Reading and Writing (2020) 33:1895–1923
https://doi.org/10.1007/s11145-020-10023-7

A longitudinal analysis of the alignment between children's early word-level reading trajectories, teachers' reported concerns and supports provided

Linda J. Graham[1] · Sonia L. J. White[1] · Haley A. Tancredi[1] · Pamela C. Snow[2] · Kathy Cologon[3]

Published online: 8 February 2020

Abstract

In this longitudinal study, the word-level reading trajectories of 118 children were tracked alongside teachers' reported concerns and types of support provided through Grades 1, 2 and 3. Results show a significant decline in composite scores relative to age norms over time, with children achieving significantly lower in phonemic decoding than word recognition at the subtest level. Five group trajectories were identified: children who achieved average or above average scores across all 3 years ($n=64$), children who consistently bordered on average ($n=11$), children who achieved below average in Grade 1 but who then achieved average or above in Grade 2 or Grade 3 ($n=7$), children who achieved average or above in Grade 1 but then declined to below average in Grade 2 or Grade 3 ($n=10$), and children who achieved below average across all 3 years ($n=26$). Appropriately, teachers' concerns were highest for students in the groups that improved, declined or remained persistently below average. However, analysis of the focus of teachers' concerns and the supports they said were provided to the children in these three groups suggests that teachers are not always accurate in their interpretation of children's presenting characteristics, resulting in the misalignment of support provision.

Keywords Early identification · Intervention · Longitudinal · Persistent reading difficulties · Support provision · Teacher knowledge

✉ Linda J. Graham
linda.graham@qut.edu.au

[1] The Centre for Inclusive Education, Faculty of Education, Queensland University of Technology (QUT), Kelvin Grove, QLD 4059, Australia

[2] School of Education, La Trobe University, Bendigo, Australia

[3] Department of Educational Studies, Macquarie University, Sydney, Australia

Learning to read is a fundamental achievement, ideally mastered by children within their first 3 years of formal schooling. Aside from the benefit of being able to escape into the world of imagination enjoyed by skilled readers, reading competency is associated with better school, further education and employment outcomes (Castles, Rastle, & Nation, 2018). If left unaddressed, early difficulties in the process of learning to read are likely to result in persistent reading difficulties and problems accessing the academic school curriculum (Partanen & Seigel, 2014; Spira, Bracken, & Fischel, 2005). Such difficulties are associated with a chain of negative consequences including poor academic self-concept (Chapman, Tunmer, & Prochnow, 2000), disruptive school behaviour (Arnold et al., 2005), academic underachievement (Cunningham & Stanovich, 1997), early school leaving (Daniel et al., 2006), and increased risk of mental health problems (Francis, Caruana, Hudson, & McArthur, 2019). High-quality initial reading instruction and the provision of evidence-based targeted supports to students exhibiting signs of early reading difficulties in the first 3 years of school are necessary to ensure that *all* students become competent readers (Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Partanen & Seigel, 2014).

Despite the prioritisation of literacy in school education, unacceptably high numbers of children do not achieve the level of reading proficiency required to function in a knowledge-based economy, placing them at increased risk of long-term unemployment and socioeconomic disadvantage. This is evident from student performance in the Progress in International Reading Literacy Study (PIRLS) and the Program of International Student Assessment (PISA). Together, these international assessments suggest that approximately one in five students internationally do not meet the baseline level of reading proficiency (18% and 20% respectively) (Mullis, Martin, Foy, & Hooper, 2017; Organisation for Economic Cooperation and Development, 2016). While it is well documented that around 5–10% of students will experience persistent reading difficulties (sometimes termed "dyslexia") despite appropriate and high-quality interventions (Partanen & Seigel, 2014), there are also a number of students who have been termed "instructional casualties" (Lyon, 2002; Snow, 2016). Instructional casualties are avoidable and occur when students are not provided with high-quality early reading instruction and/or when students with early reading difficulties are not accurately identified and provided with appropriate evidence-based supports (Vellutino, Scanlon, Zhang, & Schatschneider, 2008). It is therefore critical to pinpoint where and how this might be happening in the early years of formal schooling, as research shows that the provision of timely, evidence-based interventions as part of a multi-tiered system of support can avert negative trajectories (Galuschka, Ise, Krick, & Schulte-Körne, 2014; Swanson et al., 2017; Torgesen, 2004).

In Australia, where the present study was conducted, approaches to early reading instruction vary. The most common approach, supported by the Australian Literacy Educators' Association (ALEA, 2015), is "balanced literacy" (Hastings, 2012; Hill, 2017). While various forms of phonics-based instruction may be utilised as one element within a balanced literacy approach, there is concern that many Australian teachers do not have the knowledge to teach phonics with precision and fidelity, due to lack of emphasis on how to teach decoding skills and address learning difficulties in initial teacher education (Stark, Snow, Eadie, & Goldfeld, 2016). This perception

is supported by a recent national audit of initial teacher education unit outlines which found that only 4% of the 116 literacy units reviewed had a specific focus on early reading instruction, only 6% mentioned all five essential elements (phonemic awareness, phonics, fluency, vocabulary, and comprehension), and none mentioned the Simple View of Reading (Buckingham & Meeks, 2019). These knowledge gaps have implications not only for teachers' ability to teach all children in their class to read but to also accurately identify difficulties across the different components of reading, thereby enabling teachers to source and provide appropriate targeted support.

Investigating teachers' concerns

Evidence suggests teachers are quick to identify children who experience difficulties in school, particularly in the area of behaviour (Hecht & Greenfield, 2002). However, they often find it more difficult to identify specific factors contributing to those behaviours and/or what to do to support children once difficulties have been identified (Graham, 2015). This creates great potential for error in the identification process (Fletcher, Lyon, Fuchs, & Barnes, 2018), which involves analysis of students' presenting characteristics and is therefore vulnerable to latent assumptions about dis/ability, social background, gender, and ethnicity (Meissel, Meyer, Yao, & Rubie-Davies, 2017). There are also more and less "visible" characteristics, which can—in combination with those assumptions—cause teachers to pay attention to externalising behaviours only, potentially deflecting consideration of less observable factors, such as reading and/or language difficulties. The focus of teachers' concerns—whether, for example, they are principally concerned about learning or whether they are principally concerned about behaviour—is of critical importance because children find their way into the service-delivery system based on adults' perceptions of their primary area of difficulty (Cohen, Davine, Horodezky, Lipsett, & Isaacson, 1993). For example, it is hardly surprising that behavioural issues attract the attention of parents and teachers; yet, it is notable that when children for whom behavioural concerns have been raised undergo formal language assessments, high percentages (34.38% in Cohen et al.'s study) are identified as having clinically significant expressive and receptive language difficulties. There is, however, a bidirectional relationship between externalising behaviours and language and reading/learning difficulties (Fletcher et al., 2018) which, if unrecognised, may mean that a child only ever receives behavioural diagnostic labels and interventions, and never the primary academic support they need.

This paper aims to contribute to our understanding of this important problem by comparing the focus of teachers' concerns about individual students and the type of supports provided, with students' early word-reading trajectories. First, however, we make a distinction between *reading* and *literacy* (Stark et al., 2016). Reading refers to the ability to decode and identify printed words, with the goal being text comprehension (Castles et al., 2018a, b). Literacy is defined as "the ability to understand and use those written language forms required by society and/or valued by the individual" (Thomson, Hillman, Schmid, Rodrigues, & Fullarton, 2017), and

encompasses a broad range of knowledge and skills, including critical literacy and multiliteracies (Mills, 2015). This distinction is important because the ability to decode printed words and to comprehend text efficiently and effectively is a necessary grounding for literacy (Gough & Tumner, 1986), but does not alone constitute it.

