Australian Education Research Organisation

MONASH University

Supporting students significantly behind in literacy and numeracy

A review of evidence-based approaches

May 2023

This study was commissioned by the Australian Education Research Organisation (AERO) and undertaken by a team led by Dr Kate de Bruin at Monash University.

Authors

Dr Kate de Bruin, Dr Eugénie Kestel, Dr Mariko Francis, Professor Helen Forgasz, Ms Rachelle Fries

Acknowledgements

We would like to express our gratitude for the work of Sylvia Pilz, Education Librarian, who provided expert input into the literature search conducted during our umbrella review. We thank AERO for funding this important work, as well as its team members Tess Marslen, Greta Holmberg, Adam Inder and Daniel Carr, whose feedback on drafts and facilitation of the anonymous peer review greatly strengthened this work. We acknowledge the valuable input of the team at Monash University, in particular Jason Zheng, Joanne Tymms and Carline Hamdillah, for their technical assistance throughout the project. We also thank our families, colleagues, friends and pets for offering endless support and patience as we undertook this extensive work within a rapid timeframe.

Suggested citation

de Bruin K, Kestel E, Francis M, Forgasz H and Fries R (2023), Supporting students significantly behind in literacy and numeracy: a review of evidence-based approaches, edresearch.edu.au

Publication detail

ISBN 978-1-923066-01-4 (online)

Australian Education Research Organisation

MONASH University

Contents

Figures

Tables

Executive Summary

Background

In Australia, a substantial proportion of students start secondary school with literacy and numeracy skills that are 3 or more years below those of their peers (ACARA 2021). Evidence suggests that low literacy and numeracy skills hinder students' access to the curriculum (Shinn et al. 2016) and can result in poor progress or educational failure. These students are also more likely to face poor outcomes post-school, such as in their health and employment (Australian Institute of Health and Welfare 2020).

A large body of research has been conducted on literacy and numeracy instruction and interventions for primary school-aged students, as fundamental skills in these domains are generally taught as part of the primary school curriculum. However, less is known about the applicability or effectiveness of interventions for older students who have not mastered these foundational literacy or numeracy skills. Questions also remain about the feasibility of implementing literacy and numeracy interventions within secondary schools, where teachers are typically less familiar with teaching these basic skills and where scheduling constraints exist. In this report, we present the findings of an umbrella review to address these gaps in knowledge.

Literacy and numeracy

For the purposes of our review, we divided literacy into the domains of reading and writing, and we examined the research for interventions that targeted the components of each. For reading, the components included phonemic awareness, decoding, fluency, vocabulary and comprehension. For writing, transcription (handwriting, typing, spelling), text generation (phrases, sentences and compositions) and executive functions (planning, drafting and revising) were included. Mathematics is defined within the Australian Curriculum as the study of number and algebra, measurement and geometry, and statistics and probability, while numeracy is the application of these skills to other learning areas and to meeting societal demands. We examined interventions for older students related to the acquisition of basic skills across these strands of the mathematics curriculum.

Service delivery frameworks

We examined multi-tiered frameworks such as response to intervention (RTI) and multi-tiered system of supports (MTSS) that have replaced traditional service delivery models internationally, such as in the United States, where multi-tiered frameworks were developed. Traditionally, in Australia and many other countries, support services for underachieving students have been resourced and delivered through a categorical framework that funds and organises services based on students' identification as belonging to priority equity cohorts, such as Aboriginal and Torres Strait Islander students, students with disability, students from low-socioeconomic backgrounds, students living in regional or remote areas, and students who speak English as an additional language or dialect. This approach, while well-intended, is known to create equity gaps, as some underachieving students are ineligible for support, while others can experience significant delays before services can be provided. In the US, to address these issues with the categorical approach, research was conducted to consider how multi-tiered frameworks, such as those used in public health, could be recast into education to address the inequities and delays of categorical approaches in education. To undertake our review of contemporary approaches most appropriate for organising and delivering supports to underachieving students, we examined these multi-tiered frameworks to identify the implications for Australia.

Aims and focus

We conducted our review to inform resources being prepared by the Australian Education Research Organisation (AERO) to assist schools in supporting students significantly behind their peers in Years 7 to 9. Our aim was to identify effective practices for supporting students who are substantially behind their peers in reading, writing and mathematics. We have included systematic reviews that examined contemporary frameworks for guiding the delivery of interventions, systematic reviews of the identification and monitoring of student progress, and systematic reviews of the effectiveness of reading, writing and mathematics interventions.

Methods

We used the protocols and appraisal tools of JBI Global (Aromataris et al. 2020) to ensure the literature search was exhaustive, that bias was minimised, and that the process was methodologically rigorous. We applied 4 essential criteria in searching and screening the systematic reviews that we identified. A 2-stage process was used to determine the eligibility of systematic reviews based on their research methodology, population of interest, phenomenon of interest and context (PICo; Aromataris et al. 2020). Studies were critically appraised for quality and only those studies meeting a pre-determined threshold for quality were included. The results were extracted from all included studies. Extracted data were collated to generate findings to the 5 research questions specified by AERO.

Results

A total of 69 systematic reviews was included in the final sample. The results from each systematic review were extracted and synthesised to answer the 5 research questions set by AERO. These 5 questions, and a synthesis of the results that we extracted, are outlined below.

Research question 1

What are the contemporary supporting frameworks or models that guide the delivery of intensive supports to struggling students (for example, RTI or MTSS models)? Have these frameworks been tested or evaluated?

Multi-tiered frameworks are characterised by key elements including:

- using evidence-based instructional practices across a sliding scale of tiered supports that increase in intensity
- collaboration among school-based teams
- using student data to make decisions.

From the meta-analyses identified in which multi-tiered frameworks were evaluated (Burns and Symington 2002; Burns et al. 2005), we found that when instruction was high-quality at Tier 1, an average of 80% of students met benchmarks for progress. The majority of the remaining 20% of students, who were identified with learning gaps, benefited from effective, targeted and timely intervention at Tier 2 that was facilitated by collaborative teams using a problem-solving approach. Only 6% of students needed even more intensive and individualised intervention at Tier 3. Notably, only 1.68% of students were referred for a special education placement as compared to the national average of 5.7%.

