RAs), all of whom were blind to each child's treatment condition, using Gorilla https://gorilla.sc/. Assessments for t2 commenced immediately following the intervention (15-week from baseline to posttest) and assessments for t3 commenced following a further three-month delay. RAs observed at least one lesson and assessed fidelity using a premade checklist indexing adherence to lesson structure, delivery of core components, and pacing. Across observed sessions, approximately 80% of lesson components were delivered as intended. This level of fidelity reflects generally strong, though not perfect, adherence to the scripted programme and is comparable to fidelity levels reported in other school-based literacy interventions conducted under real-world classroom conditions (Iowa Reading Research Centre, 2024; O'Connor et al., 2010; Vaughn et al., 2000).

Measures

The primary outcome measures were word reading efficiency in Welsh and English at all three time points, measured via the Test of Word Reading Efficiency 2nd edition (TOWRE-2; Torgesen et al., 2012). Word reading efficiency is the ability to recognize written words accurately and rapidly (Perfetti et al., 2007; Torgesen et al., 2012). English word reading involved reading as many words as possible in 45 s from a list displayed in columns. The Welsh test items were constructed according to the principles of the English TOWRE-2: Welsh items are graded in difficulty progressing from high-frequency, orthographically and phonologically simple words (e.g., CV and CVC) to low-frequency, more orthographically complex words. Complexity increases as the test progresses to include longer words, irregular spellings and morphological complexity specific to the Welsh language (Ball & Müller, 1992; see

Appendix 2 for the full item list). The English TOWRE-2 has strong reported reliability (alternate-form reliability coefficients typically exceeding 0.90; Torgesen et al., 2012). For the Welsh adaptation used in this study, internal consistency at baseline was high (Cronbach's $\alpha = 0.98$; see Table 1).

We also collected data on expressive vocabulary (Welsh language only) at all three time points: A short expressive vocabulary task assessed knowledge of the vocabulary explicitly taught over the course of the programme. A subset of 18 words drawn from the 60 words taught during the RILL lessons were spoken to the child (different subsets were presented to each child). The child was then required to briefly define the meaning of each word. Children's definitions were scored on a 3-point scale (0 incorrect, or no response; 1 partially correct definition; 2 fully correct definition).

Results

Analysis plan

We assessed the effects of the intervention delivered in Welsh, on Welsh and English word reading skills as the primary outcome variable. Word reading efficiency (TOWRE-2; Torgesen et al., 2012), was measured at pre-test (t1), posttest (t2), and 3-month follow-up (t3). A post-hoc analysis split the sample into children whose home language is Welsh (L1) and children whose home language is English (L1) to examine any differences in the pattern of findings on account of native language status (see supplementary analyses). We also assessed changes in children's Welsh expressive vocabulary at pre-test (t1), posttest (t2), and 3-month follow-up (t3) to assess the efficacy of the intervention on these children's vocabulary as a proxy of oral language skill immediately following the intervention, and at follow up. For each outcome, we report Cohen's *d* based on the adjusted posttest mean difference from

Table 1 Descriptive statistics of welsh and English word reading, and welsh vocabulary at baseline (t1), post-test (t2) and follow-up (t3)

	Intervention				Wait-list Control				Reliability
	n	*M*	*SD*	Range	*n*	*M*	*SD*	Range	
Age		111.44	12.23			110.62	12.34		
Welsh Word Reading									
t1	127	33.50	16.20	0–64	124	33.12	15.40	1–65	.98[a]
t2	73	41.84	14.95	0–68	74	40.24	16.26	0–70	
t3	51	47.90	14.90	7–78	54	45.61	15.45	1–75	
English Word Reading									
t1	127	45.71	19.66	2–72	125	42.75	18.67	0–74	.98[a]
t2	73	54.78	16.65	15–79	74	51.05	17.49	1–78	
t3	50	58.26	15.83	9–81	54	54.72	15.83	4–79	
Welsh Expressive Vocabulary									
t1	127	29.03	8.36	11–52	124	27.95	7.74	10–48	.95[a]
t2	73	35.89	10.18	10–54	74	30.56	8.14	14–49	
t3	50	37.02	8.66	19–55	58	33.86	10.64	4–53	

[a]Cronbach's alpha at baseline

the ANCOVA model divided by the pooled raw standard deviation at the corresponding posttest timepoint. This provides a conservative estimate of the size of the treatment–control difference on the posttest scale and is consistent with common practice in educational intervention research (e.g., What Works Clearinghouse, 2014). As an additional exploratory analysis, we fitted a joint longitudinal mixed-effects model that simultaneously modelled Welsh and English reading across time (see Supplementary Analyses for further details).