Phonemic decoding and word recognition

A child's reading development is underpinned by their understanding of the alphabetic principle; knowledge that letters and letter patterns (graphemes) represent the sounds (phonemes) of spoken language (Castles et al., 2018a, b). In order to "crack the code", children need to quickly and easily translate graphemes into phonemes, enabling them to "sound the word out" and then match this word to an item stored in their oral lexicon (Braze et al., 2016). To ensure that *all* children fully master the alphabetic decoding skills they need to engage with the formal academic curriculum, grapheme-phoneme correspondences need to be systematically and explicitly taught; starting with the simplest and most frequently occurring and moving to the more complex and less frequently occurring (Ehri, Nunes, Stahl, & Willows, 2001). This is evidenced by the fact that even skilled readers draw on decoding skills, particularly when they encounter unfamiliar words like, for example, the word "obstreperous". To decipher meaning, a reader must phonemically decode the word, consider morpho-phonemic cues, and search for "matches" in their oral lexicon. Lack of success in this process means that unfamiliar words effectively remain "nonwords" to the reader.

While there is some literature that demonstrates teaching students a small number of high-frequency irregular words as "sight words" can support students' early reading success by contributing to orthographic knowledge (Castles et al., 2018a, b; Ehri, 1992, 2005; Shapiro & Solity, 2016), inefficient word memorisation strategies do not appear to foster skilled reading (Hudson, Pullen, Lane, & Torgesen, 2008). If novice readers are over-reliant on inefficient visual memorisation practices as a principal strategy, they are unlikely to make adequate reading progress. Such reading habits are at odds with the assumptions of reading proficiency upon which curriculum and pedagogy in the upper grades of primary school are generally based. This point was reinforced recently by Kilpatrick (2015), who noted: "[c]ontrary to any intuitions we may have about sight-word learning, a substantial amount of research shows that letter-sound knowledge is central to both phonic decoding *and* sight-word learning" (p. 84). It is therefore essential to accurately assess each skill to correctly identify problems and to source appropriate interventions early in the process of learning to read.

Standardised assessment tools that include pronounceable nonword (also known as pseudoword) reading tasks, such as the Test of Word Reading Efficiency-Second Edition (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012), are used internationally to test children's phonemic decoding skills (Language and Reading Research Consortium & Chiu, 2018; Peng et al., 2018; Vaughn et al., 2019). When presented with words for which they have no semantic association or sight word recognition,

children are required to apply their knowledge of grapheme-phoneme correspondences to "decode" the word. The same methodology underpins the Phonics Screening Check, which has been used in England since 2012 (Gibson & England, 2016) and is now in use across South Australia (Government of South Australia Department for Education, 2019). Word-level reading tasks that include both real word and pseudoword components can enable teachers to more accurately identify specific difficulties and to select interventions capable of targeting specific areas of weakness (Castles, Polito, Pritchard, Anandakumar, & Coltheart, 2018). However, the proposed use of the Phonics Screening Check across all Australian schools has been met with the claim that it "doesn't tell teachers anything they didn't know already… [or] what kind of instructional intervention their identified strugglers need" (Adoniou, 2017, np). In this paper, we analyse the word-level reading trajectories of 118 children across Grades 1–3 using the Test of Word Reading Efficiency (TOWRE-2; Torgesen et al., 2012), and then map these profiles against teachers' reported concerns and the types of supports children received over time. This analysis may help indicate whether an assessment like the Phonics Screening Check is necessary to help inform teachers' decision-making with regards to the identification and support of children experiencing reading difficulties in the early years of school.

The present study

The Supporting Behaviour in the Early Years project is a longitudinal study tracking children's development, school liking, language, learning, relationships and behaviour through the first 6 years of formal schooling in Queensland, Australia. The project, which began in 2014, explores the child and classroom characteristics that predict why some children begin to engage in disruptive behaviour with the aim of understanding which supports and/or changes in teaching practice are needed to work more productively with these children. This paper contributes to these aims by analysing word-level reading efficiency data from the 2nd, 3rd and 4th years of the study, when the children were in Grades 1, 2 and 3. We use these data to divide participating children into five longitudinal change groups and then examine teacher's reported concerns and the supports received for three of those groups: students whose word-level reading performance improved over time, students whose word-level reading performance declined over time, and students whose word-level reading performance remained below average across all 3 years.

Methods

Participants

Children and teachers from seven government primary schools servicing disadvantaged communities from the outer south metropolitan region of Brisbane, Australia, were involved in the longitudinal study. The children were recruited in the first (Preparatory) year of formal school in 2014 and will be tracked until the end of Grade

5 (2019). Information and consent forms were sent home to parents of all Prep year children attending mainstream classes in each of the seven schools and all children for whom parent consent was received participated in the project. Participating schools were between 1 and 2 SD below the national mean of 1000 (range 878–977, median 920) on the Index of Community Socio-Educational Advantage[1] (ICSEA) and were among the top 20 government primary schools in the South-East Queensland region for number of disciplinary suspensions (proportionate to enrolments) at the time of recruitment.

In each year of the project, participating children complete a suite of standardised measures assessing their development, attitudes, relationships, behaviour, oral language competence, and developing literacy and numeracy. By the fourth year of the project, a total of 118 children (55 males, 63 females) had completed the Test of Word Reading Efficiency-Second Edition (TOWRE-2; Torgesen et al., 2012) across all three waves: Grade 1 (mean age: 6.7 ± 0.3 years), Grade 2 (mean age: 7.6 ± 0.3 years) and Grade 3 (mean age: 8.7 ± 0.5 years). In each year of the study, their classroom teachers participated in a semi-structured interview probing teacher concerns about children's learning and behaviour, teaching strategies, support provided, teacher beliefs, perceived strengths and preferred areas for development, and years of experience. This analysis draws on interview data from the classroom teachers corresponding to each of the 118 children when they were in Grade 1 ($n = 28$), Grade 2 ($n = 24$) and Grade 3 ($n = 21$) respectively. Each of these teachers were interviewed in Term 4 of the respective school year. During the project, some teachers declined or were not available to participate in the study (Grade 2: $n = 3$; Grade 3: $n = 7$) or could not provide a response to specific questions (for example, if a child had left the school). When interview data is missing for an individual child, it is noted in the results where relevant.

All procedures in this research were conducted in accordance with the ethical standards of the institutional and national research committees. The study was approved by the Queensland University of Technology (QUT) Human Research Ethics Committee (Approval No. 1300000422). Approval to conduct research in state schools was granted by the Queensland Department of Education.

Measures

Child word reading efficiency

The TOWRE-2 (Torgesen et al., 2012) is a standardised assessment tool that investigates a student's ability to recognise familiar words (Sight Word Efficiency subtest) and to sound out words quickly and accurately (Phonemic Decoding Efficiency subtest). The TOWRE-2 was normed on 1717 individuals in the United States, aged six to 24 years. At the time of data collection and analysis, Australian norms were only

[1] All schools in Australia are given an ICSEA score: a calculation of the relative affluence of the school community (ACARA, 2013). ICSEA has a mean of 1000 and a standard deviation of 100. Note, as single ICSEA scores could reveal the identity of the schools, only ICSEA ranges have been provided here.

available for the TOWRE-1. Our use of the TOWRE-2 therefore required the use of US norms; although we acknowledge that the US norms for the TOWRE-1 overestimated the reading level of Australian children in the lower grades of primary school (Marinus, Kohnen, & McArthur, 2013). Overall average alternate-form coefficients were 0.91 (Sight Word Efficiency), 0.92 (Phonemic Decoding Efficiency) and 0.95 (Total Word Reading Efficiency). All coefficients exceeded the 0.9 criterion, indicating that the TOWRE-2 is a reliable measure. Both subtests were administered to participating students in a quiet room by a trained research assistant who presented stimuli words on a laminated A4 sheet, as per the TOWRE-2 examiner's manual. In the Sight Word Efficiency subtest, students were asked to read as many sight words as they could in 45 s. In the Phonemic Decoding Efficiency subtest, students were asked to read as many nonwords that they could in 45 s. Stimuli words gradually increased in length and complexity (e.g., "dat", "sploosh").