Multi-tiered frameworks have shifted in structure and scope in the 2 decades of their legislation and implementation in the US. Two systematic reviews of policy documentation in the US (Berkeley et al. 2009; 2020) show that over time, multi-tiered frameworks have become more comprehensive and the importance of effective Tier 1 instruction in the classroom to prevent learning gaps has been emphasised. Originally, multi-tiered frameworks for students making low academic progress were conceived as a narrow RTI approach, with its conceptual roots in special education. There was an early emphasis on minimising the proportions of students needing a special education placement and monitoring to identify students needing academic intervention (primarily in reading) at Tiers 2 and 3. Within RTI there was less attention to Tier 1 or to the many factors contributing to student achievement that can influence the success and impact of instruction. Over time, there has been a shift towards comprehensive MTSS that also incorporates a focus on student behaviour, engagement and wellbeing at school. Importantly MTSS is distinct from RTI. MTSS is conceptualised as a general education initiative in which it is recognised that Tier 1 is the most important tier of support as it plays the greatest role in success at school for all students.

Berkeley et al. (2009, 2020) found that following the embedding of multi-tiered frameworks within US legislation, there were critical factors that contributed to, and hindered, the system-wide scaling and consistency of implementation across the states. These researchers particularly noted the beneficial role of technical assistance centres that provided professional learning and implementation resources. Technical assistance centres were funded by the federal government and often began as university partnerships that have since become national resources, such as the Vanderbilt Iris Center. Barriers to systemic scaling and consistency of implementation across the US included the provision for state autonomy to opt in to using multi-tiered frameworks, as well as autonomy in developing state-based variations. These led to discrepancies between states in definitions and models for implementation. Our recommendations for Australia are:

- that a consistent approach be adopted from the outset to avoid the discrepant conceptual and implementation issues that arose in the US
- to emulate the successful initiatives in the US for setting up quality professional learning and technical assistance through partnerships with universities.

Research question 2

What are the recommended approaches for identifying students who require intensive support? For example, when and how frequently should students be assessed and re-assessed?

The focus of several of the systematic reviews included in our umbrella review was on curriculum-based measures (CBM) for identifying students who require intervention (Ardoin et al. 2013; Filderman et al. 2021; Jung et al. 2018; Kilgus et al. 2014; Newell et al. 2002; Reed and Cummings 2014; Shin and McMaster 2019). CBM are brief and basic assessments of students' progress through the reading, writing and mathematics curriculum that can be used frequently, such as weekly or biweekly and typically take between one and 5 minutes to administer. Students' performances can be graphed and their trajectory can show whether their progress is meeting expectations.

Most of the systematic reviews we examined on assessment focused on CBM in relation to reading. While generally used to track the progress of students in the younger years, these measures can also be used to quickly assess older students and compare their fluency to norms for expected achievement. The systematic reviews that we examined found that oral reading fluency CBM is also a robust and valid means of screening older students for general reading ability (Ardoin et al. 2013; Kilgus et al. 2014; Shin and McMaster 2019). This is because, in addition to assessing their fluency, it will reflect whether students have accumulated competency in lower-order skills such as phonemic awareness and decoding, as well as vocabulary. Oral reading fluency checks can be run at any time, making them a more useful screener for underachievement than a standardised reading comprehension test result, which can only show if students have accumulated competency across the 5 components of reading, and is more time-consuming to run.

We found that using oral reading fluency CBM to identify older students who may need Tier 2 intervention requires the schools to use a cut score as a threshold for minimum competency (Ardoin et al. 2013; Kilgus et al. 2014) in identifying which students are furthest behind. Identified students are likely to need further investigation of their underlying competency across the domains of reading instruction through a problem-solving approach (Burns and Symington 2002; Burns et al. 2005) to identify the most appropriate intervention to address their underlying skill deficits. We also found that this cut score should be based on local norms rather than national norms (Kilgus et al. 2014) and that consideration of first-language competency for students who speak English as an additional language be built into the interpretation of data (Newell et al. 2002).

While effective for screening at Tier 1, CBM in reading were not found to be sufficiently robust to monitor the progress of students receiving Tier 2 intervention (Ardoin et al. 2013). However, CBM across reading, writing and mathematics were found to be effective for teachers to use in data-based individualisation of instruction for Tier 3 support (Filderman et al. 2021; Jung et al. 2018). For example, based on student data, teachers can engage in individualisation such as adjusting the dosage or frequency of instruction, the precision or comprehensiveness of content taught, or attending to acquisition or transfer of knowledge and skills. Doing so resulted in improved reading, writing, spelling and mathematics outcomes for students needing this form of individualised Tier 3 support (Jung et al. 2018). When teachers had engaged in coaching and high-quality professional learning in how to use data in this way, the benefits for students were even stronger (Filderman et al. 2021).

Research question 3

What are the recommended approaches for supporting struggling students to improve their literacy and numeracy skills and reduce the gap between them and their peers? What level of intervention (such as duration, intensity and frequency) do these approaches require?

We found strong evidence of the effectiveness of school-based interventions to teach foundational skills in reading, writing and mathematics for secondary school students.[1] In relation to reading, lower-order reading skills (for example, phonemic awareness and decoding) were only examined in one systematic review. It was found that morphological instruction was beneficial. Intervening to improve higher-order skills (for example, fluency, vocabulary and comprehension) was the focus in the remaining systematic reviews. Instructional practices, such as explicit instruction, morphemic analysis, and instruction, were found to have clear benefits for improving students' vocabulary. There were also benefits arising from interventions to improve fluency that incorporated explicit reading instruction and repeated reading lessons. Comprehension outcomes were improved using explicit instruction to teach specific skills (for example, identifying the main idea) and strategies (for example, summarising), as well as through teaching students to use graphic organisers (for example, story maps, cognitive maps) for mapping the content of text and conceptual relationships.

Similarly, we identified evidence of effective practices to use in interventions to improve writing skills. Students' capacity to produce written text in phrases, sentences and compositions was improved through explicit instruction in particular strategies, such as planning, writing and editing. Students were found to acquire skills in spelling, generation of text and coordination of their executive functions, using effective practices such as explicit teaching, untimed tasks and graphic organisers. Students were able to learn to execute writing skills with more ease and automaticity when they were provided with interventions involving goal setting, timed writing tasks and performance feedback. Multicomponent interventions in which several practices were combined were effective, such as self-regulated strategy development (SRSD). Instructional practices were also effective when used in conjunction with graphic organisers, if students received explicit instruction in how to use them, engaged in guided practice, and received feedback to support their learning. There was some evidence that students' spelling was improved through morphological instruction.

1 There was a very large number of systematic reviews that contributed to the results reported in this section. In the interests of brevity, we have not cited each systematic review to support the findings summarised here. We therefore recommend that the results for Research question 3 be read in conjunction with the summary of strategies and supporting citations found in Appendix A.9.