Analyses examining the effect of the intervention on word reading outcomes followed an intention-to-treat approach. ANCOVA models were run in Stata 18.5 (Stata Corporation, 2023). In these models, posttest scores (t2/t3) were the dependent variables, group (0 = control, 1 = intervention) was the independent variable, and pre-test scores (t1) served as the covariate. No significant group-by-covariate interactions were found, so models are reported without this term. Means for each group are provided with each model. Data and analysis files are available at: https://osf.io/u5trk/?view_only=8bc35d1535bb4430844ac0e8cbfc7fe8.

First, we tested whether the intervention group made greater improvement in word reading by comparing adjusted posttest scores, controlling for baseline. Second, we examined longer-term effects by comparing adjusted follow-up scores between groups; again controlling for baseline.

Between t1 and t2, five schools (64 participants, 25%) withdrew, all citing staffing shortages or insufficient teacher/teaching assistant capacity to deliver the intervention. Importantly, no school withdrew due to dissatisfaction with the intervention itself, or baseline Welsh ($d = 0.19$, $p = 0.118$) or English ($d = 0.06$, $p = .657$) reading ability. These effect sizes represent differences in baseline reading scores between children retained in the study and those lost to attrition.

By the t3 follow-up an additional 84 children were lost: 26 transitioned to high school, 42 moved schools or were absent, and one further school withdrew due to staffing capacity. Attrition reflected structural/administrative factors rather than issues related to the intervention. As with t2 attrition, pupils who were present at t2 but not followed at t3 did not differ from those retained at baseline age ($d = 0.09$, $p = .860$), or t2 Welsh and English reading scores ($d = 0.10$–0.19, all $p > .05$). Within the intervention group specifically, children who were lost to follow-up did not show weaker gains between t1 and t2 that those who remained in the study in Welsh ($d = 0.14$, $p = .188$) or English ($d = 0.18$, $p = .083$) reading, indicating that drop out was not concentrated among children who were likely poor responders. Although follow-up attrition was substantial, the available comparisons suggest that drop out was largely unrelated to participant characteristics or intervention response. Reliabilities and descriptive statistics for all measures are in Table 1.

Effects of intervention on Welsh word reading

The primary outcome measure was Welsh word reading efficiency. To assess the immediate impact of the intervention on children's word reading scores, we ran ANCOVA models, comparing t2 word reading scores between children who completed the intervention and those in the wait-list control whilst controlling for t1 word reading scores. To assess the longer-term effects of the intervention on children's

word reading, we ran ANCOVA models, comparing t3 TOWRE word reading scores between children who completed the intervention and those in the wait-list control whilst controlling for t1 word reading scores. The t3 time point was on average 3-months after the intervention group completed the RILL intervention and after the wait-list control group had received the intervention. As such, this analysis provides a contrast between the gains made immediately and later and serves as an indication of the likely sustained impact of the intervention.

Immediate Effects on Welsh Word Reading. We compared t2 Welsh reading scores between children who had received the intervention and those in the wait-list control group, whilst controlling t1 Welsh reading (baseline) scores. The intervention group had significantly higher adjusted posttest scores ($M=42.70$; mean gain = 8.34) than the wait-list control ($M=39.49$; mean gain = 7.12). The difference in marginal means from the ANCOVA model = 3.22, 95% CI [0.25, 6.19], $t=2.14$, $p=.034$, $d=0.21$, constituting an educationally significant gain (Kraft, 2020; What Works Clearing House, 2014).

Long-term Effects on Welsh Word Reading. We compared t3 Welsh reading scores between children who had received the intervention first (intervention group) and those who had just finished receiving the intervention (wait-list control group), whilst controlling for their t1 Welsh reading baseline scores. Because both groups had received the intervention by t3, this analysis does not assess maintenance of a treatment–control effect. Instead, it contrasts outcomes following different instructional sequences: early intervention followed by classroom instruction versus later intervention following an extended period of business-as-usual teaching. The intervention group had significantly higher adjusted follow-up Welsh reading ($M=48.56$; mean gain = 14.40) than waitlist control group ($M=45.00$; mean gain = 12.49: ANCOVA model = 3.56, 95% CI [0.16, 6.96], $t=2.08$, $p=.040$, $d=0.23$). This difference is interpreted as a timing-dependent advantage, suggesting that early intervention may have enhanced children's subsequent responsiveness to classroom instruction.