Teacher interviews

In each year of the study, participating classroom teachers completed a semi-structured interview that proceeded in two stages. The interviews were audio recorded and transcribed verbatim, then coded using inductive content analysis (Berg, 2001). Numeric codes were entered into SPSS. In the first stage of the interview, teachers were asked a series of structured screening questions for each participating child. Relevant to our analyses are questions asking teachers whether they had any concerns about a participating child, whether that concern related to learning, behaviour, or both, whether the child was receiving additional support and, if so, which type of support (Table 1). In the second stage of the interview, teachers who reported concerns about individual children were asked a series of open-ended questions and prompts to further understand the nature of their concerns and the supports provided. This second question series was not asked of teachers who reported no

Table 1 Agreed codes for each question

	Question	Coding
1	Do you have concerns about this child?	0: No
		1: Yes
2	Do your concerns relate to learning, behavior, or learning and behaviour?	0: Learning
		1: Behaviour
		2: Learning and behavior
3	Does this child receive additional support?	0: No
		1: Yes
4	What type of support does this child receive?	0: Learning
		1: Reading
		2: Behaviour
		3: Language
		4: Other

concerns. Teachers' responses to the second stage questions were used to check and validate codes relating to the support types they had nominated in the first stage of the interview.

Types of support were coded into five categories: Learning, Reading, Behaviour, Language, and Other. General literacy and numeracy support, like 'maths and writing' were coded under 'Learning'. Support that was specific to reading, like 'withdrawal for reading support', 'Reading Eggs', 'MultiLIT', 'Levelled Literacy Instruction', 'Words their Way' and 'Fly-in Guided Reading', were coded under 'Reading'. Behaviour plans, time in the 'Responsible Thinking Room', guidance counselling, wobble chairs, social skills and 'leaving at 1:30 P.M. everyday' were coded under 'Behaviour'. English as a Second Language (ESL), speech therapy and other oral language support were coded under 'Language'. 'Other' included support from allied health professionals, such as occupational therapy and physiotherapy, as well 'special needs teacher' time. Teacher aide time was coded into the category for which teachers' open-ended responses indicated it was being used; e.g., extra help with numeracy, behaviour management. Example codes are provided in Table 1.

Results

Our analyses proceed in two parts. First, we conduct a series of quantitative analyses to investigate the word-level reading performance of three achievement groups based on children's Total Word Reading Efficiency Index (TWRE) scaled scores (above average, average, below average) in Grades 1, 2 and 3. We then draw on the results of this quantitative analysis to derive five longitudinal change groups (staying average or above, persistently below average, consistently borderline, improving, and declining) to enable group comparison of children's word-level reading trajectories against teachers' reported concerns and the provision of supports.

Quantitative analysis: three achievement groups

Analysis of students' TOWRE-2 scores were conducted as per the TOWRE-2 examiner's manual. For each subtest, students' raw scores were converted to scaled (standard) scores. Then, subtest scaled scores were added and converted to a total TWRE standard score. The TOWRE-2 scaled scores are based on a distribution with a mean of 100 and a standard deviation of 15.

All statistical analyses were completed using SPSS (ver 25.0, www.ibm.com/spss). Descriptive statistics were used to describe the TWRE scaled scores for the children when in Grades 1, 2 and 3. A repeated measures analysis of variance was used to examine the differences between grades and TWRE subtests. In the event the sphericity assumption was being violated, a Greenhouse–Geisser correction was applied; with the epsilon value (ε) and more conservative degrees of freedom reported. Effect size has been presented as partial eta-squared (η_p^2).

The TOWRE-2 specifies seven descriptive categories that align with scaled scores. Scaled scores for individual children were converted into these descriptive

categories and then further classified into three achievement groups (see Table 2). Frequencies were calculated for each achievement group (above average, average, below average) in Grade 1, Grade 2 and Grade 3, and a Pearson χ^2 Test was used to compare frequencies in each achievement group at each timepoint.

Over the 3-year period, students' TWRE scaled scores ranged from 53 to 139. The mean TWRE scaled scores were higher when the children were in Grade 1 (mean 98.52 ± 16.56) when compared to Grade 2 scaled scores (mean 95.87 ± 17.31) and Grade 3 scaled scores (mean 92.66 ± 17.14). This difference was significant $F(1.629, 190.637) = 20.275$, $\varepsilon = .815$, $p < .001$, $\eta_p^2 = .15$, with all pairwise comparisons demonstrating a significant difference. The decline in scaled scores each year was a significant decline, representing a widening gap relative to TOWRE-2 population norms.

Analysis of the separate subtests indicated a similar significant decline over time; sight word recognition, $F(1.651, 193.128) = 23.456$, $\varepsilon = .825$, $p < .001$, $\eta_p^2 = .17$, and phonemic decoding, $F(1.738, 203.374) = 10.056$, $\varepsilon = .869$, $p < .001$, $\eta_p^2 = .08$. Interestingly, the significant decline in sight word recognition was not evident when comparing Grade 1 (mean 100.72 ± 17.26) and Grade 2 (mean 98.84 ± 16.84) but became evident in Grade 3 (mean 94.24 ± 16.05). Conversely, all the pairwise comparisons for phonemic decoding were significant, declining from Grade 1 (mean 96.38 ± 15.45), to Grade 2 (mean 93.18 ± 17.57), to Grade 3 (mean 91.33 ± 18.66). Overall, phonemic decoding scores were significantly lower than sight word recognition scores, $F(1.117) = 36.497$, $p < .001$, $\eta_p^2 = .24$), however, both were declining relative to TOWRE-2 age norms at each time point.

When the scaled score descriptive categories were converted to the three achievement groups (above average, average and below average) based on the TOWRE-2 age-matched population norms, the distribution across achievement groups and across grade levels was variable. Table 2 depicts the frequency distribution using all seven TOWRE-2 scaled score descriptive categories, as well as the three achievement groups (above average, average, below average). These frequencies indicate that TWRE scores were changing from Grade 1 to Grade 3. Particularly concerning is the increasing number of children with scaled scores that were more than two standard deviations below the mean, placing them in the very poor (< 70) category. Only one child scored in this category in Grade 1; however, by Grade 3, 16 children scored in this category (Table 3).

Table 2 Categories as per TOWRE-2 total word reading efficiency index (TWRE) scaled scores

TWRE scaled score categories	Achievement groups
> 130 = very superior	Above average (≥ 111)
121–130 = superior	
111–120 = above average	
90–110 = average	Average (90–110)
80–89 = below average	Below average (≤ 89)
70–79 = poor	
< 70 = very poor	

Table 3 Frequencies for seven descriptive categories and three corresponding achievement groups

Scaled score categories	Grade one		Grade two		Grade three	
Very superior > 130	5	Above average 31	3	Above average 25	0	Above average 19
Superior 121–130	9		5		6	
Above average 111–120	17		17		13	
Average 90–110	48	Average 48	50	Average 50	59	Average 59
Below average 80–89	23	Below average 39	21	Below average 43	12	Below average 40
Poor 70–79	15		13		12	
Very poor < 70	1		9		16	

Using the three achievement groups (above average, average, below average), the Pearson χ^2 test indicated a significant difference between the distribution of children in Grade 1 and Grade 2 χ^2 (4) = 92.49, $p < .001$, and between Grade 2 and Grade 3 χ^2 (4) = 66.93, $p < .001$. These analyses indicated that there was a distinct change in the frequency of children in each of the three achievement groups from Grade 1 to Grade 2, and Grade 2 to Grade 3. Figure 1 presents the movement of individual children between the three achievement groups across Grades 1–3.