Several effective intervention practices were identified for use in mathematics interventions for secondary school students. Beneficial strategies were consistent across the targeted improvement area of general mathematical skills, as well as for specific mathematics skills such as word problem solving, algebra, fractions, basic number facts and arithmetic skills.

As for reading and writing, instruction that was explicit and systematic was found to be effective for improving overall mathematics skills, algebra, fractions and arithmetic skills.

Explicit teaching strategies were also beneficial for teaching structures and relationships, as well as word problem solving skills. Schema-based instruction, involving the explicit teaching of underlying mathematical structures, was beneficial for teaching word problem solving skills. Students also benefited from using visual or physical models in their overall mathematics performances. These models also improved students' outcomes for fractions, word problem solving skills and algebra. Using the concrete-representational-abstract (CRA) approach, improvement was found for students' basic recall of number facts, algebra, arithmetic and general mathematics skills. Interventions were also beneficial when effective instructional design features beyond explicit instruction were incorporated. For example, providing logically sequenced examples was found to be beneficial for algebra, and moving from guided to independent practice was beneficial for improving overall mathematics, fractions, algebra, word problem solving and arithmetic skills.

Research question 4

What are the most effective ways to implement these approaches, and what preconditions are necessary to support their implementation?

- **Who should be implementing these approaches? For example, primary trained teachers, classroom assistants, speech therapists.**
- **What resources and supports do schools need to implement the recommended approaches with fidelity to the duration, intensity and frequency that is required? For example, time structures, professional development opportunities and age-appropriate materials.**

One important finding was that team-based problem-solving should be a precondition for the selection and implementation of tiered interventions. These team-based collaborations should involve inter-professional collaboration, such as school educators working with school psychologists. They should also include consultation and observations (Burns and Symington 2002) as well as conferencing to select evidence-based interventions and monitor student progress based on their responsiveness to Tier 1 instruction (Burns et al. 2005).

In most of the systematic reviews included, the focus was on intervention practices as the phenomenon of interest, rather than on the implementer, the resources or the scheduling of interventions. We found no evidence to suggest that primary trained teachers, speech therapists or teaching assistants should run interventions for secondary school students underachieving in literacy or mathematics.

Two resources were identified in our umbrella review as worthy of investment by schools. The first was CBM. These simple and efficient measures for screening students' progress are based on the curriculum being taught. CBM used for screening should have sufficient predictive power to anticipate which students are likely to perform below benchmarks on standardised assessments of reading, writing or mathematics, and thus indicate if students may need further intervention. As noted above, strong evidence was found for the validity of oral reading fluency CBM used for this purpose with older students (Ardoin et al. 2013; Kilgus et al. 2014; Shin and McMaster 2019). In addition, there was also evidence that CBM for reading, writing and mathematics can be used to intensify instruction for older students needing highly individualised, Tier 3 support (Jung et al. 2018).

Our findings suggest that schools would be wiser to acquire existing CBM rather than develop their own. This is based on the caution offered by Jung et al. (2018) in their systematic review, which found no evidence for the technical adequacy of teacher-generated CBM. In the US, CBM that have been thoroughly evaluated

can be provided to schools by their districts or purchased from commercial or university-based organisations. The National Center on Intensive Intervention (n.d.) Academic Screening Tools Chart is a useful resource for schools to use in the selection of appropriate academic screening tools, as the technical adequacy of each CBM is reported and their appropriateness for particular age groups and curriculum domains easily identified.

The second resource worthy of investment by schools was quality professional learning, which can improve both teacher practice and student outcomes. This finding arose from the systematic reviews that considered the impact of fidelity of implementation of interventions or evaluated the impact of professional learning. Several systematic reviews indicated that intervention outcomes were optimised and delivered with greater fidelity when implementers had received professional learning to develop their knowledge and expertise (Fallon et al. 2015; Jung et al. 2018; Scammacca et al. 2007). Quality features of professional learning that were supported by evidence included regular sessions involving individual in-person contact (Fallon et al. 2015; Jung et al. 2018) and using active and collaborative learning approaches (Filderman et al. 2021; Jung et al. 2018) with benefits for teachers' beliefs in the value of the instructional practice as well as their skills in implementation (Filderman et al. 2021).

Based on the collective findings of these systematic reviews, schools adopting and using MTSS can optimise intervention outcomes for students underachieving in reading, writing or mathematics by investing in quality professional learning that includes face-to-face regular contact and consultation to upskill teachers.

Research question 5

Are there major gaps in the evidence base that prevent these questions being fully addressed? What research could be undertaken to fill these gaps?

Several gaps in research evidence emerged from our umbrella review. The evaluations of multi-tiered frameworks were somewhat dated, pointing to the need for more up-to-date research to gain a contemporary picture of the impact of this framework relative to other educational practices now in use. Further, systematic reviews on improving reading were dominant; that is, there were fewer systematic reviews found in which the focus was on improving writing or improving mathematics. Among the systematic reviews on reading improvements, the largest subset had improvement of comprehension skills as the main focus. There were no systematic reviews identified for screening and/or progress monitoring of writing and mathematics interventions beyond data-based instruction (DBI). Clearly more research on these topics is needed, as is research conducted in Australia, which was also lacking in our review.

An additional gap is the need for high-quality research evaluations and syntheses. Many systematic reviews were excluded as they did not meet the criteria for quality within the critical appraisal process. This process is essential to ensure the validity of the findings. Some systematic reviews were excluded due to the low research standards inherent in them. Others were excluded as some of the research studies that they reviewed were of very low quality. In both cases, there were threats to the validity of the reported results. As a consequence of the exclusions, we identified the need for more well-designed, methodologically rigorous, experimental studies and systematic reviews to identify intervention effectiveness across all domains of reading, writing and mathematics.

Conclusion

Despite the gaps identified in the research base, and the dated systematic reviews in some areas, there is sound evidence to support the introduction of MTSS in Australian school systems. The evidence suggests that with consistent use of effective instruction at Tier 1, a team-based problem-solving approach to selecting evidence-based intervention and implementing these with fidelity at Tier 2, 95% of students could meet academic benchmarks. This would reserve intensive Tier 3 support only for a small proportion of students in need and reduce the number of special education placements, as was found in the US.

Our findings point to the importance of consistency across jurisdictions to ensure system-wide scaling of MTSS is successful. They also emphasise the importance of ensuring the impact of instruction and intervention for underachieving students is maximised through robust CBM and professional learning.