Effects of intervention on English word reading

The outcome measure here was English word reading efficiency. Although there was a small baseline imbalance between the intervention and control groups ($d=0.15$, $p=.288$), this difference was not statistically significant. Furthermore, our ANCOVA models statistically adjust for any baseline differences, providing an estimate of the treatment effect that accounts for initial group variation. We assessed the immediate and long-term effects of English word reading using the same approach we used to assess the effects of the intervention on Welsh word reading.

Immediate Effects on English Word Reading. We compared t2 English reading scores between the intervention and the waitlist control group, whilst controlling for their t1 English reading baseline scores. The intervention group had significantly higher adjusted posttest English reading scores ($M=54.38$; mean gain = 9.07) than waitlist control children ($M=50.99$; mean gain = 8.3). The difference in marginal means from the ANCOVA model = 3.40, 95% CI [0.01, 6.78], $t=1.98$, $p=.049$, $d=0.20$ and similar to gains in Welsh reading.

Long-term Effects on English Word Reading. We compared t3 English reading scores between children who had received the intervention first (intervention group) and those who had just finished receiving the intervention (wait-list control group), whilst controlling for their t1 baseline scores. Intervention group children did not have significantly higher follow-up English reading scores ($M = 57.57$; mean gain = 12.55) compared with waitlist control children ($M = 55.07$; mean gain = 11.97). The difference in marginal means from the ANCOVA model = 2.50, 95% CI [−2.02, 7.01], $t = 1.10$, $p = .275$, $d = 0.16$, is reported descriptively and cannot be interpreted as evidence of continued treatment advantage. Because both groups had received the intervention by follow-up, this convergence should not be interpreted as a loss of intervention effects, but rather as reflecting delayed intervention in the waitlist group and the absence of continued differential instruction between groups.

Effects of intervention on Welsh expressive vocabulary

The outcome measure here was Welsh expressive vocabulary knowledge (the ability to define spoken words that had been taught in the intervention). We assessed the immediate and long-term effects using the analytic approach described above.

Immediate Effects on Welsh Expressive Vocabulary. We compared t2 Welsh vocabulary scores between children in the intervention group and those allocated to the waitlist control group, whilst controlling for their t1 Welsh vocabulary baseline scores. Children in the intervention group made greater gains in their Welsh vocabulary accuracy ($M = 35.49$; mean gain = 6.82) than children in the wait-list group ($M = 30.96$; mean gain = 2.45). The difference in marginal means from the ANCOVA model = 4.53, 95% CI [2.47, 6.69], $t = 4.36$, $p < .001$, $d = 0.71$.

Long-term Effects on Welsh Expressive Vocabulary. We compared t3 Welsh vocabulary accuracy scores between children who had received the intervention first (intervention group) and those who had just completed the intervention (wait-list control group). We again controlled for children's t1 Welsh vocabulary baseline scores to assess potential sustained, growing, or fade-out effects of the intervention. The Intervention group children made greater gains in Welsh vocabulary ($M = 36.86$; mean gain = 7.84) than waitlist control group children ($M = 34.00$; mean gain = 5.89). The difference in marginal means from the ANCOVA model = 2.85, 95% CI [0.04, 5.65], $t = 2.01$, $p = .047$, $d = 0.44$.

Discussion

The aim of this study was to determine, using an RCT design, whether structured literacy instruction delivered in Welsh would improve Welsh word reading, produce cross-language transfer to English decoding, and whether the timing of intervention influenced longer-term learning trajectories. We also examined these effects in the context of children with poorer reading skills.

The intervention led to immediate improvements in Welsh word decoding at t2. At follow up, children who received the intervention earlier continued to show higher Welsh reading scores than children who received the same intervention later.

Importantly, this pattern does not reflect maintenance of a treatment–control contrast, but rather suggests that early intervention may confer a timing-dependent advantage. One plausible interpretation is that the intervention strengthened foundational decoding skills in a way that allowed children to benefit from subsequent classroom instruction.

A similar pattern was also found for Welsh expressive vocabulary. Notably, receiving the intervention earlier conferred an advantage to growth in vocabulary skills for these bilingual children with weak literacy, aligning with accounts in which early strengthening of core linguistic skills can alter learning trajectories in transparent orthographies (Caravolas et al., 2013; Tobia & Bonifacci, 2015; Torppa et al., 2016).