Observing group frequencies over time (Table 3 and Fig. 1), there is a decline in the number of children who are achieving above average, together with an increase in the number of children who are achieving within the average range and a

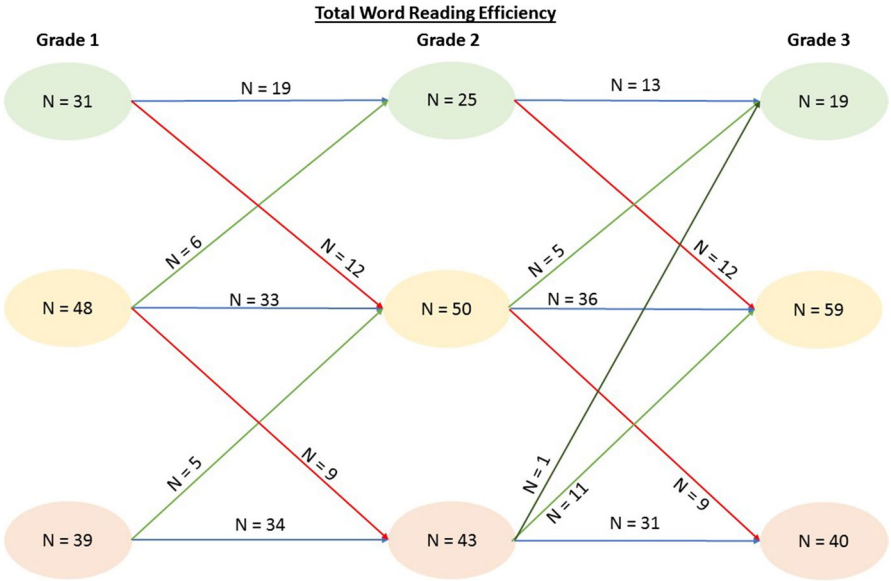

Fig. 1 Change through Grades 1 to 3, organised into three achievement groups (above average: green ellipse, average: yellow ellipse, below average: red ellipse). Change in child achievement is colour coded. Improvement: green, decline: red, and no change: blue. (Colour figure online)

persistently high number of children who are achieving below average. This pattern is explained by the movements of children represented by arrows in Fig. 1. Although the majority of our sample stayed within their respective achievement group from Grade 1 to Grade 3, 21 children (17.80%) declined to a lower achievement group at each transition point (Grade 1–Grade 2, and Grade 2–Grade 3). By contrast, only 11 children (9.32%) improved enough from Grade 1 to Grade 2 to join a higher achievement group. From Grade 2 to Grade 3, 17 children (14.41%) improved enough to join a higher achievement group. Seven of these 17 children (41.18%) spoke English as a Second Language.

Our use of these three achievement groups is valuable in practical terms, as above average, average and below average categories are regularly used by schools and classroom teachers to identify individual children requiring support or extension. However, the patterns of stable achievement versus improvement and decline identified through our quantitative analyses suggest that additional groupings are needed to both identify and account for movement (or lack of movement) of individuals over time. We therefore re-grouped our sample according to their pattern of achievement in Grades 1–3, identifying five distinct longitudinal change groups whose word-level reading trajectories can be analysed qualitatively in parallel with teachers' reported concerns and the provision of supports.

Qualitative analysis: five change groups

Based on our statistical analyses of word-level reading performance as measured by the TOWRE-2, individual children were re-grouped based on how their TWRE achievement group (above average, average, below average) did or did not change from Grade 1 to Grade 2 to Grade 3. This resulted in five groupings that we described as: staying average or above, persistently below average, improving, declining, or consistently borderline. Children's scores over time dictated group allocation. To be allocated to any one of the three "no change" groups (staying average or above, persistently below average, and consistently borderline), children's scores must have remained within the same respective category for all 3 years of data collection. With respect to the two "change" groups, children were allocated to the improving group if they achieved a below average score in Grade 1 but improved to average (or above average) in either Grade 2 or Grade 3. Children were allocated to the declining group if they achieved an average (or above average) score in Grade 1 but declined to below average in either Grade 2 or Grade 3. To manage the potential for marginal improvements/declines, we applied a minimum criterion of 2 standard points (calculated as 10% of the 20 standard points that is the average range 90–110) into the average/below average group to determine clear improvement or decline. An improvement into the average group required a scaled score of 92 or better and a decline into the below average group required a scaled score of 88 or less. Other standard scores (89–91) were captured within the consistently borderline group. The consistently borderline group was characterised by marginally fluctuating performance that alternated between average and below average over the 3 years of data

collection. This group also included improvements/declines that were borderline in nature with scaled scores all in the range of 89–91.

The distribution of children over the five longitudinal change groups is summarised in Table 4. Boys accounted for less than half the overall sample (46.61%) but more than two thirds of the persistently below average group (69.23%). Girls typically achieved higher scores than boys with girls accounting for more than half the full sample (53.39%) but less than one-third of students in the persistently below average group (30.77%). Students who spoke English as a Second Language (ESL) accounted for just over one-third of both the overall sample (35.59%) and the staying average or above group (39.06%), just under half of the children in the improving (42.86%) and declining groups (40.00%), and less than one in five children in the persistently below average group (19.23%). Students from an English language background (ELB) accounted for less than two-thirds of the full sample (64.41%), but more than four fifths of students in the persistently below average group (80.77%).

In the following section, we analyse teachers' reported concerns across our five longitudinal change groups. We then examine three of those change groups more closely: students with below average scores that improved to average or above average (improving group), students with scores that declined from average to below average (declining group), and students with scores that stayed below average across the Grade 1–Grade 3 period (persistently below average group). We then examine the supports provided to these children.

Teachers' reported concerns and type of concern

Teachers' reported concerns about children in their class fluctuated over time for all five groups (Table 5); however, the direction and magnitude of that fluctuation differed across groups. Appropriately, teachers' concerns were highest for students in the improving, declining and persistently below average groups. Our subsequent analyses focus on these three groups, beginning with the improving group.

Improving group ($n = 7$)

Students in the improving group all had composite TWRE scores in the below average range in Grade 1 but improved to average or above average in either Grade 2

Table 4 Demographics of the five change groups

Change group	Males	Females	ELB	ESL
Stayed average or above $n = 64$	27	37	39	25
Consistently borderline $n = 11$	3	8	7	4
Improving $n = 7$	3	4	4	3
Declining $n = 10$	4	6	6	4
Persistently below average $n = 26$	18	8	21	5
Total $n = 118$	55	63	76	42

Table 5 Teachers' reported concerns across groups over three time points

Group	Grade one			Grade two			Grade three		
	No concerns %	Some concerns %	Missing %	No concerns %	Some concerns %	Missing %	No concerns %	Some concerns %	Missing %
Stay above ($n=64$)	60.94	39.06	–	71.88	23.44	4.68	59.38	28.12	12.50
Borderline ($n=11$)	54.55	33.33	22.22	72.73	27.27	–	72.73	27.27	–
Improving ($n=7$)	14.29	85.71	–	28.57	66.67	14.29	14.29	83.33	14.29
Declining ($n=10$)	30	70	–	50	44.44	10	40	50.00	20
Persistently below average ($n=26$)	0	100	–	15.38	84.62	–	7.69	90.47	19.23

Table 6 Distribution of subtest scaled scores for students in the improving group ($n=7$) over time

Descriptive categories	Sight word efficiency						Phonemic decoding efficiency					
	Grade 1		Grade 2		Grade 3		Grade 1		Grade 2		Grade 3	
Superior 121–130	0	Above average 0	0	Above average 0	0	Above average 1	0	Above average 0	0	Above average 0	1	Above average 1
Above average 111–120	0		0		1		0		0		0	
Average 90–110	2	Average 2	4	Average 4	6	Average 6	1	Average 1	2	Average 2	3	Average 3
Below average 80–89	3	Below average 5	3	Below average 3	0	Below average 0	2	Below average 6	3	Below average 5	3	Below average 3
Poor 70–79	2		0		0		4		2		0	
Very poor < 70	0		0		0		0		0		0	

or Grade 3. As shown in Table 6, these seven children improved in *both* sight word recognition and phonemic decoding.

