



Reciprocal relationships among reading and vocabulary over time: a longitudinal study from grade 1 to 5

Callula Killingly¹ · Linda J. Graham¹ · Haley Tancredi¹ · Pamela Snow²

Accepted: 11 February 2024
© The Author(s) 2024

Abstract

Reading comprehension is contingent on both oral language comprehension and word-level reading ability, skills that are thought to be intrinsically related in the early school years. However, while previous studies examining bidirectional relationships among oral vocabulary and reading development have generally found an association between word recognition and subsequent vocabulary skills, they have yielded inconsistent results regarding whether vocabulary is linked to later word recognition. In this longitudinal study ($n = 176$), we assessed reciprocal relationships between expressive and receptive vocabulary and word recognition measured yearly from Grade 1 to 5. We compared results from a cross-lagged panel model (CLPM) to a random-intercept cross-lagged panel model (RI-CLPM), an analytic approach which permits disaggregation of variance between children in the constructs of interest from year-to-year fluctuations within children. Expressive and receptive vocabulary were examined in separate models. Results showed that the RI-CLPMs provided a better fit to the data in contrast to the CLPMs, and provided evidence of reciprocal relationships between word recognition and both receptive and expressive vocabulary. Significant between-child variance was shown for vocabulary but not word recognition. These findings confirm that reading and vocabulary skills build upon one another across the primary years, and each is an important area of classroom instruction focus. Early and regular screening measures may serve to inform the development of timely and appropriate support and intervention.

Keywords Vocabulary · Word recognition · Longitudinal · Reciprocal · Reading

✉ Callula Killingly
c.killingly@qut.edu.au

¹ Centre for Inclusive Education, Faculty of Creative Industries, Social Justice and Education, Queensland University of Technology (QUT), Level 3, E Block, Kelvin Grove Campus, Brisbane, QLD 4059, Australia

² School of Education, La Trobe University, Victoria, Australia

Introduction

Reading is a foundational skill acquired in the early years of school and necessary for ongoing access to the curriculum throughout school, as well as effective participation in employment and civic engagement (Graham et al., 2020; Snow, 2020). Successful reading comprehension is underpinned by the product of effective word recognition and oral language comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990; Hoover & Tunmer, 2018); hence, the beginning reader must learn to both decode alphabetic symbols and store their representations in long term memory via a process known as orthographic mapping (Ehri, 2022) and apply their developing oral language comprehension skills in order to gain meaning from text. While most often recognised for its contribution to reading comprehension, oral vocabulary has also been shown to play a key role in word-level reading processes, and the relationship among these processes has been the subject of extensive longitudinal, cross-sectional, and experimental research (for review, see Wegener et al., 2022). Research shows that early vocabulary knowledge for children aged 16–24 months predicts later foundational reading skills (Duff et al., 2015), and studies of school-age children typically find that vocabulary knowledge explains significant variance in word-level reading skill (Ouellette, 2006; Scarborough, 2001).

Several complementary theories have been proposed to explain why oral vocabulary supports reading development (Wegener et al., 2022). The lexical quality hypothesis describes knowledge of a word as being represented at the phonological, orthographic, and semantic levels, each of which may differ in the quality of representation (Perfetti & Hart, 2001). Having an extensive oral vocabulary is thought to support the individual's word-level reading by helping them to establish and enhance the links between a word's phonology and its orthography. Strong oral vocabulary has also been proposed to support word reading by allowing the reader to self-correct mispronunciations using semantic knowledge (Dyson et al., 2017; Tunmer & Chapman, 2012). Another mechanism by which vocabulary is thought to support reading applies even prior to print exposure, at which point it is hypothesised that children may develop what are termed 'orthographic skeletons', where they anticipate the likely spelling of a word based on their knowledge of the relationship between speech sounds and their corresponding spelling patterns (Wegener et al., 2018).

Reading skills, as they develop, may then serve to bolster vocabulary development. By acquiring knowledge of phoneme-grapheme correspondences throughout the early years of school, children are equipped to decode written words with increased automaticity (fluency and accuracy), skills which enable them to read more widely with ease and provide a mechanism by which they strengthen and extend their word knowledge (Cain & Oakhill, 2011; Duff et al., 2015b). Proficiency in word-level reading frees up cognitive resources, allowing the individual to focus more on overall text meaning and infer context-specific meaning from words (Cunningham & Stanovich, 1991; Perfetti, 2010; Smith et al., 2021; Verhoeven et al., 2011). The relationship between reading and vocabulary development is thought to increase in strength over time: initially, children learn new words largely through oral, rather than written exposure, as early reading materials include simpler words that are already familiar; however, as children progress through school, they are exposed to many

more new words via print as they read texts of increasing complexity (Georgiou et al., 2023; Verhoeven et al., 2011). Therefore, word-level reading and vocabulary appear to be reciprocally related: oral vocabulary supports the development of reading skill, while being a proficient reader facilitates the acquisition of new words (Nippold, 2016; Perfetti, 2010).

Longitudinal studies examining the co-development of word-level reading skills and vocabulary knowledge have yielded inconsistent results (see Table 1). Two relevant early studies focused on the contribution of oral language to later word reading ability. As part of a two-year longitudinal study, Muter and colleagues (2004) investigated receptive vocabulary at school entry (4 years 9 months) and its relationship with later word recognition (6 years 9 months) in a sample of 90 English-speaking children and did not observe a significant relationship. Nation and Snowling (2004) examined the development of reading skills across a wider span of time in 72 English-speaking children. Specifically, they assessed the variance explained by early oral language skills (Time 1; age 8.5 years) on later word recognition (Time 2; age 13 years). In contrast to Muter and colleagues, Nation and Snowling found that vocabulary explained unique variance in later word recognition, even when controlling for word recognition skills at Time 1 and decoding skills at Time 1 and 2.

Several other studies have examined reciprocal relationships among vocabulary and word reading over time. Notably, three of the five studies were conducted in Chinese, a morphosyllabic language. Hulme and colleagues (2018) assessed these skills at five timepoints across 6-month increments between Grades 1 to 3, in a sample of 143 Mandarin-speaking children. Results from latent growth curve modelling indicated that early word recognition predicted growth in early and later expressive vocabulary knowledge, while vocabulary did not predict the growth in word recognition. Dulay and colleagues (2021) obtained similar results in a sample of Grade 1 students ($n=160$) across a different timeframe (one year apart) and using different analytic techniques (cross-lagged panel model), by which it was determined that vocabulary did not predict later word recognition, but word recognition predicted later vocabulary. In contrast, Yan and colleagues (2021) assessed vocabulary and word recognition (reading Chinese characters) once each year from Grades 1 to 3 in a sample of 186 children, finding that vocabulary predicted character recognition from Grade 1 onwards, while character recognition also predicted vocabulary from Grade 2 onwards. Notably, the expressive vocabulary measure used by Yan et al. was a more complex task targeting word definitions, thereby providing a deeper assessment of lexical knowledge, in comparison to the other two studies, where the measures of vocabulary required picture naming only.