In English, the untrained language, the intervention yielded short-term improvements in decoding that were comparable in size to the gains observed in Welsh at t2. This provides strong evidence for cross-linguistic transfer, supporting theoretical models of bilingual literacy that posit shared underlying cognitive and phonological processes across languages (Cummins, 1979; Koda, 2008). However, these effects were not sustained at t3, once both groups had received the intervention. This attenuation should not be interpreted as failure to maintain treatment effects, but instead reflects convergence following delayed intervention in the waitlist group and the absence of continued difference exposure.

The pattern of findings therefore differs systematically across the language of instruction (Welsh) and the untrained language (English), In Welsh, the intervention produced a clear causal effect at posttest, with children who received the intervention outperforming those in the waitlist control group. At follow-up, when both groups had received the intervention, children who received it earlier continued to show higher Welsh reading outcomes. This pattern suggests that the timing of intervention matters for outcomes in the language of instruction, likely because these children can make subsequent better use of classroom instruction following the intervention.

Despite being selected for poor reading skills, children in the intervention group showed immediate gains in English decoding at posttest, providing clear evidence of short-term transfer from Welsh to English. This finding aligns with and extends prior work demonstrating that explicit instruction in foundational decoding skills can support cross-language generalisation, even in populations with relatively weak literacy skills (e.g., Kim & Piper, 2019; Kim et al., 2025). From a theoretical perspective, this supports interactive models of bilingual literacy development, in which shared phonological and orthographic processes can be activated through instruction in one language and temporarily support performance in another, even when underlying representations are relatively fragile.

At follow-up, no significant between-group difference was observed in English reading. Convergence in performance at t3 is compatible with an account in which ongoing transfer processes operating in both groups, rather than indicating attenuation or loss of transfer. The distinction between Welsh and English outcomes suggests that timing effects may be more readily observed in the language of instruction, where subsequent classroom input *congruent with the intervention language* enables children to consolidate transferred skills more systematically. Future studies should examine whether linguistic factors may further contribute to this asymmetry. Welsh's relatively transparent orthography may allow early decoding gains to be

more directly used during later instruction, whereas English's deeper orthography may require sustained, language-specific practice for transferred skills to translate into durable performance differences.

These findings are consistent with work by Kim and colleagues in demonstrating that structured literacy instruction can support cross-language transfer of decoding skills (Kim & Piper, 2019; Kim et al., 2025). In the present study, children with reading difficulties showed evidence of transfer from Welsh to English at posttest, despite being selected for poor reading skills. We were also able to show for the first whether the *timing* of intervention confers additional advantages for transfer. We show that receiving the intervention earlier did not lead to enhanced cross-language outcomes at follow-up, even though children in both groups were subsequently exposed to English literacy instruction in school. One plausible interpretation is that early gains achieved through instruction in a transparent orthography may support immediate transfer but are not sufficient, on their own, to alter longer-term transfer trajectories in the absence of continued, language-congruent instructional support. Particularly for children with fragile phonological and orthographic representations.

Supplementary analyses provide additional light on these findings. Multigroup models indicated that neither immediate intervention effects nor timing-related follow-up differences were moderated by home language background, suggesting that cross-language transfer patterns were not driven by native language status. In addition, a joint longitudinal mixed-effects model revealed parallel growth trajectories in Welsh and English across the study period and no evidence of differential intervention-related growth between languages. Together, these analyses support the conclusion that the intervention exerted broadly similar effects across learner profiles and languages.

From a clinical and educational perspective, this pattern of findings suggests that bilingual children with reading difficulties can benefit from structured decoding instruction delivered in the language of schooling, both directly and through short-term cross-language transfer. However, the results also indicate that timing and sequencing of instruction are critical, particularly if the goal is to support longer-term outcomes across both languages. Sequential intervention models—in which stable decoding skills are first established in the primary instructional language, followed by targeted support in the second language—may therefore be more effective than single-phase approaches for children with fragile literacy skills. Such models acknowledge the presence of cross-linguistic transfer mechanisms while recognising their limits in the absence of continued, language-specific instructional support.

While the results are promising, they should be interpreted in light of several limitations. First, attrition between posttest and follow-up reduced the available sample size at t3, limiting statistical power to detect smaller effects and increasing uncertainty around follow-up estimates. Replication in larger samples with more complete longitudinal data is therefore essential to establish the robustness of the timing-dependent patterns observed here.