Our findings further highlight the effectiveness of developing and resourcing technical assistance centres, as well as using materials developed by those exemplary states in the US that are regarded as the benchmark for quality MTSS.

Clear evidence is presented to indicate how secondary schools can best intervene to close achievement gaps in reading, writing and mathematics. The effective practices that were identified can be considered an excellent starting point for use in secondary schools seeking to support these students.

In addition to upskilling secondary school staff in using screening data well, we recommend professional learning for teachers in effective practices for reading, writing and mathematics interventions, such as:

- explicit instruction
- strategy instruction
- using graphic organisers.

We recommend that this investment be made schoolwide, rather than being reserved for the implementation of interventions. Instruction at Tiers 2 and 3 of support should not be fundamentally different from instruction at Tier 1. The effective practices we identified for reading, writing and mathematics interventions are not special educational practices; they are quality teaching practices and offer benefit in all classrooms for all students, as well as benefiting those needing more support.

The MTSS framework is a general education initiative. It is reliant on strong, regular classroom instruction at Tier 1 for all students as the most important foundation, and this should be sufficient for most students to succeed at school. If MTSS were implemented at scale across schools, our research suggests this would minimise the number of students reaching secondary school needing support. At the present time, too many students arrive in secondary school substantially behind their peers, and too many of these are students who are already disadvantaged. Our review offers clear advice about how this can be addressed and shows that it is never too late to teach these students.

To achieve this, effective intervention selected by collaborative teams and implemented by appropriately skilled staff is needed. Students should be identified as quickly as possible, using data from effective screening measures. These data should be examined by multidisciplinary teams and a problem-solving approach adopted to select evidence-based interventions aligned to the identified skills needing improvement. This will help the most vulnerable and disadvantaged students to be as well-prepared as possible to flourish in life beyond school, and make our schools places where every student can thrive.

Introduction

Background to the project

In this report, we present the findings of an umbrella review conducted for the Australian Educational Research Organisation (AERO). In its project 'implementing effective tiered interventions in secondary schools', AERO notes that schools describe a lack of clarity or guidance regarding how they should support underachieving older students in literacy and numeracy. Specifically, secondary schools report confusion about how best to identify those students who are the furthest behind in literacy and numeracy. They also report a lack of clear direction regarding resourcing and implementating interventions, and educational services targeting foundational skills, to address gaps in learning.

In addition, minimal system guidance on how funds should be allocated for running intervention for students, and the specific resources that support its implementation, also contribute to confusion among teachers and leaders on how to intervene. Accordingly, we conducted this review by locating high-quality systematic reviews and synthesising the findings to inform guidance for schools to deliver educational supports and interventions for students in Years 7 to 9. In our review, we focus on what is suggested by the evidence regarding how schools should identify underachieving students, what interventions they should provide and the conditions that produce the optimal outcomes.

Literature review

Literacy and numeracy

Underachievement in literacy and numeracy hinders secondary school students' potential to learn and succeed at school, and places them at risk of educational failure. As well, their long-term outcomes in adult life, such as in health, employment and community participation, are jeopardised (Australian Institute of Health and Welfare 2020).

Closing achievement gaps for these students is vital if they are to graduate from school equipped with essential literacy and numeracy skills. The most efficient means to address literacy and numeracy achievement gaps is to prevent them from widening in the primary school years. Preventing gaps is best accomplished by providing the highest-quality early literacy and numeracy instruction that is effective for the most diverse cohort to acquire foundational literacy and numeracy skills, so students can execute these skills fluently.

The instructional practices recommended to achieve this for younger students are to teach skills and strategies systematically, explicitly, directly and following a clear scope and sequence. This also includes incorporating daily instruction with opportunities to practise through a gradual release of responsibility (Foorman et al. 2016; Frye et al. 2013; Gersten et al. 2009; Siegler et al. 2010; Shanahan et al. 2010), with regular progress screening (Gersten et al. 2008; Gersten, Beckman et al. 2009). Providing effective instruction using these practices supports about 95% of children to learn (Foorman et al. 1998), although some may need supplementary and carefully targeted early intervention (Vellutino et al. 1996) to make satisfactory progress.

Without effective intervention and support, students who do not master foundational literacy and numeracy skills in the early years of school do not catch up (Francis et al. 1996; Juel 1988; Shaywitz et al. 1999). Students who are underachieving in reading, writing and mathematics therefore require targeted or intensive intervention to effectively address skill deficits and close achievement gaps; this becomes even more critical for older students (de Haan 2021).

In the upper primary and secondary years of schooling, instruction moves from teaching basic skills to a content-based curriculum, such as English, mathematics, science, languages and the arts. Success is therefore much more challenging for students who lack competency in foundational literacy and numeracy skills (Shinn et al. 2016). Because they are assumed, there is typically no more instruction in these foundational skills and learning is expected to progress through using these skills.

By secondary school, skills such as reading and writing are embedded in learning and assessment across all curriculum areas. Basic mathematics skills are also assumed across many domains, such as the sciences and humanities. Students lacking foundational skills can become significantly behind their peers, as compromised literacy or numeracy hinders their ability to access the curriculum or to make good progress in many other key learning areas. For older students experiencing learning gaps, this means targeted and timely support is particularly urgent to minimise gaps and enable them to attain the highest possible chance of post-school success.

To provide older students with support, secondary schools need to quickly identify those in need and provide them with the most efficient and effective educational supports and intervention. This requires a coordinated framework of service delivery for screening students, allocating resources, and delivering educational services and intervention (Triano 2000). For this reason, service delivery frameworks are a central focus of our review, through which we consider the evidence that should inform the guidance provided to schools. In the remainder of this section, we provide a literature review on literacy and numeracy instruction and intervention, the identification of underachieving students, and on service delivery frameworks.

We note that literacy has received more attention in research and policy than numeracy, and therefore constitutes a larger component of the review presented in this report. One reason for this increased attention is that an expanded research focus on literacy was catalysed by the United Nations (UN) Literacy Decade as part of the Education for All initiative (United Nations Educational 1990). This initiative sought to address global issues in the education of children, with particular attention paid to disadvantaged cohorts such as those living in poverty, from culturally and linguistically diverse communities, and those living with disability. Literacy and numeracy are broadly conceptualised terms and are often defined in contested ways. What is generally agreed is that they are deeply embedded within the human right to education, as articulated in the Universal Declaration of Human Rights (UN 1948). Literacy and numeracy are also conceptualised as essential within the basic foundations of education, taking the form of being able to read, understand and create written text, and to use mathematical concepts and processes to calculate answers to mathematical problems (Moretti and Frandell 2013). Given their foundational role in learning, for the purposes of this project we focus on reading and writing as the main components of literacy for older students who are achieving well below the expected standard and for whom these foundations are not in place. We also focus on the mathematical concepts and processes involved in calculation as constituting numeracy for the same reasons.