To date, two studies have examined the co-development of vocabulary and reading in an alphabetic orthography. In a large longitudinal study of Dutch-speaking children ($n=2,790$), Verhoeven and colleagues (2011) tracked the development of reading skills each year from Grade 1 to 6, with seven timepoints of word recognition and eight timepoints of receptive vocabulary (additional timepoints within Grade 1 and 2). Word recognition predicted later vocabulary from Grades 2 to 3 and 4 to 5, while vocabulary predicted later word recognition within the Grade 1 and 2 timepoints. Most recently, Georgiou and colleagues (2023) conducted a study with English-speaking children ($n=172$) each year from Grades 1 to 3, examining the

Table 1 Prior studies¹ investigating longitudinal and/or reciprocal relationships among vocabulary and reading development

Study	N	Age/Grade	Timepoint (interval)	Language	Analysis	Measures	Vocabulary predicts later word recognition	Word recognition predicts later vocabulary
Muter et al. (2004)	90	4yrs 9mths, 6yrs 9mths	2* (2 yrs)	English	SEM	RV; WR	No	Not assessed
Nation and Snowling (2004)	72	8.5yrs, 13yrs	2 (4.5yrs)	English	HR	EVa; WR	Yes	Not assessed
Verhoeven et al. (2011)	2790	Grades 1–6	8; 7**	Dutch	CLPM	RV; WR	Yes – Grades 1–1; 2–2	Yes – Grade 2–3; 4–5
Hulme et al. (2018)	143	Grades 1–3	5 (6mth-ly)	Chinese	LGCM	EVa; WR	No	Yes – predicts growth
Dulay et al. (2021)	160	Grade 1	2 (yearly)	Chinese	CLPM	EVa & RV; WR	No	Yes
Yan et al. (2021)	186	Grades 1–3	3 (yearly)	Chinese	CLPM	EVb; WR	Yes – Grade 1–2; 2–3	Yes – Grade 2–3
Georgiou et al. (2023)	172	Grades 1–3	3 (yearly)	English	CLPM	EVb; WR & NWR	No	Yes – Grade 1–2; 1–3

HR=Hierarchical Regression; SEM=Structural Equation Modelling; CLPM=Cross-Lagged Panel Model; LGCM=Latent Growth Curve Model; EVa=Expressive Vocabulary (labelling task); EVb=Expressive Vocabulary (definition task); RV=Receptive Vocabulary; WR=Word Recognition; NWR=Nonword Reading. *This study included a middle timepoint involving additional measures not relevant to this relationship

**2 timepoints within Grade 1 for WR; 2 timepoints within Grade 1 and 2 for RV

¹These studies have referred variably to word identification, word recognition, and decoding, sometimes distinguishing between the two in terms of the measures used (e.g., Georgiou et al., 2023) and sometimes referring to measures of word recognition as decoding in the broader sense (e.g., Verhoeven et al., 2011). In the present study, we refer to decoding as the conversion of written symbols into sounds, e.g., sounding out words, while word recognition reflects the identification of familiar words, which may be enacted through a process of decoding or orthographic mapping. To isolate knowledge of grapheme-phoneme correspondences, measures of decoding use nonwords (sometimes referred to as pseudowords), while measures of word recognition use high frequency, familiar words.

cross-lagged relations among decoding and expressive vocabulary (using a measure targeting word definitions), and word identification and expressive vocabulary. Like Verhoeven et al., they found that both word recognition and decoding in Grade 1 predicted vocabulary in subsequent grades; in contrast, vocabulary did not predict subsequent word recognition or decoding.

In summary, while it has been consistently shown that word recognition (and decoding of pseudowords; see Georgiou et al., 2023) predicts subsequent vocabulary, controlling for earlier vocabulary skill, there have been mixed findings concerning whether vocabulary predicts later reading throughout the primary school years. There are several plausible reasons for these discrepant findings, such as the difference in sample age and/or time lag between measurements. For example, Nation and Snowling (2004) found vocabulary to predict later reading over a span of 4.5 years; conversely, Muter and colleagues (2004) did not observe a significant relationship, yet their study spanned a period of 2 years and was conducted with a younger sample. The other two studies investigating these relationships in alphabetic orthographies used different timeframes again, with the Grades 1–2 measurements taken six-monthly (Verhoeven et al., 2011) or yearly (Georgiou et al., 2023). Such discrepancies in age and measurement timing potentially have important implications when investigating developmental phenomena such as literacy and oral language skills. There are also differences in the type of vocabulary measures used across studies: while all measures are reflective of vocabulary breadth as opposed to depth (Ouellette, 2006), some studies have used receptive and others have used expressive measures, which could conceivably affect measurement validity and the strength of the cross-lagged relationships. Finally, studies have varied in the type of word reading task, with some using timed tasks (e.g., Verhoeven et al., 2011) and some untimed (e.g., Georgiou et al., 2023; Nation & Snowling, 2004).

Another consideration in the interpretation of these results concerns the analytic approach taken. Most studies have employed a cross-lagged panel model (CLPM), a popular technique frequently used in the examination of reciprocal relationships over time. Given two or more variables, measured across at least two waves of data, the CLPM allows one to assess the association between a variable x at time 1 and a second variable y at time 2, controlling for the prior effects of y (time 1). Simultaneously, one can evaluate the relationship between y at time 1 and x at time 2, controlling for time 1 x . In this way, it is possible to observe the extent to which these variables are co-related over time. In the case of reading and vocabulary, one can therefore see the extent to which reading predicts subsequent vocabulary knowledge, controlling for prior vocabulary knowledge, and vice versa.

Despite its widespread use, the CLPM has been criticised (e.g., Hamaker et al., 2015), due to its inherent conflation of between- and within-unit effects. Between-unit effects are concerned with whether, on average across time, one individual is characterised by higher or lower levels of the construct of interest in relation to other individuals, therefore reflecting stable ‘trait-like’ individual differences (Curran & Hancock, 2021). Within-unit effects reflect temporal changes of the individual in relation to their own average. Without disaggregating between- and within-unit effects, the assumption underpinning the CLPM is that all individuals in the sample vary around the same means over time; hence this approach does not account for

individual differences (Hamaker et al., 2015). In the context of reading and language development, for example, it may be that some children have stronger vocabulary skills than others, on average across time, reflecting between-unit effects. Yet generally, the more pressing question for those studying longitudinal transactional relationships is the relationship between time-to-time individual-level change, or the within-unit effect: has an individual's vocabulary score increased over time in relation to their typical vocabulary level? Further, how does this individual-level change relate to subsequent individual-level change in a second presumably related variable?

While these questions are not answerable using traditional CLPM techniques, recent years have seen development of the random intercept cross lagged panel model (RI-CLPM; Hamaker et al., 2015). Like the CLPM, the RI-CLPM allows examination of how two or more constructs are co-related over time; however, in the RI-CLPM, data are treated as multi-level, such that timepoints are nested within individuals. Under this approach, between- and within-unit effects are decomposed, by incorporating random intercepts to represent stable individual differences in the overall constructs and modelling the autoregressive and cross-lagged parameters at the level of the residuals to reflect individual-level fluctuations over time (Hamaker et al., 2015). The RI-CLPM requires at least three waves of data for the model to be identified and is among several longitudinal modelling techniques that disaggregate between- and within-effects (for review, see Usami et al., 2019; Curran & Hancock, 2021).