Teachers had concerns about six of these seven children in Grade 1 (Table 7). While some teachers did not provide a response or declined to be interviewed in Grades 2 and 3, leading to missing data for one child (Student 67), the reported data indicates some consistency in concern type for students in this group. For example, across years, teachers consistently expressed concerns about "learning and behaviour" for Student 101 and "learning" for Student 16. Student 194's teachers also consistently raise concerns about his behaviour but not always about his learning. In contrast, teachers raise no concerns about Student 236 until Grade 3 when concerns are raised about both learning and behaviour. Student 80's Grade 1 teacher raises concerns about learning but—possibly reflecting improvements in achievement—concerns are not raised by either her Grade 2 or 3 teachers.

Declining group ($n = 10$)

Students in the declining group all had composite TWRE scores in the average range in Grade 1 but slipped to below average by Grade 3. At the subtest level, this decline was more severe in phonemic decoding than sight word recognition (Table 8). In other words, not only were children in the declining group progressively slipping from average to below average in terms of composite scores, but some children were also slipping from the "below average" category to either the "poor" or "very poor" category in either or both subtests.

As shown in Table 9, teachers had concerns about seven of these 10 children in Grade 1 but only four of the same 10 children in Grades 2 and 3. While some teachers could not provide a response or declined to participate in the study, leading to missing data for a small number of children, the available data still indicates that the number of children for whom teachers expressed no concerns in Grades 2 and 3 increased relative to Grade 1. Further, concerns appear to differ by teacher because the individual children about whom teachers express concerns are not the same children in each successive year. For example, in Grade 1 teachers expressed concerns about two children in the declining group for behaviour (Students 4 and 63) and another child for both learning and behaviour (Student 9). In Grade 2, however, the concern for Student 9 is purely behaviour (and not learning), Student 63's teacher has no concerns, and the student for whom a teacher has concerns about *both* learning and behaviour is now Student 153. By Grade 3, the pattern has changed again.

Table 7 Variability in teachers' reported concerns over time for children in the improving group ($n = 7$)

Concern	Grade one	Grade two	grade three
Behaviour		194	194
Learning and behaviour	67, 101, 194	101	101, 236
Learning	16, 80, 226	16, 226	16, 226
No concerns	236	80, 236	80
Missing		67	67

Table 8 Distribution of subtest scaled scores for students in the declining group ($n = 10$) over time

Scaled Score Categories	Sight Word Recognition						Phonemic Decoding					
	Grade 1		Grade 2		Grade 3		Grade 1		Grade 2		Grade 3	
Above average 111–120	0	Above average 0	0	Above average 0	0	Above average 0	1	Above average 1	0	Above average 0	0	Above average 0
Average 90–110	9	Average 9	8	Average 8	5	Average 5	8	Average 8	2	Average 2	1	Average 1
Below average 80–89	1	Below average 1	2	Below average 2	2	Below average 5	1	Below average 1	5	Below average 8	3	Below average 9
Poor 70–79	0		0		3		0		1		3	
Very poor < 70	0		0		0		0		2		3	

Table 9 Variability in teachers' reported concerns over time for children in the declining group ($n = 10$) over time

Concern	Grade 1	Grade 2	Grade 3
Behaviour	4, 63	4, 9	54, 76
Learning and behaviour	9	153	4
Learning	54, 76, 153, 220	220	13
No concerns	13, 96, 206	13, 63, 76, 96, 206	9, 63, 206, 220
Missing	–	54	96, 153

Student 9's teacher registers no concerns and Students 54 and 76 are the students about whom teachers have concerns for both learning and behaviour. The only student among the 10 children in the declining group about whom teachers consistently raise similar concerns is Student 4. A similar pattern of irregularity is evident in the learning category.

Persistently below average group ($n = 26$)

Students in the persistently below average group all achieved below average composite TWRE scores at each of the three timepoints. As with the declining group, many children slipped from the "below average" category to either the "poor" or "very poor" categories (Table 10). For example, in Grade 1 none of the 26 children in the persistently below average group were in the "very poor" category but, by Grade 3, 11 children had fallen into the very poor category for sight word recognition. Again, like the declining group, these declines to lower categories were more evident in the area of phonemic decoding. None of the children in the persistently below average group were the "very poor" category in Grade 1 but, by Grade 3, 15 children had fallen into this category.

As shown in Table 11, teachers had concerns about all 26 children in the persistently below average group in Grade 1, but by Grade 2 this had declined to 22 children with the remaining four raising no concerns. By Grade 3, teachers of two students in the persistently below average group (Students 87 and 208) stated that they had no concerns. These students were not the same students about whom the Grade 2 teachers had no concerns (Students 3, 17, 26 and 33).

Teachers' type of concern was more consistent across years and for individual children for the persistently below average group than for the declining group, although some variability in teachers' concerns is still evident (Table 11). For example, Student 3's Grade 1 teacher expresses concern for his behaviour but not his learning. His Grade 2 teacher expresses no concerns, whereas his Grade 3 teacher is concerned about *both* his learning and behaviour. A similar swinging pattern is evident for Student 33, whose Grade 1 teacher expresses concerns about both learning and behaviour, but her Grade 2 teacher has no concerns and her Grade 3 teacher expresses concerns only about her learning. Teachers' concerns are consistent across years for only seven of the 26 children in the persistently below average group: Students 34, 36, 48, 134, 223 and 224 in the learning and behaviour category, and Student 232 in the learning category. All other children for whom teacher interview

Table 10 Distribution of subtest scaled scores for students in the persistently below average group ($n = 26$) over time

Scaled score categories	Sight word recognition						Phonemic decoding					
	Grade 1		Grade 2		Grade 3		Grade 1		Grade 2		Grade 3	
Average 90–110	1	Average 1	1	Average 1	3	Average 3	1	Average 1	0	Average 0	0	Average 0
Below average 80–89	17	Below average 25	8	Below average 25	4	Below average 23	14	Below average 25	6	Below average 26	2	Below average 26
Poor 70–79	8		13		8		11		12		9	
Very poor < 70	0		4		11		0		8		15	

Table 11 Variability in teachers' reported concerns over time for children in the persistently below average ($n=26$) group

Concern	Grade 1	Grade 2	Grade 3
Behaviour	3, 199	1, 199	27
Learning and behaviour	1, 27, 33, 34, 36, 48, 60, 74, 77, 108, 129, 132, 134, 223, 224	8, 27, 30, 34, 36, 48, 60, 74, 134, 223, 224	1, 3, 34, 36, 48, 70, 77, 132, 134, 199, 223, 224
Learning	8, 17, 26, 30, 70, 87, 100, 208, 232	70, 77, 87, 100, 108, 129, 132, 208, 232	17, 26, 33, 108, 129, 232
No concerns		3, 17, 26, 33	87, 208
Missing			8, 30, 60, 74, 100

data is available change category at least once, although it is possible that Grade 3 teachers for whom interview data is not available may have reported similar concerns each year. Importantly however, teachers also report *no* concerns for six of the 26 children in the persistently below average group, at least once.

Support provided

As we noted in the introduction, teachers' categorisation of concern is important because the perceptions of key adults in children's lives direct service-delivery (Cohen et al., 1993). In both stages of our two-stage teacher interviews, we asked teachers whether children for whom they had concerns were receiving additional support and, if so, what types of support those children were receiving. The interviewer was instructed not to specify what "additional" meant in the context of support and no teachers sought clarification. All teachers interpreted additional support to mean support external to their own classroom teaching such as guidance counselling or extra teacher aide time, or participation in social skills or literacy programs, except when it came to behaviour plans. Some teachers also included wobble chairs and part-time attendance as additional support. No teachers described making reasonable adjustments in the context of inclusive practice as an example of additional support. The most commonly reported additional support was floating teacher aide time. In the following section, we examine students' achievement in the TWRE against teachers' stated concerns and supports provided for the improving, declining and persistently below average groups (Table 12). Our analyses reveal clear discrepancies between teachers' reported concerns and the provision of support, however, the frequency counts shown in Table 12 do not reveal the full extent of these discrepancies as frequencies do not detect exchanges at the individual student level. In some cases, students about whom teachers had no concerns were receiving additional support, while some students about whom teachers did have concerns were receiving none. The final section of this paper therefore unpacks supports received at the individual level.