Literacy: reading and writing

Literacy is often described as comprising a broad range of skills such as speaking, listening, reading and writing. The Australian Curriculum, Assessment and Reporting Authority[2] (ACARA) defines literacy as 'the knowledge and skills students need to access, understand, analyse and evaluate information, make meaning, express thoughts and emotions, present ideas and opinions, interact with others and participate in activities at school and in their lives beyond school' (ACARA n.d.). This aspirational statement outlines what literate students should be able to do; however, it lacks detail in the specific subskills that should be taught to students. In our report, we examine literacy by dividing it into reading and writing, and separately review the research on how each is taught.

Of all the areas examined in this report, reading draws on the largest research base and therefore receives more attention. This is because a significant volume of research was generated following 3 national inquiries examining reading: the *Report of the National Reading Panel* in the US (National Institute of Child Health and Human Development 2000), the National Inquiry into the Teaching of Literacy in Australia (Rowe 2005),

2 ACARA is the independent authority responsible for the development of a national curriculum (Kindergarten to Year 12) and a national assessment program, including managing the National Assessment Program – Literacy and Numeracy (NAPLAN).

and the 2 systematic reviews commissioned by the United Kingdom Department for Education and Skills in 2006 (Rose 2006; Torgersen et al. 2006). The authors of these reports collectively agreed that research from diverse disciplines, including education and cognitive science, affirmed that learning to read within an alphabetic writing system such as English involves 5 instructional components. Based on the reports, the 5 essential and interconnected components of reading are listed below.

- **Phonemic awareness:**[3] knowledge of, and capacity to manipulate, the smallest distinct sounds (phonemes) in spoken words. It is a subset of phonological awareness skills and is the last to develop.
- **Phonics:**[4] learning and using the relationships between sounds and letter-symbols to sound out (decode) written words.
- **Fluency:**[5] the ability to read accurately, quickly and expressively. Fluent readers can focus on reading for meaning.
- **Vocabulary:**[6] the words children need to know to comprehend and communicate. Oral vocabulary is the words children recognise or use in listening and speaking. Reading vocabulary is the words children recognise or use in reading and writing.
- **Comprehension:** extracting and constructing meaning from written text using knowledge of words, concepts, facts and ideas (Hempenstall 2016).

A theoretical explanation of how reading skills are acquired is provided by Hoover and Tunmer (2018) in their Simple View of Reading (SVR) model. The SVR proposes that *reading comprehension*, the capacity to extract meaning from print, consists of 2 components: word recognition (the capacity to recognise printed words) and language comprehension (the ability to derive meaning from spoken words). The relationship between the 2 components is conceptualised in the SVR as:

decoding x language comprehension = reading comprehension

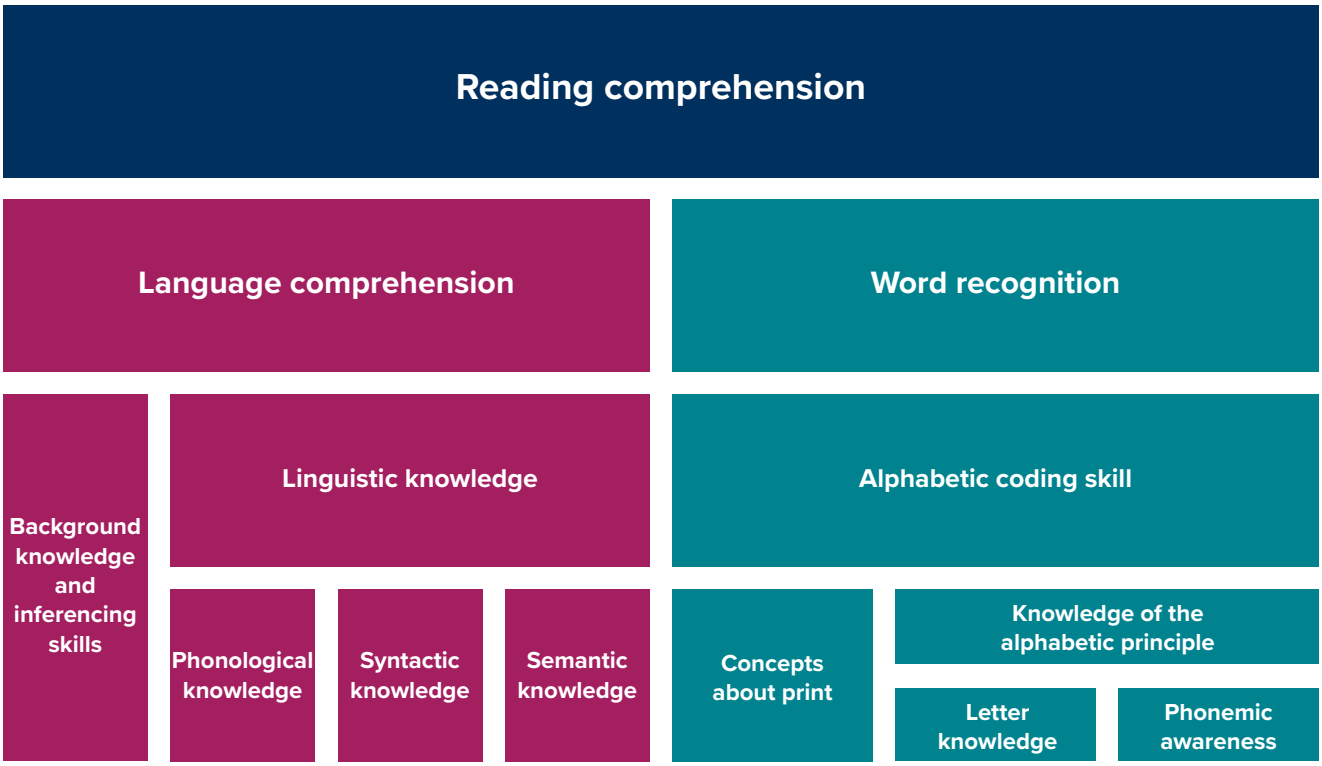
The SVR has been expanded within the Cognitive Foundations Framework (Tunmer and Hoover 2019) to illustrate the lower- and higher-order cognitive skills needed to achieve reading comprehension (see Figure 1). The word recognition domain consists of components including phonemic awareness and knowledge of letters and sounds, while the language comprehension domain consists of components including background knowledge, vocabulary and language structures (Tunmer and Hoover 2019). These 2 domains of reading work together and, as competency is achieved over time, readers become more skilled in reading accurately and fluently, resulting in competency in reading comprehension. Tunmer and Hoover (2019) suggest that at least some mastery of the lower-order skills involved in word recognition is required in order to develop higher-order skills. For example, unless students have understood the relationship between letters and sounds, they will not be able to decode printed words and extract meaning from written texts.