Comparisons of the CLPM and RI-CLPM have shown that the cross-lagged effects can change in direction and/or significance once random intercepts have been included, because the temporal relationships being modelled are fundamentally different (Mulder & Hamaker, 2021). Specifically, while autoregressive relationships in the CLPM represent the rank-order stability of individuals between adjacent timepoints for the same construct, much of this stability is captured by the random intercept in the RI-CLPM, and the autoregressive component instead reflects additional year-to-year variation in the construct at the individual level. In the CLPM, the cross-lagged component is based on individual deviation in relation to the sample mean; however, under the RI-CLPM, individuals are instead compared to their own expected scores, permitting an examination of within-person change from year to year (Hamaker et al., 2015). The main advantage of the RI-CLPM, therefore, is the ability to distinguish between-child differences from within-child year-to-year development. This approach is of great use when examining how developmental processes unfold and influence one another over time; hence, it has recently gained traction for similar research questions (e.g., Hwang et al., 2023; Willard et al., 2021).

The present study

In this paper, we report findings from a longitudinal study spanning the first six years of school, in which standardised measures of expressive and receptive vocabulary and word recognition were obtained each year. We investigate whether vocabulary and word recognition are reciprocally related over time from Grade 1 to 5, extending on prior research that examines this question by (1) including measures of both

expressive and receptive vocabulary, and (2) comparing results from a traditional CLPM approach to a RI-CLPM in which between- and within-child effects are disaggregated. We hypothesised that, when modelled using RI-CLPMs, word-level reading and vocabulary would be reciprocally and positively related across all adjacent timepoints, such that reading predicts later vocabulary and vocabulary predicts later reading.

Method

Participants

Participants were recruited from seven government schools in south-east Queensland as part of a 6-year longitudinal study ($n=250$). Students were invited to participate in the study through a letter sent home to their parent or guardian, via their school; all whose parents provided consent participated in the project. While most students were recruited in their first formal year of schooling (Preparatory), some consented to participate in later waves of recruitment across the period. Analyses in the present paper include a subsample of students with English as their first language who completed reading and vocabulary measures in at least one wave from Grades 1 to 5 ($n=176$; see Table 2 for demographic characteristics). Students with a language background other than English were excluded from the sample, consistent with previous studies examining relationships among reading and vocabulary. Of the sample, 10.23% identified as Aboriginal or Torres Strait Islander.

Students were individually tested in each year of the project on a range of measures spanning language, literacy, numeracy, behaviour, and social development. Each measure was administered in a quiet room by a trained research assistant. Research was conducted according to the ethical standards stipulated by the institution and national guidelines and received approval from the Queensland University of Technology (QUT) Human Research Ethics Committee (approval no. 1300000422) and Queensland Department of Education. All children and their parent(s)/guardian(s) provided informed consent to participate in the research.

Measures

Expressive vocabulary

The Expressive Vocabulary Test 2nd Edition (EVT-2; Williams, 2008) is a standardised assessment of oral vocabulary knowledge taking approximately 10–20 min to administer. For each item, students are read a stimulus question and presented with a picture, to which they are asked to respond verbally with a word that provides an

Table 2 Sample demographic characteristics in each wave and age in months at time of assessment

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
% female	56.3	51.7	54.5	54.5	54.0
M_{age} (SD)	79.75 (3.41)	91.86 (3.56)	104.50 (5.77)	116.00 (3.73)	127.52 (3.85)

appropriate label or synonym for the item or provides the answer to a specific question. In accordance with the manual, a 10 s window was provided for participants to provide each response, and stimulus words were only repeated once as required. The EVT-2 has been shown to have good reliability (alternate-form alphas ranging from 0.83 to 0.91; Williams, 2008).

Receptive vocabulary

The Peabody Picture Vocabulary Test 4th Edition (PPVT-4; Dunn & Dunn, 2007) is a standardised measure of receptive vocabulary knowledge. This test is untimed, and generally takes approximately 10–15 min to administer. For each item, the child is asked to indicate which of four images (presented on a coloured card) represent the stimulus word spoken by the administrator. For each form of the test, internal consistency is high (α s=0.94, 0.95), as is alternate-form reliability (α s=0.87, 0.93; Dunn & Dunn, 2007).

Word reading

The Test of Word Reading Efficiency 2nd Edition (TOWRE-2; Torgesen et al., 2012) is a standardised measure of a child's ability to recognise known words (Sight Word subtest) and sound out phonically legal pseudowords fluently and accurately (Phonemic Decoding subtest). For each subtest, words are presented on a laminated A4 sheet in accordance with the examiner's manual. The Sight Word subtest requires students to read as many words as possible within 45 s, while the Phonemic Decoding subtest requires them to read as many pseudowords as possible within 45 s. The TOWRE-2 has good reliability, with alternate-form coefficients of >0.9 on each subtest (Sight Word, 0.91; Phonemic Decoding, 0.92; Torgesen et al., 2012). The Sight Word subtest was used in analyses.

Analytic plan

Using a structural equation modelling framework, a series of cross-lagged analyses were compared to examine the reciprocal relationships between vocabulary and word recognition. In alignment with previous studies, a traditional cross-lagged panel model (CLPM) was first conducted. To account for stable between-person differences, data were also fitted to a random-intercept cross-lagged panel model (RI-CLPM; Hamaker et al., 2015). Comparison of the CLPM to the RI-CLPM was achieved using a chi-bar-square difference test (Hamaker et al., 2015). To determine the most parsimonious model, and to evaluate whether the developmental processes were consistent over time, we systematically added equality constraints to the RI-CLPMs beginning with autoregressive and cross-lagged parameters, and then residual covariances and variances, using chi-square difference tests and the Akaike Information Criterion (AIC) to determine whether constraints were tenable (Mulder & Hamaker, 2021). Maximum likelihood estimation with robust Huber-White standard errors and a scaled chi-square (χ^2) statistic was employed in all analyses (MLR). Model fit was assessed using a range of statistics, where good fit is indicated ideally by a

non-significant χ^2 test, Comparative Fit Index (CFI) > 0.95, Tucker Lewis Fit Index (TLI) > 0.95, Root Mean Square Error of Approximation (RMSEA) < 0.06, and Standardised Root Mean Square Residual (SRMR) < 0.08 (Hu & Bentler, 1999; West et al., 2012). The following effect size values have been proposed as a means of interpreting cross-lagged parameters for both CLPM and RI-CLPM: 0.03 (small); 0.07 (medium); and 0.12 (large; Orth et al., 2022). All analyses were conducted in R with the *lavaan* package (ver. 0.6–15; Rosseel, 2012).

While most students commenced participation during Preparatory, recruitment remained open across all years of the project; hence, some students have data for later waves, while, due to school mobility, some participants only have data during earlier waves, with just over a third of cases missing across the entire sample over all timepoints (37%). Missing data on the relevant variables were determined to be missing completely at random (MCAR; Little's test non-significant, $\chi^2(194) = 221.87$, $p = .084$); therefore, Full Information Maximum Likelihood (FIML) was used in all analyses. FIML permits all available data to be used in analysis, by estimating a likelihood function for each individual based on their available observations. Importantly, FIML has been found to yield unbiased standard errors and parameter estimates when data are MCAR (or missing at random; Enders & Bandalos, 2001). Therefore, all 176 students in the sample were included in analyses.