Table 12 Summary of support received for the improving, declining and persistently below average groups

Improving group	Grade 1 ($n=7$)	Grade 2 ($n=6$)	Grade 3 ($n=6$)
Students teachers express concerns about	6	4	5
No support	4	4	1
Support received	3	2	4
Behaviour-related support	0	0	0
Reading-related support	1	0	3
Declining group	Grade 1 ($n=10$)	Grade 2 ($n=9$)	Grade 3 ($n=8$)
Students teachers express concerns about	7	4	4
No support	4	5	5
Support received	6	4	3
Behaviour-related support	0	0	0
Reading-related support	1	0	1
Stay below group	Grade 1 ($n=26$)	Grade 2 ($n=26$)	Grade 3 ($n=21$)
Students teachers express concerns about	26	22	19
No support	5	10	7
Support received	21	16	12
Behaviour-related support	7	2	2
Reading-related support	10	7	4

Numbers do not total as children could be receiving more than one type of support. Only behaviour and reading-related supports are disaggregated. Teacher interview data was unavailable for some children; this is reflected in the n values in each column

Improving group ($n=7$)

Not all students in the improving group about whom teachers expressed concern received support (Table 12). Of those who did receive support, most received English as a Second Language (ESL) support and one received speech language therapy. Only four students in the improving group received support that was specifically related to reading at any time across the 3 years: Student 226 in Grade 1, and Students 16, 101 and 236 in Grade 3 (Table 13).

Declining group ($n=10$)

Across Grades 1–3, teachers reported that around half the students in the declining group were not receiving additional support each year (Table 12). Of the six students who were receiving some form of support in Grade 1, two were receiving support for speech, another was receiving oral language support, one was receiving generic

Table 13 Students in the improving group ($n=7$) who scored in the "below average", "average" and "above average" categories over time with those receiving reading related support highlighted in bold italics

TWRE category	Grade 1	Grade 2	Grade 3
Above average (111–120)			**226**
Average (90–110)		101, 236	***16***, 67, **80**, ***101***, **194**, ***236***
Below average (80–89)	16, *67*, **80**, 101, **194**, ***226***, 236	16, *67*, **80**, **194**, **226**	

Bold denotes students with English as a Second Language (ESL). Children with missing interview data are in plain italics

support for literacy, numeracy and gross motor skills, and one teacher indicated that there was a floating teacher aide in her room. In Grades 2 and 3, the number of children receiving support declined to four and then three children with teachers again describing generic supports for numeracy, and social and motor skills, coupled with floating teacher aide support. Two children in Grade 2 received more specific supports: Student 220 received English as a Second Language (ESL) support and Student 13 spent time in the Special Education Unit for vision impairment support in both Grades 2 and 3. Only one child received support that was specifically reading-related in any year and this was the same child (Student 54) in both Grades 1 and 3 (Table 14).

Persistently below average group ($n=26$)

Across the 3 years of interviews, teachers reported that the number of students in this group who received no support increased relative to Grade 1 (Table 12). While most children in this group were receiving some form of support in Grade 1 (80.77%), teachers described mainly generic supports (e.g., extra teacher aide time, social skills, weekly school counsellor visits, wobble chairs). Only 10 students in the persistently below average group (38.46%) received support that was specifically reading-related at any one timepoint, and this number declined across the 3 years (Table 15). Only two children (Students 223 and 224) consistently received

Table 14 Students in the declining group ($n=10$) who scored in the "average" and "below average" categories over time with those receiving reading related support highlighted in bold italics

TWRE category	Grade 1	Grade 2	Grade 3
Average (90–110)	4, 9, **13**, ***54***, 63, **76**, 96, 153, **206**, 220	63, 96, 153, **206**, 220	
Below average (80–89)		4, 9, **13**, *54*, **76**	4, 9, **13**, ***54***, 63, **76**, *96*, *153*, **206**, 220

Bold denotes students with English as a Second Language (ESL). Children with missing interview data are in plain italics

Table 15 Students in the persistently below average group ($n=26$) who scored in the "below average", "poor" and "very poor" categories with those receiving reading related support highlighted in bold italics

TWRE category	Grade 1	Grade 2	Grade 3
Below average (80–89)	***1***, ***3***, ***8***, **26**, ***33***, 60, 70, 74, **87**, ***129***, 134, 199, **208**, 232	8, 33, ***87***, **129**, 199, ***208***	33, 199, **208**
Poor (70–79)	**17**, 27, 30, 34, ***36***, 77, 100, ***108***, ***132***, ***223***, ***224***	1, **17**, **26**, 30, 36, 60, 70, 74, 100, 132, ***223***	1, 3, *8*, ***17***, **26**, 27, 132
Very poor (<70)	48	3, 27, 34, 48, ***77***, 108, ***134***, ***224***, ***232***	*30*, ***34***, 36, 48, *60*, 70, *74*, 77, **87**, *100*, 108, **129**, 134, ***223***, ***224***, 232

Bold denotes students with English as a Second Language (ESL). Children with missing interview data are in plain italics

reading-related support from Grades 1 to 3, although it is possible that the five students for whom teacher interview data is not available in Grade 3 were receiving support and that it was reading-related. However, even if this were the case, it is notable that only one of these five children (Student 8) received reading-related support in Grade 1, and none received reading-related support in Grade 2.

Discussion

This study examined the word-level reading trajectories of 118 children through Grades 1–3, as measured by the Test of Word Reading Efficiency-Second Edition (TOWRE-2; Torgesen et al., 2012), alongside teachers' reported concerns and the types of support teachers report students receiving. Results from our quantitative analyses of Total Word Reading Efficiency Index (TWRE) scaled scores from the 118 children in our sample show a significant decline in word-level reading scores over time relative to age norms. This decline was also evident at the subtest level in both phonemic decoding and word recognition, however children's phonemic decoding skills were significantly weaker. These findings are deeply concerning given that both phonemic decoding and word recognition skills are necessary for children to achieve the level of reading proficiency required to function in a modern knowledge-based economy. Poor phonemic decoding skills and overreliance on inefficient visual memorisation strategies together burden working memory, impairing children's ability to access the academic school curriculum as they progress through school grades. Indeed, recent results from a longitudinal study indicate that early decoding and language skills explained 99.7% of the variance in reading comprehension at 7 years of age (Hjetland et al., 2019). Weaker skills in phonemic decoding across all our student groups highlight explicit phonics-based instruction as a potentially neglected component of reading instruction. This is a common criticism of the "balanced literacy" approach; an approach that informs early reading instruction in most Australian schools, including the seven schools participating in this study.

While our analyses did highlight one group that demonstrated improvement over time, the number of children in the improving group was small and almost half were children from a language background other than English. Only four of the seven children in the improving group received reading-related support at any one time, with others receiving English as a Second Language (ESL) support or speech language pathology services. The declines in TWRE scaled scores for the larger declining and persistently below average groups are an important finding, for several reasons. First, our results show that children in both these groups were doing better in Grade 1 relative to age norms than they were two full school years later, which suggests that the reading instruction and support they were receiving during that period did not have the intended or necessary impact. Second, some of the most at-risk children fell from the "below average" to "poor" or "very poor" categories in both word recognition and phonemic decoding, indicating that their initial difficulties were becoming more severe and entrenched over time. Finally, the declines in

achievement detected by our analyses challenge social background as an explanation for these children's reading difficulties.

While it is true that there is an association between socioeconomic disadvantage, child development and reading achievement (Aikens & Barbarin, 2008), that relationship exists well before children begin formal schooling and all the children in our study were attending schools serving communities of similar social status. It is not the case in this study that the children achieving average and above average scores were from more affluent backgrounds than the children scoring below. The profiles of the children in our declining group especially call the social background explanation into question, given they were all achieving average or above average scores in their second year of school and declined to below average in the third and fourth years. Further, while previous research suggests that word-reading growth among at-risk children accelerates first but then decelerates in the intermediate grades (Peng et al., 2018), this does not mean it is typical for growth to reverse or that children should do more poorly over time relative to their starting point. Finally, the overall better performance of children from a language background other than English challenges English language learning status as an explanation for poor reading proficiency (Geva & Farnia, 2012).