The evidence is clear that students need to be able to engage with these foundational reading skills and read words accurately several times before letters, sounds and words are *overlearned* and orthographically mapped[7], so reading is executed fluently and comprehension supported. These basic skills should be a key focus in the first 3 years of schooling. Research from the US shows that when this occurs, the percentage of students reading proficiently significantly improves (Biancarosa and Snow 2006). Effective reading instruction for students beyond

3 For more information, see https://www.readingrockets.org/teaching/reading101-course/modules/phonological-and-phonemic-awareness-introduction
4 See also, National Center on Improving Literacy (n.d.). https://improvingliteracy.org/kit/alphabetic-principle-phonics
5 See also, National Center on Improving Literacy (n.d.). https://improvingliteracy.org/kit/fluency-text
6 Further reading: National Center on Improving Literacy (n.d.). https://www.readingrockets.org/article/teaching-vocabulary
7 Orthographic mapping is a process involved in developing fluency. It involves the permanent storage of connections between sounds, letters and spelling within the oral language processing centre in the brain. The fusing of these connections frees up working memory and removes the hard work of decoding so that students ultimately recognise words instantly.

the early years should target higher-order skills to ensure progress in fluency and vocabulary development, as well as in advanced phonics and word study skills (Biancarosa and Snow 2006; Sedita 2011). For students who are still struggling with the more basic skills after Year 3, their deficits in these skills can compromise their ability to develop the higher-order skills, delaying their vocabulary development and their capacity to learn strategies for comprehending text. In our review, we examine what research shows about instruction in these reading skills for older, underachieving students.

Figure 1 The Cognitive Foundations Framework



Note: Reproduced with permission from Tunmer and Hoover (2019).

Compared to reading, there has been relatively less research attention paid to writing. This led to a call for a research agenda on writing (Graham and Perin 2007). Writing is widely understood to involve instruction in sentence construction and text construction. It involves knowledge of grammar, spelling and transcription (for example, handwriting and typing). To create sentences, paragraphs and larger texts such as essays and stories, students need to master skills in the creation of texts such as planning, drafting and organising ideas (Graham 2020). These are essential skills for students. In our research review, we consider what is known about writing skills and how skills are best taught, particularly to older students who are underachieving and for whom writing difficulties have become persistent.

Numeracy: mathematical skills and knowledge

The terms 'numeracy' and 'mathematics' are often inappropriately used synonymously in the Australian context. For example, the numeracy test included in the National Assessment Program – Literacy and Numeracy (NAPLAN) is actually a test of mathematics aligned with the Australian Curriculum: Mathematics: 'In the NAPLAN numeracy tests, the proportion of questions from each strand matches the proportion of the strand content in the Australian Curriculum: Mathematics' (ACARA 2016).

Numeracy is one of 7 general capabilities in the Australian Curriculum (ACARA n.d. b), and students are said to 'become numerate as they develop the knowledge and skills to use mathematics confidently across other learning areas at school and in their lives more broadly' (ACARA n.d. c). Numeracy is defined as encompassing 'the knowledge, skills, behaviours and dispositions that students need to use mathematics in a wide range of situations. It involves students recognising and understanding the role of mathematics in the world and having the dispositions and capacities to use mathematical knowledge and skills purposefully' (ACARA n.d. c). That is, numeracy involves the capacity 'to identify the knowledge and capabilities required to accommodate the mathematical demands of private and public life and to participate in society as informed, reflective, and contributing citizens' (Geiger et al. 2015:531).

By contrast, mathematics is an academic discipline. It is an abstract science involving number, quantity, and space; whether statistics is included under the umbrella of mathematics has been debated in academic circles. In the Australian Curriculum: Mathematics, students are provided 'with essential mathematical skills and knowledge in number and algebra, measurement and geometry, and statistics and probability' (ACARA n.d. a). In this report, we use the term mathematics (not numeracy) to represent the proficiencies that high-school students are expected to acquire in accordance with the Australian Curriculum: Mathematics (ACM).

As noted earlier, there has been less research and policy attention paid to mathematics for underachieving students than there has been on reading and writing. In 2006, the Human Capital Working Group of the Council of Australian Governments[8] (COAG) commissioned a review to inform best practices in teaching mathematics. Culminating in the *National Numeracy Review Report* (NNRR; Stanley 2008), it was found that an unacceptable proportion of students in Australia were failing to achieve basic proficiency levels, 'particularly but certainly not only amongst Indigenous students' (p. xii). The NNRR highlighted that achievement of competency in numeracy [that is, mathematics] by Australian students has important educational and labour market implications.

Despite receiving less attention from researchers, effective strategies and principles of intervention for students with learning difficulties and disabilities have been well established; however, the focus of most of the research studies has involved primary school-aged children (for example, Gersten, Beckman et al. 2009). It is important to establish whether the same strategies and intervention principles that are effective for younger children are also effective for students in secondary school. Accordingly, in this report we review research syntheses in which effective methods of assisting students underachieving in mathematics were investigated.

8 Until 2020, the COAG was the peak intergovernmental forum in Australia and comprised the prime minister, state premiers, territory chief ministers and the president of the Australian Local Government Association.

Identifying underachieving secondary school students

International data from the Programme for International Student Assessment (PISA, Thomson 2019) and the Trends in International Mathematics and Science Study (TIMSS, Thomson 2020) show that underachievement in reading and mathematics is a persistent and growing issue, particularly for secondary school students in Australia. These and other data sets, such as that from NAPLAN, indicate that this underachievement disproportionately affects priority equity cohorts including Aboriginal and Torres Strait Islander students, those from remote locations, low-socioeconomic backgrounds, and from language backgrounds other than English (ACARA 2021). Collectively, the data highlight the importance of intervening to address inequity and to reduce disadvantage. Recent research commissioned by AERO highlights that this will involve a mix of interventions targeting groups of disadvantaged students, such as those from priority equity cohorts, as well as individual students (AERO 2022). A focus on how these individual students might best be identified is included in our research review.