Marginal distributions were normal according to visual inspection and skewness and kurtosis statistics (see Table 3). We determined the impact of any univariate outliers, as identified based on extreme z scores (± 3.29 ; Tabachnick et al., 2013), finding that each measure (EVT, PPVT, TOWRE Sight Word subtest) contained one outlier in one of the waves (representing three different participants). To assess residual distributions and multivariate outliers (using Mahalanobis' and Cook's distance), multiple regression analyses were conducted for adjacent waves of data (e.g., Grade 3 EVT predicted by Grade 2 EVT and TOWRE Sight Reading subtest). Residuals were generally normal, but some outliers were detected in several waves for models using both EVT and PPVT (standardized residual > 3). Analyses were run with and without outlying participants ($n = 173$ for models using EVT; $n = 170$ for models using PPVT), and, as substantive results were consistent both ways, all reported analyses include the full sample.

Results

Bivariate correlations among all predictors at each time point indicated significant, positive relationships (Table 3). These relationships tended to be slightly stronger between word recognition and expressive vocabulary (EVT), compared to receptive vocabulary (PPVT). Considering cross-domain bivariate associations between adjacent waves (see bolded coefficients in Table 3), the coefficients between word recognition and expressive vocabulary were comparable across the five waves for word recognition and subsequent vocabulary ($r_s = 0.435 - 0.530$), and vocabulary and subsequent word recognition ($r_s = 0.383 - 0.497$). For receptive vocabulary, the associations between word recognition and later vocabulary were generally similar in

Table 3 Bivariate correlations among vocabulary and word recognition

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. EVT t1	--														
2. EVT t2	0.727**	--													
3. EVT t3	0.849**	0.850**	--												
4. EVT t4	0.827**	0.831**	0.837**	--											
5. EVT t5	0.790**	0.817**	0.820**	0.868**	--										
6. PPVT t1	0.775**	0.760**	0.730**	0.716**	0.700**	--									
7. PPVT t2	0.629**	0.750**	0.661**	0.634**	0.654**	0.774**	--								
8. PPVT t3	0.666**	0.641**	0.680**	0.711**	0.728**	0.718**	0.719**	--							
9. PPVT t4	0.725**	0.691**	0.710**	0.743**	0.732**	0.769**	0.831**	0.814**	--						
10. PPVT t5	0.642**	0.621**	0.642**	0.665**	0.729**	0.585**	0.709**	0.724**	0.737**	--					
11. WR t1	0.419**	0.504**	0.452**	0.448**	0.493**	0.349**	0.321**	0.226*	0.273*	0.320**	--				
12. WR t2	0.383**	0.489**	0.466**	0.449**	0.475**	0.331**	0.368**	0.267*	0.306**	0.291*	0.845**	--			
13. WR t3	0.367**	0.403**	0.520**	0.435**	0.438**	0.269*	0.273*	0.371**	0.330**	0.299**	0.747**	0.903**	--		
14. WR t4	0.357**	0.460**	0.497**	0.469**	0.530**	0.313**	0.345**	0.424**	0.451**	0.340**	0.692**	0.843**	0.897**	--	
15. WR t5	0.409**	0.484**	0.499**	0.470**	0.523**	0.357**	0.362**	0.420**	0.417**	0.316**	0.673**	0.809**	0.853**	0.937**	--
<i>n</i>	142	118	112	99	87	142	118	112	99	88	142	118	112	99	88
<i>M</i>	82.88	90.72	97.13	104.58	112.02	105.23	120.16	134.30	146.91	155.10	27.74	42.75	52.11	59.09	64.77
<i>SD</i>	13.34	15.05	14.27	14.09	14.48	17.57	19.20	16.66	15.22	14.38	16.25	17.87	16.84	16.70	15.69
Minimum	46	40	66	67	79	55	27	88	110	117	0	0	4	7	10
Maximum	115	122	135	144	152	158	167	168	192	191	65	73	76	88	92
Skewness	0.08	-0.41	0.39	0.19	0.36	-0.02	-0.90	-0.25	-0.20	0.05	0.49	-0.45	-0.90	-1.19	-1.40
Kurtosis	-0.18	0.55	0.28	0.95	0.27	-0.15	3.82	-0.64	0.42	0.24	-0.71	-0.64	0.14	1.32	2.65

* $p < .05$; ** $p < .01$

magnitude ($r_s=0.267-0.340$) to those of vocabulary and subsequent word recognition ($r_s=0.273-0.424$).

Expressive vocabulary and word recognition

The CLPM yielded a relatively poor fit to the data according to most indicators (e.g., $RMSEA>0.08$; see Table 4 for full model fit statistics). While the poor fit indicates these results should be interpreted with caution, we include a summary here to enable comparison. Results of this model indicated significant, positive autoregressive relationships between consecutive waves for both vocabulary and word recognition. Cross-lagged relationships were also significant between some adjacent waves: word recognition at Grade 1 was positively associated with vocabulary at Grade 2; this relationship was again significant from Grade 2 to 3, and 4 to 5. Vocabulary in Grade 2 predicted word recognition in Grade 3. No other cross-lagged effects were observed (see Fig. 1a for regression parameters).

When modelling these data with a RI-CLPM, the random intercept for word recognition produced negative variance (a Heywood case), potentially indicating an absence of stable between-child effects in reading or simply that the model is too complex for the data; therefore this parameter and its covariance with the random intercept for vocabulary were removed from the model (Mulder & Hamaker, 2021). The model including only a random intercept for vocabulary produced a good fit to the data (see Table 4), significantly better than the CLPM according to a chi-square difference test, $\bar{\chi}^2 = 6.16$, $p<.001$. This RI-CLPM was then compared to models in which autoregressive and cross-lagged components were constrained to equality, with chi-square difference tests indicating that model fit did not decline; hence the constraints were retained. Further constraints to residual variances and covariances resulted in significantly poorer fit; thus, the final model included equality constraints for unstandardised autoregressive and cross-lagged relationships only. Standardised parameter estimates typically still vary over time even when unstandardised estimates are constrained to equality (see Mulder & Hamaker, 2021); hence we report both unstandardised (b ; in-text) and standardised estimates (β ; Fig. 1b) and use the latter to evaluate the strength of the cross-lags in relation to one another.