Second, the use of a waitlist crossover design constrains the interpretation of follow-up effects. Although this design permits strong causal inference at posttest, group comparisons at follow-up reflect differences in the *timing* and *sequencing* of intervention exposure rather than sustained treatment–control contrasts. Conse-

quently, the absence of between-group differences at t3—particularly for English outcomes—cannot be interpreted as evidence of attenuation or loss of transfer, nor can follow-up differences be taken as evidence of maintained treatment superiority. Future studies employing designs with extended untreated control periods, or alternative sequencing structures, would be required to directly evaluate longer-term persistence of intervention effects.

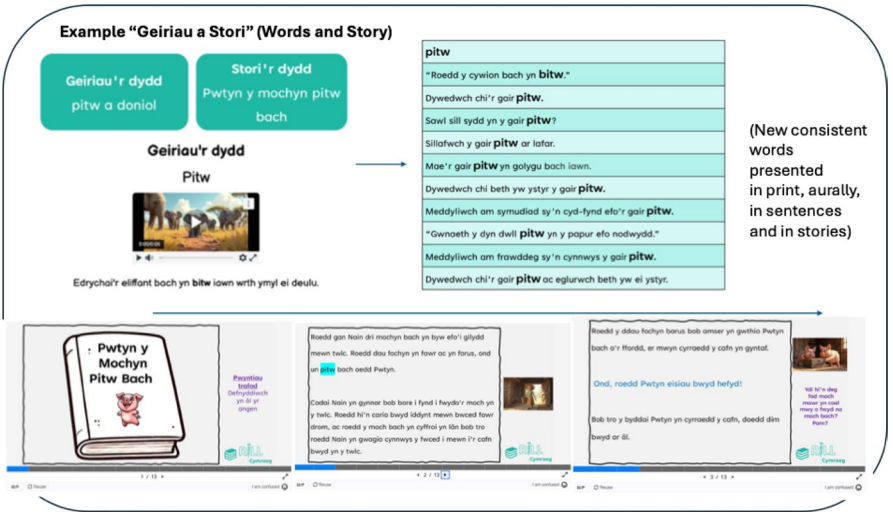
Third, participants were selected on the basis of poor reading skills, which strengthens the clinical relevance of the findings but limits generalisability. It remains unclear whether similar timing and transfer patterns would be observed in typically developing bilingual readers within the same educational context, or whether these patterns reflect boundary conditions specific to children with fragile phonological and orthographic representations.

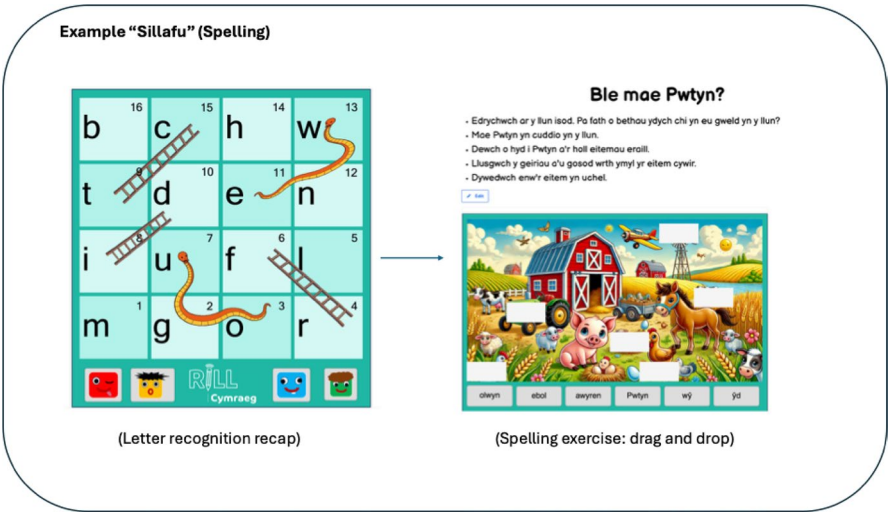
Finally, the intervention was delivered as a multicomponent programme, precluding identification of the specific instructional elements most responsible for cross-language transfer. Future research should use component-level or adaptive designs to isolate the mechanisms underlying transfer and to determine whether targeted reinforcement in the untrained language enhances longer-term cross-language outcomes.