To offer interventions to individual underachieving students, secondary schools require effective processes to identify these students so appropriate intervention, aligned to each student's needs, is provided. While ACARA suggests that NAPLAN data are considered a reliable measure for the identification of individual students who are underachieving (ACARA 2016), not all agree. For example, some researchers have cautioned against applying the lowest bands of NAPLAN as a measure of underachievement, arguing that the bands are 'set too low to identify the stragglers' (Goss and Sonneman 2016:3).

Others argue that the wide margin of error within individual students' NAPLAN achievement scores makes these too unreliable for use in identifying underachieving students (Wu 2015). Moreover, there are large gaps in the data set as a number of students do not sit NAPLAN (Elliott, Davis and Kettler 2012) and many of these are students with disability (Dempsey and Davies 2013). While students with disability are identified under the National School Reform Agreement (NSRA; Council of Australian Governments 2018), with associated targets for improving their educational outcomes, their under-representation in NAPLAN means that for this priority equity cohort it is not always possible to identify students needing intervention through NAPLAN data. Thus, it seems clear that NAPLAN data are not ideal for schools to use to screen for underachieving individuals.

Another way that secondary schools can potentially determine which students are underachieving upon enrolment, is to look at the students' A to E assessment and reporting history from primary school in the provided transition documentation. In primary school, particularly the first 3 years, the foundational skills of reading, writing and mathematics are expected to be taught and assessed. This means that early assessments should, for example, indicate whether student underachievement in mathematics is due to a lack of mastery of basic mathematics facts, or if poor reading can be explained by weaknesses in phonemic awareness. However, using the curriculum to assess students' progress in mastering foundational skills in this way requires that students be taught following a clear scope and sequence, from simple to complex content, to identify exactly where the weaknesses lie.

Despite recent improvements to the Australian Curriculum, this clear scope and sequence is not currently embedded in that document (ACARA 2022), nor in the National Literacy and Numeracy Progressions (ACARA 2020a; ACARA 2020b). Schools and teachers can assess against their own teaching progression, but not all schools have a scope and sequence in place for reading, writing and mathematics. Moreover, once a student transitions to upper primary, basic reading skills from the early years are an expected prerequisite. Foundational reading skills, such as phonemic and phonological awareness, are no longer assessed. Instead, students are assessed on their application of reading skills.

Likewise, foundational mathematics skills, such as counting and number sense, are also no longer assessed. This means that while sustained poor attainment in the upper primary years for a student may indicate to secondary schools that students need additional support and intervention, it will not indicate which skills or knowledge are weak in ways that can inform decisions about appropriate support or intervention. Thus, early primary years reports are not a reliable means of screening students to identify skill and knowledge deficits in the primary or in the secondary years. In our review, we therefore considered the research evidence that might provide indications of the best way for secondary schools to screen students to identify those needing intervention, and which skill deficits need to be addressed.

Frameworks for the delivery of targeted educational services and support

Categorical service delivery models in Australia and the US

In Australia, the approach used to identify students needing intervention has long been implemented through a categorical framework (de Bruin et al. 2020). Under this approach, students are assessed for their eligibility to access funding for additional educational services that are offered to address disadvantage associated with categories of personal characteristics or circumstances. These frameworks have long been defined and embedded across all jurisdictions within Australia's education systems. Under these frameworks, students must be identified as belonging to a particular priority equity cohort, such as Aboriginal and Torres Strait Islander students, or holding a verified diagnosis within one of the state-approved categories of disability, before accessing funding and support (for example, see Department of Education and Training Victoria 2022).

The frameworks of resource allocation and service delivery in Australia have been associated with a range of equity concerns and issues. There are many students with disability who need support but do not fit into an approved category, or do not meet eligibility requirements for targeted funding and support: for example, students with language and attentional difficulties,[9] or those with specific learning difficulties or disorders.[10] In several reviews, it has been found that this ineligibility for funding was often confused with ineligibility for support services, meaning that many went without (Commonwealth of Australia 2016; Poed et al. 2020).

Even where resource allocation moves beyond a categorical framework, the assessment of need and the funding of support may not be perfectly aligned. One example is Victoria's 'Equity (Catch Up)' initiative (Department of Education and Training Victoria 2022). Under this initiative, funding is allocated to secondary schools based on the size of the student cohort who fail to meet the National Minimum Standard (NMS) on the reading assessment within NAPLAN at Year 5. While well-intentioned in allocating resources to secondary schools to support underachieving students, this results in a delay as the funds do not flow to support the students until 2 or more years after the identified underachievement. Moreover, students who meet the reading but not numeracy standards do not attract additional funding. These are all examples of delays and restrictions that cause inefficiencies in delivering support and intervention.

Internationally, categorical frameworks were widely used from the 1970s, such as in the US. During the 1980s, concerns were raised in the US about the inefficiency and ineffectiveness of using a categorical model to deliver services. One concern was the delays in service delivery (as described above), as students had to 'wait to fail' before they would be referred for an assessment of their eligibility to access support (Ysseldyke and Marston 1999). These delays resulted from a requirement that eligibility for support could only be determined once students' academic achievement was substantially below that of their same-age peers, and a referral for diagnostic assessment made. This assessment generally entailed diagnosing a disability that was considered to satisfactorily explain low achievement and would trigger the release of funds to access services. For those students who did not have an intellectual disability, or any other impairment that was considered to explain underachievement, the difference between their intelligence quotient (IQ) and their achievement in reading or mathematics was used to diagnose them with a learning disability (LD) and provide access to services under that category. This became known as the IQ–achievement discrepancy approach for diagnosing LD. The delay in waiting for a pattern of low achievement to be established before a diagnosis could be made and eligibility for accessing services received, resulted in a widening of learning and achievement gaps.