Autoregressive relationships were generally smaller in magnitude compared to those produced by the CLPM, and particularly for expressive vocabulary, for which relationships between consecutive waves were non-significant for all grades ($b=0.22$, $SE=0.13$, $z=1.75$, $p=.080$). This absence of autoregressive effects is presumably due to these effects being captured by the random intercept for EVT, in which significant variance was observed, indicating stable, between-child differences in vocabulary (variance=108.67, $SE=15.13$, $z=7.18$, $p<.001$). For word recognition, autoregressive effects remained significant across all waves ($b=0.80$, $SE=0.040$, $z=20.18$, $p<.001$), and of a similar magnitude to the CLPM, as expected, given that the between-subject random intercept for word recognition was not included in the model. The RI-CLPM indicated significant bidirectional relationships between vocabulary and word recognition. Word recognition predicted vocabulary in the following year across timepoints ($b=0.32$, $SE=0.06$, $z=5.71$, $p<.001$), with large effects that increased in magnitude over time (all $\beta_s>0.12$). Vocabulary was posi-

Table 4 Model fit statistics for CLPM and RI-CLPM with RI for EVT only

	χ^2 (df)	CFI	TLI	RMSEA	SRMR	AIC	
CLPM	75.62** (24)	0.950	0.907	0.161	0.055	8187.8	
RI-CLPM (RI EVT only)	30.97 (23)	1.000	1.000	0.000	0.033	8141.3	
<i>RI-CLPM equality constraints:</i>							$\Delta \chi^2$ (df)
Autoregressive (AR)	33.51 (29)	1.000	1.000	0.000	0.034	8136.0	4.22 (6)
AR, cross lags (CR)	43.86 (35)	1.000	1.000	0.000	0.076	8133.8	11.02 (6)
AR, CR, residual (co)variances	87.02** (44)	0.986	0.986	0.062	0.094	8160.3	45.49** (9)

* $p < .05$; ** $p < .001$. Retained model in bold

tively associated with word recognition in successive waves ($b=0.22$, $SE=0.07$, $z=3.22$, $p=.001$), albeit with weaker effects of a similar magnitude over time (still medium to large; $\beta_s > 0.07$).

Receptive vocabulary and word recognition

Similar to expressive vocabulary, the CLPM modelling receptive vocabulary and word recognition did not fit the data well according to most metrics (see Table 5). Significant autoregressive relationships were observed for both constructs, and significant cross-lagged associations were observed in both directions in some waves (Grades 2–4 for vocabulary to word recognition; Grades 2–3 for word recognition to vocabulary; see Fig. 2a). A RI-CLPM provided a significantly better fit for the data, $\bar{\chi}^2 = 6.43$, $p < .001$, yielding a non-significant chi-square value and excellent indices of absolute fit. As with the EVT, the random intercept of PPVT explained significant variance, indicating stable between-child differences in receptive vocabulary (variance = 188.99, $SE = 36.56$, $z = 5.16$, $p < .001$). The word recognition random intercept was non-significant (variance = 27.17, $SE = 94.06$, $z = 0.29$, $p = .773$), as was the covariance between the random intercepts (30.78, $SE = 42.74$, $z = 0.72$, $p = .471$); hence, these effects were removed from the model per recommendations (Mulder & Hamaker, 2021).

Equality constraints were then imposed on the RI-CLPM, first to autoregressive parameters, then adding in cross-lagged parameters; these constraints were found to be tenable. However, as with the EVT, further constraints resulted in a significantly poorer model fit, and so the model with constraints only on lagged parameters was retained. In this final model, autoregressive associations were non-significant for vocabulary ($b=0.17$, $SE=0.11$, $z=1.53$, $p=.127$), while for word recognition, they were significant across all grades ($b=0.81$, $SE=0.04$, $z=21.84$, $p < .001$; see Fig. 2b). Cross-lagged associations were significant from PPVT to word recognition ($b=0.19$, $SE=0.06$, $z=3.19$, $p=.001$) with effects of a similar magnitude across time ($\beta_s=0.10$ – 0.13), and for the reverse effect of word recognition to vocabulary ($b=0.28$, $SE=0.06$, $z=4.56$, $p < .001$), with large effects observed ($\beta_s > 0.12$).

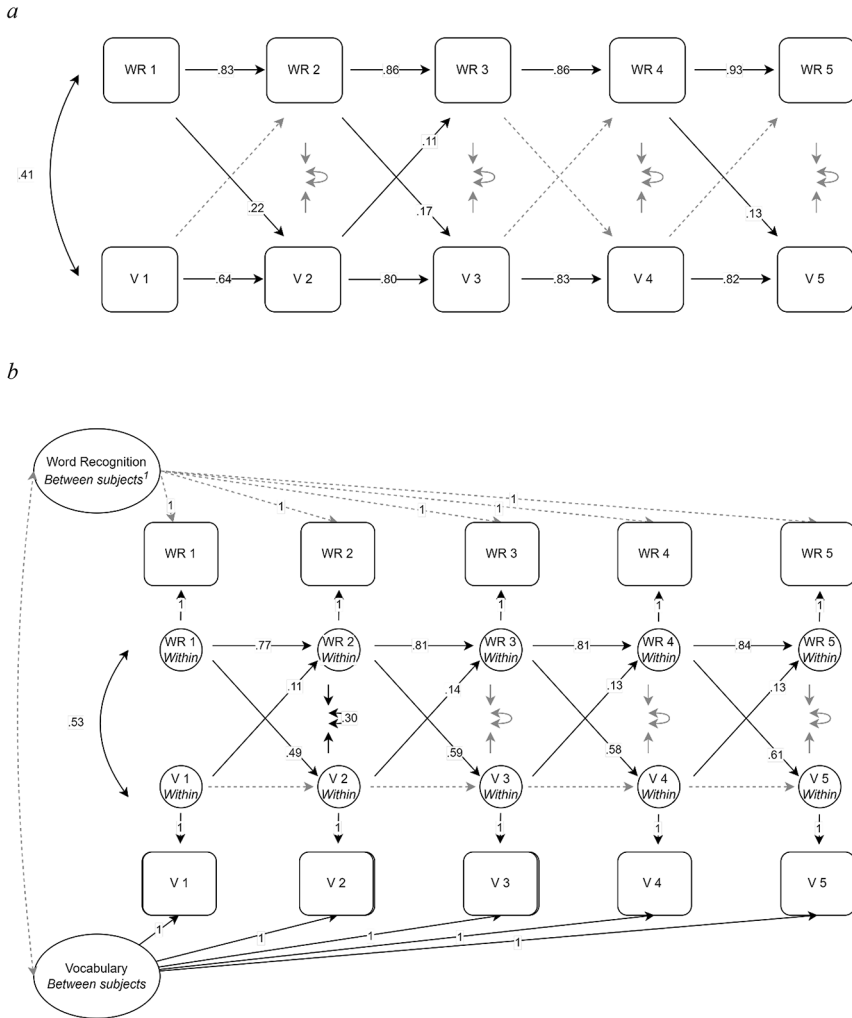


Fig. 1 a. cross-lagged panel model for expressive vocabulary and word recognition. b. Random-intercept cross-lagged panel model for expressive vocabulary and word recognition

Note. In both models, rectangles denote observed variables, single-headed arrows indicate regression coefficients, and double-headed arrows indicate covariances. In the RI-CLPM (Panel B), large ovals indicate between-child effects (random intercepts), and small circles indicate within-child effects. Standardised regression coefficients displayed for significant paths ($p < .05$); non-significant paths indicated with dotted lines. ¹In the RI-CLPM model, the random intercept for Word Recognition and the covariance between intercepts were constrained to 0. Diagrams were created using the JavaScript program drawio (v 22.0.8 release; <https://github.com/jgraph/drawio>)