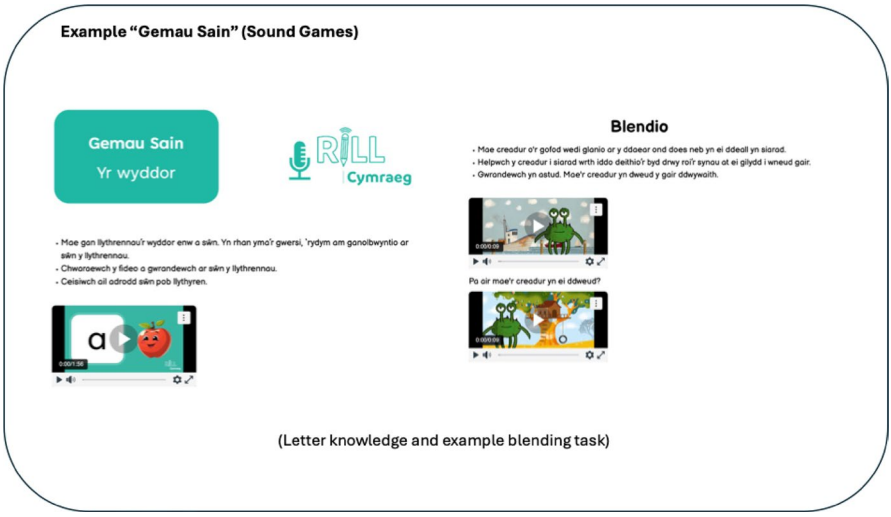
Conclusion

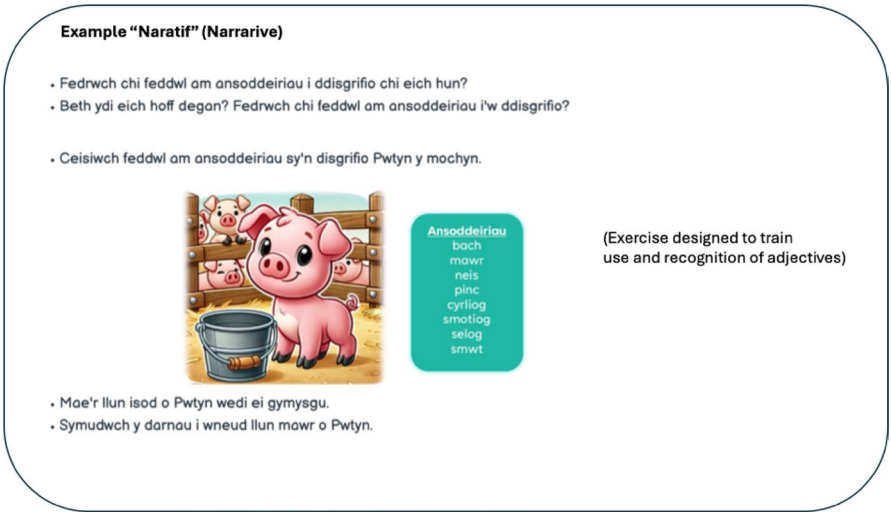
This study provides causal evidence that a structured, monolingual literacy intervention delivered in Welsh can produce immediate improvements in decoding for bilingual children with poor literacy skills, alongside cross-language transfer to English. In the language of instruction, children who received the intervention earlier showed advantages at follow-up consistent with a timing-dependent effect, suggesting that early strengthening of decoding skills may support children's subsequent responsiveness to classroom instruction, when it is delivered in the same language as the intervention. In contrast, the absence of between-group differences in English at follow-up reflects convergence following delayed intervention and shared instructional exposure. Together, these findings underscore the interactive nature of bilingual literacy systems and highlight the importance of considering instructional timing and sequencing when supporting bilingual readers with learning difficulties. Future work should examine how language congruence, and targeted follow-up support shape the longer-term expression of cross-linguistic transfer across languages and learner profiles.

Appendix 1: example visuals of the lessons presented in lesson 1









Appendix 2

See Table 2.