Our examination of teachers' ability to identify children experiencing difficulties found considerable variability, as well as inconsistency, in the focus of their concerns. Teachers also rarely nominated reading as an area of concern. This is problematic, as teachers' understanding of the reasons underlying children's difficulties directly influence the supports that children receive. Of specific note—consistent with the work of Cohen et al. (1993)—is the fact that some teachers nominated student behaviour, and not learning, as the reason for their concern, even with respect to students identified through this research as persistently performing below average in word-level reading. We also identified considerable variability between individual children's subsequent grade teachers with respect to the type of concern identified over time, which may be influenced by individual teachers' perceptions of and skills in identifying learning needs and managing behaviour. While we did not have the scope to provide an analysis of the reasons for teachers' reported concerns in this paper, we note that it was rare during interview for teachers to consider behaviour as a possible indicator of underlying academic difficulties. Rather, teachers were typically of the view that behaviour affected learning. This is consistent with previous Australian research highlighting a similar pattern of belief (Childs & McKay, 2001). However, when teachers make pedagogical and resourcing decisions based on their interpretation of students' surface-level behaviours, the presence of early language and/or reading difficulties may be overlooked, despite strong evidence of a bidirectional relationship between the two (Fletcher et al., 2018). This possibility has been noted in previous research from the state of Queensland, which found that "ADHD-like behaviours" acted as a "red-herring" in primary school, resulting in the provision of behavioural responses—as opposed to academic and/or language support—to students who were later "verified" under the Speech Language Impairment disability support category (Graham, 2008). Future analyses of this 6-year longitudinal dataset will examine our initial observations of a similar pattern in relation to reading and other academic support using both qualitative and quantitative analyses.

Another important discrepancy emerged where teachers expressed concerns about students yet reported many of these children as not receiving any support. A common view among Queensland teachers, including those participating in this study, is that students require a departmentally verified diagnosis in one of six limited categories of disability—intellectual impairment, physical disability, sensory disability, speech language impairment, or autism spectrum disorder—in order to receive additional support (Graham & Tancredi, 2019). This is a widely-held, yet false belief, that runs counter to international human rights and national antidiscrimination legislation, which entitles children with *any* type of disability—whether diagnosed or imputed—across Australia to support in the form of reasonable adjustments (Graham, Tancredi, Willis, & McGraw, 2018). However, our individual level analyses show that even when participating teachers did have concerns and children did receive support, it was rare for that support to be reading-related. This points to the possibility of a general weakness in the ability of teachers to accurately assess and interpret students' presenting characteristics. Such knowledge is essential for classroom teachers, as it is necessary for them to understand and accurately interpret learner characteristics if they are to implement appropriate adjustments (de Bruin, Graham, & Gallagher, 2020).

The findings of this study have important implications for teacher education and professional learning. Solid understanding of child development and the science of reading is essential for teachers to accurately interpret children's presenting characteristics, as well as determine how best to respond when progress is sub-optimal (Klug, Bruder, & Schmitz, 2016). Our findings are consistent with a recent call for a review of the emphasis placed on the practical components of teacher education to ensure adequate grounding in the theoretical knowledge necessary to understand and teach all five components of reading, including phonics, explicitly and with fidelity (Buckingham & Meeks, 2019). A necessary complement is knowledge of child behaviour that goes beyond the mechanics of classroom management strategies, as behaviour is widely regarded as a form of communication that can help guide teachers' practice, but only if they know how to interpret it (Wolff, Jarodzka, & Boshuizen, 2017). Better integration of units in child development, behaviour management, and inclusive education across undergraduate teacher education programs, as well as high-quality ongoing professional learning in language development and reading instruction, could help strengthen teachers' knowledge and skills in these critically important areas (Stark et al., 2016). These improvements in teacher education and professional learning would be supported by the use and interpretation of validated assessment tools to assist teachers to identify reading problems at the sub-component level. Education departments and governments can play a vital role by implementing a word-level reading task similar to the Test of Word Reading Efficiency (TOWRE-2; Torgesen et al., 2012), such as the Phonics Screening Check, to assist teachers to identify and address decoding weaknesses (Wheldall, Bell, Wheldall, Madelaine, & Reynolds, 2019), and by commissioning research to independently assess the effectiveness of common reading practices and promoting greater use of those found to be effective.

Conclusion

Early difficulties in reading are linked to difficulties accessing the academic school curriculum with the potential to lead to disengagement, disruptive behaviour, and early school leaving. In addition to providing high-quality initial reading instruction, it is critical that schools accurately identify and address early reading difficulties using relevant and evidence-based interventions that appropriately target children's learning needs. This process begins with classroom teachers who are responsible for making reasonable adjustments and referring children for further intervention. Our findings point to possible cracks in school support systems with some children about whom teachers expressed concerns and who persistently achieved below average not being provided any additional support. Variability in teachers' reported concerns suggests that some children's externalising behaviours may distract teachers from reading difficulties. Our data on the supports provided suggests these children receive generic or behaviour-related supports, like floating teacher aide time, wobble chairs or a behaviour plan, as opposed to skilled reading-related support. Findings from this research suggest that more fine-grained evidence-based assessment tools are necessary to identify children experiencing difficulties early in the process of learning to read. Such tools need to highlight specific weaknesses to ensure that strengths in visual word memorisation at the earliest stages of reading do not mask problems in phonemic decoding. Further, it is critical that teachers are provided with initial teacher education and professional learning opportunities that will support them to correctly interpret and respond to children's presenting characteristics. Finally, more research is urgently needed to investigate possible links between teachers' perceptions of and responses to children's behaviour and gaps in the provision of support for early learning difficulties, including reading.

Acknowledgements This research was funded by the Financial Markets Foundation for Children (2013-030) and the Australian Research Council Discovery Projects Scheme (DP160100319). We also thank the anonymous reviewers of this manuscript for their comprehensive, insightful and valuable feedback.

References

Adoniou, M. (2017). How the national phonics test is failing England and why it will fail Australia too [Blog post]. Retrieved from https://www.aare.edu.au/blog/?p=2533.

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100,* 235–251. https://doi.org/10.1037/0022-0663.100.2.235.

Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., et al. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of Abnormal Child Psychology, 33,* 205–217. https://doi.org/10.1007/s10802-005-1828-9.

Australian Literacy Educators' Association. (2015). Literacy in the 21st Century: The ALEA Declaration. Retrieved from https://www.alea.edu.au/aboutus/alea-declaration.

Berg, B. L. (2001). *Qualitative research methods for the social sciences*. Boston, MA: Allyn and Bacon.

Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., et al. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 29,* 435–451. https://doi.org/10.1007/s11145-015-9608-6.

Buckingham, J., & Meeks, L. (2019). *Short-changed: Preparation to teach reading in initial teacher education.* Research Report July 2019. Retrieved from https://www.fivefromfive.org.au/wp-content/uploads/2019/07/ITE-REPORT-FINAL.pdf.

Castles, A., Polito, A., Pritchard, S., Anandakumar, T., & Coltheart, M. (2018a). Do nonword reading tests for children measure what we want them to? An analysis of year 2 error responses. *Australian Journal of Learning Difficulties, 23,* 153–165. https://doi.org/10.1080/19404158.2018.1549088.

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

Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2000). Early reading-related skills and performance, reading self-concept, and the development of academic self-concept: A longitudinal study. *Journal of Educational Psychology, 92,* 703–708. https://doi.org/10.1037/0022-0663.92.4.703.

Childs, G., & McKay, M. (2001). Boys starting school disadvantaged: Implications from teachers' ratings of behaviour and achievement in the first two years. *British Journal of Educational Psychology, 71,* 303–314. https://doi.org/10.1348/000709901158532.