Table 5 Model fit statistics for CLPM and RI-CLPM

	χ^2 (df)	CFI	TLI	RMSEA	SRMR	AIC	
CLPM	64.16** (24)	0.952	0.911	0.151	0.059	8492.6	
RI-CLPM	18.28 (21)	1.000	1.000	0.000	0.041	8449.1	
RICLPM (RI for EVT only)	18.56 (23)	1.000	1.000	0.000	0.040	8445.7	
<i>RI-CLPM equality constraints:</i>							$\Delta \chi^2$ (df)
Autoregressive (AR)	32.89 (29)	1.000	1.000	0.000	0.067	8449.8	12.56 (6)
AR, cross lags (CR)	33.87 (35)	1.000	1.000	0.000	0.070	8440.2	1.82 (6)
AR, CR, residual (co)variances	82.92 (44)	0.968	0.967	0.092	0.090	8480.0	42.58** (9)

* $p < .05$; ** $p < .001$

Discussion

We investigated the co-development of vocabulary and word-level reading from Grade 1 to 5 in a cohort of Australian children. Results provide evidence of bidirectional relationships between vocabulary and word recognition in the early years of school, on both measures of expressive and receptive vocabulary. Specifically, results indicate that when stable between-child differences are accounted for in the model, individual variation in word recognition is positively and significantly associated with subsequent vocabulary knowledge throughout the early years of school, and vocabulary is similarly associated with subsequent word reading. These effects appear to be relatively stable across development, as the models indicated the lagged effects to be time-invariant. Significant between-child variance was observed for vocabulary in both models, but not for word recognition skill, indicating an absence of stable individual differences among the sample in word recognition. These results provide support for the ongoing contribution that oral vocabulary knowledge makes to reading acquisition throughout pivotal school years and the ways in which word-level reading skills help to expand oral language vocabulary.

Findings from this study are consistent with the lexical quality hypothesis, which emphasises not only the quality of word representations for the individual child, but also the breadth of vocabulary knowledge and its influence on reading (Perfetti & Hart, 2001). Accordingly, having an extensive vocabulary is proposed to foster reading skill by supporting the reader to develop stronger links between phonological and orthographic representations of a word (Perfetti & Hart, 2001; Wegener et al., 2022). Stronger reading skills may then facilitate vocabulary growth: firstly, by allowing the individual greater exposure to new words through wider reading (Duff et al., 2015), and secondly, by freeing up cognitive resources during reading, so that existing connections between the meaning and form of a word may be enhanced, and the meaning of new words may be inferred based on their context (Cunningham & Stanovich, 1991; Perfetti, 2010; Verhoeven et al., 2011).

Previous longitudinal studies have consistently shown that word recognition ability is predictive of later oral vocabulary, in alphabetic (Georgiou et al., 2023; Verhoeven et al., 2011) and non-alphabetic written codes (e.g., Dulay et al., 2021; Hulme et al., 2018; Yan et al., 2021). In this study, these findings were generally confirmed. The RI-CLPMs, which provided the best fit to the data, indicated significant cross-lagged relationships predicting vocabulary from reading, and these were somewhat

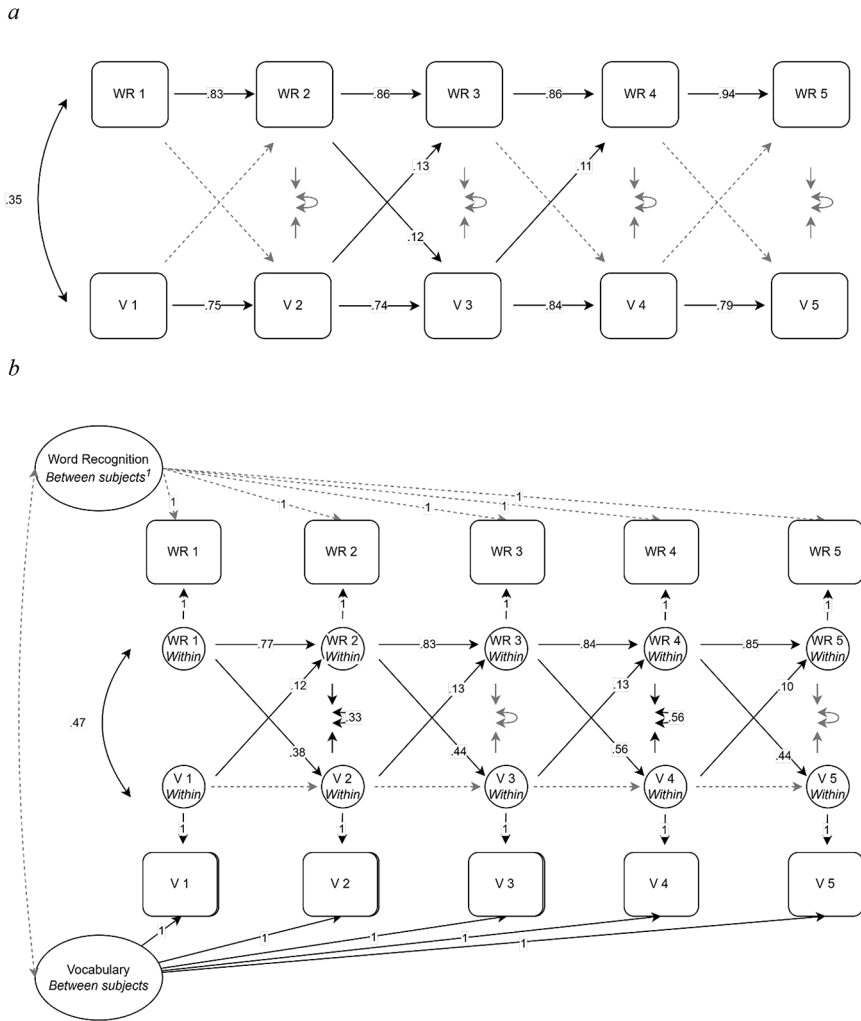


Fig. 2 a. cross-lagged panel model for receptive vocabulary and word recognition **b.** Random-intercept cross-lagged panel model for receptive vocabulary and word recognition

Note. In both models, rectangles denote observed variables, single-headed arrows indicate regression coefficients, and double-headed arrows indicate covariances. In the RI-CLPM (Panel B), large ovals indicate between-child effects (random intercepts), and small circles indicate within-child effects. Standardised regression coefficients displayed for significant paths ($p < .05$); non-significant paths indicated with dotted lines. ¹In the RI-CLPM model, the random intercept for Word Recognition and the covariance between intercepts were constrained to 0. Diagrams were created using the JavaScript program drawio (v 22.0.8 release; <https://github.com/jgraph/drawio>)

stronger from Grade 2 onwards for both measures of vocabulary. Similarly – while providing a much poorer fit to the data – the CLPMs produced significant cross-lagged associations in both models of vocabulary knowledge. As noted by Georgiou et al., new vocabulary in the early grades is acquired more readily through oral

language interactions, while word reading skills are still being learned, potentially reflected in a stronger relationship between word recognition and later vocabulary in the later primary years.