Table 2 Items used in the Welsh adapted version of the TOWRE-2

Welsh item	English translation	Welsh item	English translation	Welsh item	English translation	Welsh item	English translation
na	no (adv.)	sydd	who/which is (v.)	nifer	number (n.m.f.)	sefyllfa	scenario, situation (n.f.)
at	to, by, towards (prep)	beth	what (adv.)	tebyg	similar (adj.)	dywedodd	said (v.)
da	good (n.m.)	tri	three (adj.)	gwaith	job, work (n.m.)	diweddar	late, recent (adj.)
fi	me (pron.)	peth	thing (n.m.)	hanner	half (n.m.)	ystyried	to consider, reflect (v.)
mae	is (v.)	math	brand, sort (n.m.)	gwybod	knowledge, studies (n.m.), to know (v.)	athrawon	teaching staff (n.pl.)
ond	but (conj.)	bwyd	food (n.m.)	derbyn	receiving, to accept, approve (v.)	diwethaf	final, last, ultimate (adj.)
gan	from, by (prep.)	creu	to create (v.)	heddiw	today (adv.)	adroddiad	account, narration, recital (n.m.)
fel	as, like (conj.)	llai	fewer, less, smaller (adj.)	addysg	education, schooling (n.f.)	defnyddio	to adopt, deploy, exploit, use (v.)
neu	or (conj.)	tref	town (n.f.)	cafodd	to get, to have (v.)	cyffredin	accepted, common, general (adj)
dim	any, no, nothing (n.m.)	gwyn	white (adj.)	edrych	to check, examine, look (v.)	cynllunio	planning, design, layout (v.)
lle	area (n.m.)	tair	three (feminine form; adj.)	diwedd	end, ending (v.)	gwybodaeth	information, knowledge (n.f.)
dod	to come (v.)	cwbl	all, everything (n.m.)	cenedl	nation, race (n.f.)	cymedithas	association, guild, society (n.f.)
heb	without (prep.)	marw	dead (n.m.) to die (v.) dead (adj.)	ymateb	feedback, reaction, response (n.m.), answer, respond (v.)	gwasanaeth	service, use (n.m.)
tro	turn, time, bend (n.m.)	bore	morning (n.m.)	unwaith	once (adv.)	disgyblion	pupil, disciple (n.m.)
byw	live (v.) alive (adj.) life (adj.)	iddi	to her (v.)	wythnos	week (n.f.)	dechreuodd	started (v.)

Table 2 (continued)

Welsh item	English translation	Welsh item	English translation	Welsh item	English translation	Welsh item	English translation
oes	age (n.f.)	cadw	to keep, preserve, retain (v.)	efallai	maybe, perhaps (adv.)	eisteddfod	session (n.f.)
efo	with (prep.)	codi	to arise, build, charge, issue, life, set up (v.)	aelodau	limb, member (n.m.)	cyffredinol	common, general, universal (adj.)
mor	as, so (adv.) sea (n.m.)	roedd	was (adv.)	ychydig	very little (n.m.), slight (adj.)	llywodraeth	government (n.f.)
deg	ten (adj.)	gweld	to see, view (v.)	sicrhau	to affirm, assure, confirm (v.)	cymdeithasol	social (adj.)
tua	approx. (adv.) some (adv.)	iaith	language (n.f.)	cyrraedd	to arrive, attain, meet (v.)	cenedlaethol	national (adj.)
enw	name (n.m.)	plant	children (n.f.)	gweithio	to work (v.)		
oedd	was (adv.)	dweud	to indicate, say, state (v.)	rhywbeth	something (n.m.)		
mewn	in (prep.)	rhwng	between (prep.)	blwyddyn	year (n.f.)		
cael	to get, receiving, to have (v.)	merch	daughter (n.f.) female (n.) girl (n.f.)	cyfarfod	gathering (n.), meeting (n.m.) to meet (v.)		
mynd	to go (v.)	corff	body (n.m.)	datblygu	to develop (v.)		
iawn	right (n.m.) accurate (adj.)	gallu	ability (n.) capability, capacity, power (n.m.) to be able (v,)	gwahanol	different (adj.), distinctive (adj.)		
mawr	big (adj.)	teulu	family (n.m.)	ysgolion	ladder, school (n.f.)		
rhai	ones (n.m.) some (pron.)	golwg	appearance (n.) glance, look, sight (n.m.f.)	arbennig	special (adj.)		
bydd	will (adv.)	angen	destitution, necessity, need, want (n.m.)	oherwydd	because (conj.)		

Note that English translations are only supplied for the readership but were not presented in the test

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11145-026-10795-4.

Acknowledgements We wish to thank the children, parents, and schools who took part in the project. We also wish to thank Mererid Owen for her support in developing materials, Abi Jones and Helena Griffith for their support in testing and scoring, and Dr Richard Watkins for his help in recruitment.

Author contributions MJ, and CD secured funding, and MJ, CD and CH oversaw the design and implementation. GEJ conducted project management and GEJ and TH carried out much of the data collection and day-to-day project work. CD conducted the data analysis. MJ took the lead in writing the manuscript, with contributions from CD and CH. All authors reviewed and approved the final version of the manuscript.

Data availability Data and analysis files are available at: [https://osf.io/u5trk/?view_only=8bc35d1535bb4430844ac0e8cbfc7fe8].