Cohen, N. J., Davine, M., Horodezky, N., Lipsett, L., & Isaacson, L. (1993). Unsuspected language impairment in psychiatrically disturbed children: Prevalence and language and behavioral characteristics. *Journal of the American Academy of Adolescent Psychiatry, 32,* 595–603. https://doi.org/10.1097/00004583-199305000-00016.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934–945. https://doi.org/10.1037/0012-1649.33.6.934.

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities, 39,* 507–514. https://doi.org/10.1177/00222194060390060301.

de Bruin, K., Graham, L. J., & Gallagher, J. (2020). What is the NCCD and what does it mean for my practice? In L. Graham (Ed.), *Inclusive Education in the 21st Century* (Chap. 6, pp. 122–155). Sydney: Allen & Unwin.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167–188. https://doi.org/10.1207/s1532799xssr0902_4.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71,* 393–447. https://doi.org/10.3102/00346543071003393.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention* (2nd ed.). New York, NY: The Guilford Press.

Francis, D. A., Caruana, N., Hudson, J. L., & McArthur, G. M. (2019). The association between poor reading and internalising problems: A systematic review and meta-analysis. *Clinical Psychology Review, 67,* 45–60. https://doi.org/10.1016/j.cpr.2018.09.002.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making "secondary intervention" work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing: An Interdisciplinary Journal, 21,* 413–436. https://doi.org/10.1007/s11145-007-9083-9.

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0089900.

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing: An Interdisciplinary Journal, 25,* 1819–1845. https://doi.org/10.1007/s11145-011-9333-8.

Gibson, H., & England, J. (2016). The inclusion of pseudowords within the year one phonics 'Screening Check' in English primary schools. *Cambridge journal of education, 46,* 491–507. https://doi.org/10.1080/0305764X.2015.1067289.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10. https://doi.org/10.1177/074193258600700104.

Government of South Australia Department for Education (2019). *Phonics screening check fact sheet*. Retrieved from https://www.education.sa.gov.au/sites/default/files/2018-phonics-screening-check-fact-sheet.pdf.

Graham, L. J. (2008). Drugs, labels and (p) ill-fitting boxes: ADHD and children who are hard to teach. *Discourse: Studies in the Cultural Politics of Education, 29*(1), 85–106. https://doi.org/10.1080/01596300701801377.

Graham, L. J. (2015). A little learning is a dangerous thing: Factors influencing the increased identification of special educational needs from the perspective of education policy-makers and school practitioners. *International Journal of Disability, Development and Education*, *62*(1), 116–132.

Graham, L. J., & Tancredi, H. (2019). In search of a middle ground: the dangers and affordances of diagnosis in relation to attention deficit hyperactivity disorder and developmental language disorder. *Emotional and Behavioural Difficulties*, *24*(3), 287–300.

Graham, L. J., Tancredi, H., Willis, J., & McGraw, K. (2018). Designing out barriers to student access and participation in secondary school assessment. *The Australian Educational Researcher*, *45*(1), 103–124.

Hastings, P. (2012). Early career teachers' self-efficacy for balanced reading instruction. *Australian Journal of Teacher Education, 37*(6), 55–72. https://doi.org/10.14221/ajte.2012v37n6.2.

Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing: An Interdisciplinary Journal, 15,* 789–809. https://doi.org/10.1023/A:1020985701556.

Hill, S. (2017). Early reading and practice-inspired research. *Early Child Development and Care, 187,* 322–330. https://doi.org/10.1080/03004430.2016.1234468.

Hjetland, H. N., Lervåg, A., Lyster, S.-A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology, 111,* 751–763. https://doi.org/10.1037/edu0000321.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly, 25,* 4–32. https://doi.org/10.1080/10573560802491208.

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties*. Hoboken, NJ: Wiley.

Klug, J., Bruder, S., & Schmitz, B. (2016). Which variables predict teachers' diagnostic competence when diagnosing students' learning behavior at different stages of a teacher's career? *Teachers and Teaching, 22*(4), 461–484. https://doi.org/10.1080/13540602.2015.1082729.

Language and Reading Research Consortium, & Chiu, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education, 39,* 289–303. https://doi.org/10.1177/0741932518762055.

Lyon, G. R. (2002). *Learning disabilities and early intervention strategies.* Retrieved from http://reidlyon.com/edpolicy/11-LEARNING-DISABILITIES-AND-EARLY-INTERVENTION-STRATEGIES.pdf.

Marinus, E., Kohnen, S., & McArthur, G. (2013). Australian comparison data for the test of word reading efficiency (TOWRE). *Australian Journal of Learning Difficulties, 18,* 199–212. https://doi.org/10.1080/19404158.2013.852981.

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education, 65,* 48–60. https://doi.org/10.1016/j.tate.2017.02.021.

Mills, K. A. (2015). *Literacy theories for the digital age: Social, critical, multimodal, spatial, material and sensory lenses*. Bristol: Multilingual Matters.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/pirls2016/international-results/.

Organisation for Economic Cooperation and Development. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing. https://doi.org/10.1787/9789264266490-en.

Partanen, M., & Siegel, L. S. (2014). Long-term outcome of the early identification and intervention of reading disabilities. *Reading and Writing: An Interdisciplinary Journal, 27,* 665–684. https://doi.org/10.1007/s11145-013-9472-1.

Peng, P., Fuchs, D., Fuchs, L. S., Elleman, A. M., Kearns, D. M., Gilbert, J. K., et al. (2018). A longitudinal analysis of the trajectories and predictors of word reading and reading comprehension development among at-risk readers. *Journal of Learning Disabilities, 52,* 1–14. https://doi.org/10.1177/0022219418809080.

Shapiro, L. R., & Solity, J. (2016). Differing effects of two synthetic phonics programmes on early reading development. *British Journal of Educational Psychology, 86,* 182–203. https://doi.org/10.1111/bjep.12097.

Snow, P. C. (2016). Elizabeth Usher Memorial Lecture: Language is literacy is language - Positioning speech-language pathology in education policy, practice, paradigms and polemics. *International Journal of Speech-Language Pathology*, *18*(3), 216–228.

Spira, E. G., Bracken, S. S., & Fischel, J. E. (2005). Predicting improvement after first-grade reading difficulties: The effects of oral language, emergent literacy, and behavior skills. *Developmental Psychology, 41,* 225. https://doi.org/10.1037/0012-1649.41.1.225.

Stark, H. L., Snow, P. C., Eadie, P. A., & Goldfeld, S. R. (2016). Language and reading instruction in early years' classrooms: The knowledge and self-rated ability of Australian teachers. *Annals of Dyslexia, 66*(1), 28–54.

Swanson, E., Stevens, E. A., Scammacca, N. K., Capin, P., Stewart, A. A., & Austin, C. R. (2017). The impact of tier 1 reading instruction on reading outcomes for students in Grades 4–12: A meta-analysis. *Reading and Writing: An Interdisciplinary Journal, 30,* 1639–1665. https://doi.org/10.1007/s11145-017-9743-3.

Thomson, S., Hillman, K., Schmid, M., Rodrigues, S., & Fullarton, J. (2017). *Reporting Australia's results: PIRLS 2016.* Retrieved from Australian Council for Educational Research website: https://research.acer.edu.au/pirls/1/.

Torgesen, J. K. (2004). Preventing early reading failure. *American Educator, 28*(3), 6–9.

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

Vaughn, S., Roberts, G., Capin, P., Miciak, J., Cho, E., & Fletcher, J. M. (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children, 85,* 180–196. https://doi.org/10.1177/0014402918782618.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing: An Interdisciplinary Journal, 21,* 437–480. https://doi.org/10.1007/s11145-007-9098-2.

Wheldall, K., Bell, N., Wheldall, R., Madelaine, A., & Reynolds, M. (2019). Performance of Australian children on the English Phonics Screening Check following systematic synthetic phonics instruction in the first two years of schooling. *Australian Journal of Learning Difficulties*. https://doi.org/10.1080/19404158.2019.1635500.

Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education, 66,* 295–308. https://doi.org/10.1016/j.tate.2017.04.015.