In general, these results are somewhat consistent with the only other prior CLPM study including Grades 4 and 5 data and a receptive vocabulary measure (Verhoeven et al., 2011), although in their study, significant paths were observed only from Grade 2 reading to Grade 3 vocabulary (larger effect; $\beta=0.37$), and Grade 4 to 5 (smaller effect; $\beta=0.04$). To understand and reconcile findings produced by these studies, it is important to consider that cross-lagged parameters in the RI-CLPM are interpreted differently to those of the CLPM (Hamaker et al., 2015). These models provide estimates of different effects, as the cross-lagged parameters are modelled at fundamentally different levels; therefore, it is unsurprising, and quite common, for parameters between the two models to change in significance or even direction (Mulder & Hamaker, 2021). It is also noteworthy in the present study that significant variance was observed in the random intercept for vocabulary, but not for word recognition. Hence, in the final RI-CLPMs, individuals varied around the same mean in word recognition (Hamaker et al., 2015; Mulder et al., 2021), but were assumed to fluctuate around their own average vocabulary level. Overall, these findings indicate a positive relationship between word recognition and subsequent vocabulary, suggesting that children with stronger reading skills, relative to the sample mean (in this case), tend to have stronger vocabulary knowledge, relative to their own mean, measured in the following year.

When considering the reverse relationship, that vocabulary predicts subsequent word reading, previous research has yielded inconsistent results, with some studies finding no evidence of cross-lagged effects (Georgiou et al., 2023) and others observing significant effects (Nation & Snowling, 2004; Verhoeven et al., 2011; Yan et al., 2021). When modelling these relationships using RI-CLPMs in the present study, we found significant positive cross-lagged associations of a similar magnitude to the CLPMs across adjacent waves on expressive and receptive vocabulary. These results indicate that stronger vocabulary knowledge, relative to a child's own mean, is related to better word recognition skill, relative to the sample mean, in the subsequent year, during early primary school. By accounting for individual differences in vocabulary, the link between vocabulary knowledge and subsequent reading ability may be stronger than implied by previous research.

Previous research using CLPM has found significant cross-lagged effects for expressive vocabulary and later word reading between Grade 1 and 3 (Yan et al., 2021; $\beta_s=0.20$), and the present study extends this by showing these relationships in the RI-CLPM up to Grade 5, with effects of a similar magnitude. For receptive vocabulary, the only prior study to find significant cross-lagged effects was Verhoeven et al.; however, effects were only significant for the relationship between measures taken 6 months apart during Grades 1 and 2 (β_s ranging from 0.03 to 0.09), while the yearly increments between Grades 2 and 5 were non-significant. In contrast, this study showed that receptive vocabulary predicted subsequent word recognition with the strength of these effects relatively consistent across grades. Again, it is possible that these discrepant findings are explained by the approach to analysis: when model-

ling these cross-lagged relationships at the within-participant level, it appears there is a stronger relationship between vocabulary and later word recognition.

Educational implications

These analyses highlight the importance of vocabulary knowledge in word reading development and vice versa, and therefore signal the need for early identification of difficulties and the provision of high-quality instruction and support in both domains. Research shows that even after controlling for multiple risk factors such as socio-economic status and family stress, oral vocabulary at 24 months remains a significant predictor of both academic and behavioural outcomes upon kindergarten entry (Morgan et al., 2015). Comprehensive and preventative mechanisms for support and intervention, such as Multi-tiered Systems of Support (MTSS), enable children at risk of reading difficulties to be detected early (De Bruin et al., 2024). The inclusion of measures which assess both early reading and vocabulary knowledge may be beneficial to universal screening processes, so that appropriate, timely intervention can be put in place.

Interventions targeted at improving vocabulary knowledge have been shown to enhance both word knowledge and reading comprehension (Elleman et al., 2009; Marulis & Neuman, 2010). Supporting vocabulary development in classroom instruction is also important for fostering reading development, particularly during the early years of school, where oral language and vocabulary instruction often receive limited attention (Wright & Neuman, 2014). Much of children's early so-called "Tier 1" vocabulary knowledge (Beck et al., 2013) is acquired through repeated exposure to oral language in activities such as child directed speech and being read books (e.g., Wasik et al., 2016; Weisleder & Fernald, 2013). However, vocabulary can and should be intentionally and explicitly taught in the classroom by focusing on word meanings, word knowledge, and active processing tasks (McKeown, 2019).

Strengths and limitations

Longitudinal studies offer great insight into the nature of child development, allowing a deeper understanding of how processes such as oral language and literacy skills emerge and influence one another over time. As an observational study, it is important to note that the present findings are *suggestive* of causal relationships yet cannot provide causal evidence, for which experimental research is required. The sample size in the present study was also relatively small; hence future research should replicate these results with a larger sample. This study was strengthened by the inclusion of both expressive and receptive measures of vocabulary, allowing a more direct comparison with the effects produced in previous studies examining this question. It is worth noting that, in contrast to the only other study examining cross-lagged relationships in English (Georgiou et al., 2023), the present study used a timed measure of word recognition, and consequently it is unknown whether the difference in type of measure contributed to the difference in findings. Future researchers could consider the inclusion of both timed and untimed measures of reading.

Finally, longitudinal studies such as this one typically employ standardised tests to assess both vocabulary and word recognition. As a result, word stimuli used on the separate measures do not necessarily correspond to one another, nor to words subsequently tested the following year; yet these transactional effects are thought to exist at the level of specific words, as indicated by item-level experimental studies. That is, knowledge of the meaning of a specific word is thought to support its decoding / recognition by further fusing the links between a word's pronunciation, meaning, and orthography (Ehri, 2022; Perfetti, 2010; Wegener et al., 2022). In the present study, despite using non-aligned measures across years, reciprocal relationships were still observed, potentially suggesting a more general effect that acquiring vocabulary has on reading development, and vice versa.

Conclusion

The present findings showed that oral vocabulary and word reading ability are transactional in nature, each serving to enhance growth in the other over time. This study provides a novel contribution to the field, by investigating oral vocabulary and word recognition across five years of school in a sample of English-speaking children, providing insight into the development of these processes over a pivotal developmental window. Moreover, this study is the first known attempt to investigate reciprocal relationships among these constructs using both CLPM and RI-CLPM, with the latter approach allowing a deeper investigation of how oral vocabulary and reading co-develop at an individual level, while simultaneously accounting for stable individual differences. Future researchers should confirm these findings with a larger sample size and consider how other variables affect the strength of these relationships: for instance, language background other than English. Importantly, these findings signal the need for high quality classroom instruction in both reading skill and vocabulary knowledge, and early and regular screening measures to inform the development of timely and appropriate support and intervention.