Declarations

Conflict of Interest The authors declare no competing interests.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Beck, I. L., McKeown, M. G., & Kucan, L. (2001). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., Götz, K., & 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.

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science, 24*(8), 1398–1407.

Caravolas, M., Lervåg, A., Mikulajova, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading, 23*(5), 386–402.

Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics, 50*, 149–161.

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

Comeau, L., Cormier, P., Grandmaison, É., & Lacroix, D. (1999). A longitudinal study of phonological processing skills in children learning to read in a second language. *Journal of Educational Psychology, 91*(1), 29–43.

Department for Education and Employment. (2000). *The National Literacy Strategy: Grammar for writing*. Department for Education and Employment.

Downing, C., Evans-Jones, G., Calabrich, S. L., Wynne, C., Cartin, R., Dunton, J., Elliott, R., Caravolas, M., Hulme, C., & Jones, M. (2025). Literacy instruction from afar: Evidence for the effectiveness of a remotely delivered language-rich reading programme. *Reading and Writing, 38*(1), 77–91.

Duff, F. J., Fieldsend, E., Bowyer-Crane, C., Hulme, C., Smith, G., Gibbs, S., & Snowling, M. J. (2008). Reading with vocabulary intervention: Evaluation of an instruction for children with poor response to reading intervention. *Journal of Research in Reading, 31*(3), 319–336.

Durgunoğlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*(3), 453–465.

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of Educational Research, 88*(2), 243–284.

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

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry, 45*(2), 338–358.

Iowa Reading Research Center. (2024). *How does fidelity of implementation relate to student reading outcomes?* University of Iowa. https://irrc.education.uiowa.edu/blog/2024/11/how-does-fidelity-implementation-relate-student-reading-outcomes.

Kim, Y. S. G., & Piper, B. (2019). Cross-language transfer of reading skills: An empirical investigation of bidirectionality and the influence of instructional environments. *Reading and Writing, 32*(4), 839–871.

Kim, Y. S. G., Stern, J., Mohohlwane, N., & Taylor, S. (2025). Instruction influences cross-language transfer of reading skills: Evidence from a longitudinal randomized controlled trial. *Reading and Writing, 38*(1), 171–194.

Koda, K. (2005). Learning to read across writing systems: Transfer, metalinguistic awareness, and second language reading development. In K. Koda & A. Zehler (Eds.), *Learning to read across writing systems* (pp. 311–334). Routledge.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253.

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review, 33*, 959–987.

O'Connor, R. E., Harty, K. R., & Fulmer, D. (2010). Improving reading outcomes for struggling first graders: The efficacy of Tier 2 intervention. *Journal of Learning Disabilities, 43*(4), 322–338.

Perfetti, C. A., Landi, N., & Oakhill, J. (2007). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch13

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Journal of Pharmacology and Pharmacotherapeutics, 1*(2), 100–107.

Spencer, L. H., & Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology, 94*(1), 1–28.

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing, 28*, 939–957.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency* (2nd ed.). Pro-Ed.

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

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2000). Why intensive instruction improves reading outcomes for students with reading difficulties. *Reading Research Quarterly, 35*(4), 426–441.

Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics, 28*(3), 425–439.

What Works Clearinghouse. (2014). *Interpreting effect sizes of education interventions*. U.S. Department of Education. https://ies.ed.gov/ncee/wwc/.

Zhang, D., Chin, C. F., & Li, L. (2017). Metalinguistic awareness in bilingual children's word reading: A cross-lagged panel study on cross-linguistic transfer facilitation. *Applied Psycholinguistics, 38*(2), 395–426.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29.

Koda, K., & Reddy, P. (2008). Cross-linguistic transfer in second language reading. *Language Teaching*, *41*(4), 497-508.

Authors and Affiliations

Manon Jones[1] · Gwennant Evans-Jones[1] · Tesni Huws[1] · Catherine Clark[1] · Charles Hulme[2,3] · Cameron Downing[4]

✉ Manon Jones
manon.jones@bangor.ac.uk

Gwennant Evans-Jones
g.m.evans@bangor.ac.uk

Tesni Huws
tsh19dnj@bangor.ac.uk

Catherine Clark
c.clark@bangor.ac.uk

Charles Hulme
charles.hulme@education.ox.ac.uk

Cameron Downing
cameron.downing@york.ac.uk

[1] Bangor University, Bangor, UK

[2] OxEd and Assessment, Oxford, UK

[3] School of Psychology, Oxford Brookes University, Oxford, UK

[4] Department of Education, University of York, York, UK