Acknowledgements This research was supported by the Australian Government through the Australian Research Council's *Discovery Projects* funding scheme (project DP160100319). The views expressed herein are those of the authors and are not necessarily those of the Australian Government or Australian Research Council.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Competing interests The authors report there are no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use

is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.
- Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*(5), 431–443. <https://doi.org/10.1177/0022219411410042>
- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*(2), 264. <https://doi.org/10.1037/0022-0663.83.2.264>
- Curran, P. J., & Hancock, G. R. (2021). The challenge of modelling co-developmental processes over time. *Child Development Perspectives, 15*(2), 67–75. <https://doi.org/10.1111/cdep.12401>
- De Bruin, K., Killingly, C., & Graham, L. (2024). What are multi-tiered systems of support (MTSS)? In L. Graham (Ed.), *Inclusive education for the 21st century: Theory, policy, and practice* (2nd ed., pp. 157–175). Routledge.
- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015a). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848–856. <https://doi.org/10.1111/jcpp.12378>
- Duff, D., Tomblin, J. B., & Catts, H. (2015b). The influence of reading on vocabulary growth: A case for a Matthew effect. *Journal of Speech Language and Hearing Research, 58*(3), 853–864. https://doi.org/10.1044/2015_JSLHR-L-13-0310
- Dulay, K. M., Law, S. Y., McBride, C., & Ho, C. S. H. (2021). Reciprocal effects of morphological awareness, vocabulary knowledge, and word reading: A cross-lagged panel analysis in Chinese. *Journal of Experimental Child Psychology, 206*, 105100. <https://doi.org/10.1016/j.jecp.2021.105100>
- Dunn, L. M., & Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test* (fourth ed.). NCS Pearson, Inc.
- Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training mispronunciation correction and word meanings improves children's ability to learn to read words. *Scientific Studies of Reading, 21*(5), 392–407. <https://doi.org/10.1080/10888438.2017.1315424>
- Ehri, L. C. (2022). What teachers need to know and do to teach letter–sounds, phonemic awareness, word reading, and phonics. *The Reading Teacher, 76*(1), 53–61. <https://doi.org/10.1002/trtr.2095>
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1–44. <https://doi.org/10.1080/19345740802539200>
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5
- Georgiou, G., Inoue, T., & Parrila, R. (2023). Are vocabulary and word reading reciprocally related? *Scientific Studies of Reading, 27*(2), 160–168. <https://doi.org/10.1080/10888438.2022.2123275>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. <https://doi.org/10.1177/07419325860070010>
- Graham, L. J., White, S. L., Tancredi, H. A., Snow, P. C., & Cologon, K. (2020). A longitudinal analysis of the alignment between children's early word-level reading trajectories, teachers' reported concerns and supports provided. *Reading and Writing: An Interdisciplinary Journal, 33*(8), 1895–1923. <https://doi.org/10.1007/s11145-020-10023-7>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. <https://doi.org/10.1037/a0038889>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160. <https://doi.org/10.1007/BF00401799>
- Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education, 39*(5), 304–312. <https://doi.org/10.1177/0741932518773154>

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hulme, C., Zhou, L., Tong, X., Lervåg, A., & Burgoyne, K. (2018). Learning to read in Chinese: Evidence for reciprocal relationships between word reading and oral language skills. *Developmental Science*, 22(1), e12745. <https://doi.org/10.1111/desc.12745>
- Hwang, H., McMaster, K. L., & Kendeou, P. (2023). A longitudinal investigation of directional relations between domain knowledge and reading in the elementary years. *Reading Research Quarterly*, 58(1), 59–77. <https://doi.org/10.1002/rrq.481>
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, 80(3), 300–335. <https://doi.org/10.3102/0034654310377087>
- McKeown, M. G. (2019). Effective vocabulary instruction fosters knowing words, using words, and understanding how words work. *Language Speech and Hearing Services in Schools*, 50(4), 466–476. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0126
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, 86(5), 1351–1370. <https://doi.org/10.1111/cdev.12398>
- Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(4), 638–648. <https://doi.org/10.1080/10705511.2020.1784738>
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40(5), 665–681. <https://doi.org/10.1037/0012-1649.40.5.665>
- Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4), 342–356. <https://doi.org/10.1111/j.1467-9817.2004.00238.x>
- Nippold, M. A. (2016). *Later language development: School-age children, adolescents, and young adults*. 4th edition PRO-ED Inc.
- Orth, U., Meier, L. L., Bühler, J. L., Dapp, L. C., Krauss, S., Messerli, D., & Robins, R. W. (2022). Effect size guidelines for cross-lagged effects. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000499>
- Ouellette, G. P. (2006). What's the meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554. <https://doi.org/10.1037/0022-0663.98.3.554>
- Perfetti, C. (2010). Decoding, vocabulary, and comprehension: The golden triangle of reading skill. In M. G. McKeown, & L. Kucan (Eds.), *Bringing reading research to life* (1st ed., pp. 291–303). The Guilford.
- Perfetti, C. A., & Hart, L. (2001). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–214). John Benjamins.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 1, pp. 97–110). Guilford Press.
- Smith, R. J., Snow, P. C., Serry, T. A., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, 42(3), 214–240. <https://doi.org/10.1080/02702711.2021.1888348>
- Snow, P. C. (2020). SOLAR: The science of language and reading. *Child Language Teaching and Therapy*, 37(3), 222–233. <https://doi.org/10.1177/0265659020947817>
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics*. Pearson.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency* (2nd ed.). Pro-ed.
- Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading*, 16(2), 122–140. <https://doi.org/10.1080/10888438.2010.542527>
- Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24(5), 637. <https://doi.org/10.1037/met0000210>

- Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*(1), 8–25. <https://doi.org/10.1080/1088438.2011.536125>
- Wasik, B. A., Hindman, A. H., & Snell, E. K. (2016). Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly, 37*, 39–57. <https://doi.org/10.1016/j.ecresq.2016.04.003>
- Wegener, S., Wang, H. C., de Lissa, P., Robidoux, S., Nation, K., & Castles, A. (2018). Children reading spoken words: Interactions between vocabulary and orthographic expectancy. *Developmental Science, 21*(3), e12577. <https://doi.org/10.1111/desc.12577>
- Wegener, S., Beyersmann, E., Wang, H. C., & Castles, A. (2022). Oral vocabulary knowledge and learning to read new words: A theoretical review. *Australian Journal of Learning Difficulties, 27*(2), 253–278. <https://doi.org/10.1080/19404158.2022.2097717>
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modelling* (pp. 209–231). Guilford Press.
- Willard, J. A., Agache, A., Kohl, K., Bihler, L. M., & Leyendecker, B. (2021). Longitudinal interrelations between nonword repetition and vocabulary from age three to five: Evidence for within-child processes? *Developmental Psychology, 57*(9), 1423–1438. <https://doi.org/10.1037/dev0001230>
- Williams, K. T. (2008). *Expressive vocabulary test Second Edition (EVT™ 2)*. NCS Pearson.
- Wright, T. S., & Neuman, S. B. (2014). Paucity and disparity in kindergarten oral vocabulary instruction. *Journal of Literacy Research, 46*(3), 330–357. <https://doi.org/10.1177/1086296X14551474>
- Yan, M., Li, Y., Sun, X., Zhou, X., Hui, Y., & Li, H. (2021). The roles of decoding and vocabulary in Chinese reading development: Evidence from a 3-year longitudinal study. *British Journal of Educational Psychology, 91*(1), 300–314. <https://doi.org/10.1111/bjep.12365>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.