9 For more information, see here: https://research.qut.edu.au/accessibleassessment/2022/03/26/what-are-we-learning-from-students-with-language-and-or-attentional-difficulties/

10 Examples include difficulties in reading, writing, spelling or mathematics that tend to be persistent and lifelong. For more information, see here: https://psychology.org.au/for-the-public/psychology-topics/understanding-specific-learning-disorders

There were other concerns raised that were associated with categorical frameworks. Predominantly, these concerns related to how special education service provisions were enacted for identified students. One concern was the presumption, within categorical frameworks, that low achievement is biologically determined or attributable to personal characteristics, without considering whether the quality of teaching could be an explanatory factor in student underachievement (Ysseldyke and Marston 1999). Special education thus took the form of remediation of student deficits and was not necessarily connected to the curriculum. Yet a further concern stemmed from the commonly used pull-out format of the service delivery and its relative ineffectiveness. Typically, and for secondary school students in particular, this pull-out service delivery placed students into special education settings (Chard 2013). Here, students were frequently taught by teachers or support staff without content knowledge in the general curriculum and were often offered an alternative curriculum that focused on basic skills or on functional 'life' skills (Zigmond and Sansone 1986). The content of this instruction typically lacked any correspondence with the general education curriculum and resulted in a loss of time spent accessing the academic curriculum and lowered the likelihood of successful graduation (Tindal et al. 1987).

Researchers have found that when a more consultative model for the identification of underachievers was used, students' achievements improved (Tindal et al. 1987), although the consultative model was less often implemented in schools (Zigmond and Sansone 1986). Accordingly, a number of recommendations were made for improving and reforming the delivery of support services to scale up the use of the consultation model. Recommendations included increasing collaboration and ensuring that educational support services complemented, rather than supplanted, instruction in the general curriculum (Tindal et al. 1987). An additional important recommendation was to use students' responsiveness to basic instruction in the curriculum as the basis for making decisions whether additional support and intervention was required; this was termed a problem-solving model.

During the 1980s, pre-referral intervention team models (PIT) were developed to reduce the number of children referred for special education placements and services (Burns and Symington 2002). The PIT models were a precursor to what later became known as response to intervention (RTI), which was endorsed as a preferred approach to identifying students with LD by the President's Commission on Excellence in Special Education. Appointed by President Bush in 2001 in the US, this Commission offered 9 key recommendations for reforming special education in its final report (Yell and Drasgow 2007). A particular focus was on reforming the process of determining eligibility for special education and reducing the number of students referred for special education who were behind on learning and did not have a disability. These recommendations led to profound changes in the delivery of educational supports and interventions, and ultimately led to the development of multi-tiered frameworks as a replacement for the categorical approach.

Multi-tiered frameworks in the US

Response to intervention

Multi-tiered frameworks were introduced in the US when response to intervention was written into the *Individuals with Disabilities Education Improvement Act of 2004* (IDEA[11]). This framework built on the previous decades of research outlined above and leveraged the findings from the studies that were used to address the recommendations of the Commission. The law enacted changes to how states could identify students with learning disabilities, following immense dissatisfaction with the IQ–achievement discrepancy model described previously that had been used in the overidentification of children for special educational services (Preston et al. 2016). The new changes in IDEA placed consideration on students' responsiveness to instruction and intervention at the heart of any determination of student underachievement, prior to referrals for assessment or

11 The *Individuals with Disabilities Education Improvement Act* (IDEA) is a federal law that guarantees all eligible children with disabilities between the ages of 3 and 21 (or until the child graduates) the right to a free appropriate public education designed to meet their individual needs. It legislated that RTI could be used as an alternative approach for identifying students with LD, meaning that states had the choice to move away from using the IQ discrepancy as a diagnostic measure for LD, and first determine whether students had received effective instruction.

special education placement. The changes to the law involving the use of RTI also permitted funds to be released to provide early intervention for students without a disability diagnosis, representing a clear shift away from the categorical framework (National Joint Committee on Learning Disabilities [NJCLD] 2005). This represented a radical reform of special education, with special educational services becoming supports that could be provided to students in general education schools and settings. The approach contrasts with how special education is understood in Australia, where special education continues to be offered for students with disability in entirely segregated classrooms and schools.

RTI was originally conceived as a continuum of educational services for students experiencing academic difficulties. There were several core components of RTI specified at that time. The components included a coordinated system of support across a sliding scale of tiers of support that increased in intensity, with evidence-based instruction and intervention practices implemented with fidelity at each level of support (generally known as Tiers 1, 2 and 3). Foundational to this framework was (and remains) the importance of Tier 1, also known as the universal tier, which operates to prevent learning gaps for the majority of students by providing the highest quality foundational instruction in the curriculum, designed for delivery to a diverse student cohort. The components also included universal screening and progress monitoring for skills acquisition to inform data-based decision-making about which students may need support. Additionally, a collaborative approach in using these data was adopted to design intervention to underachieving students, to implement intervention and to closely monitor progress for students receiving Tier 2 and 3 supports (Berkeley et al. 2009; Chard 2013; NJCLD 2005).

Following the reauthorisation of IDEA, increased funding was provided for research into RTI. Proper training of administrators and educators in the contents of the federal special education assessment guidelines and in using evidence-based interventions and frameworks such as RTI were recommended (Barrio et al. 2015). To support the system-wide scaling of RTI implementation, the US Department of Education Office of Special Education Programs (OSEP) developed a federal website (https://ies.ed.gov/ncee/WWC) as a knowledge clearinghouse and funded the creation of technical assistance centres[12] to build capacity for change within the system (Fixsen et al. 2009). These federally-funded technical assistance centres have expanded considerably in scope and number since that time (OSEP n.d.), with most states promoting state-run support centres[13] in addition to those run federally.

In the early literature, 2 broad approaches for RTI were delineated: the problem-solving model and the standard protocol. The problem-solving model involves the implementation of individually designed evidence-based interventions. For each student, a team of stakeholders (teachers, administrators, parents and school psychologists) define the problem, plan and implement an intervention, and evaluate the student's progress. The standard protocol model involves the implementation of specific standardised, evidence-based interventions for groups of students with similar difficulties (for example, reading fluency deficits) for a defined period (Berkeley et al. 2009).

Multi-tiered systems of support

Over time, the way that multi-tiered frameworks are conceptualised changed and matured to become much more comprehensive. RTI was traditionally articulated as a framework exclusively devoted to academic instruction and intervention and designed to address issues in the diagnosis of LD and lower special education placements. In addition to these aims, the problem-solving and standard protocol approaches emphasised the 'intervention' component, thus focusing much attention on Tiers 2 and 3 with less attention to the universal tier. Together, these foci meant that RTI remained, early on, somewhat rooted in a special education paradigm with less attention to the implications of what goes on in general education classrooms. Moreover, RTI as articulated within policy and legislation did not explicitly account for the parallel research efforts taking place at the same time, in developing multi-tiered frameworks focused on behaviour, known as Positive Behaviour Intervention and Supports (PBIS; Sailor 2015).

12 See here: https://osepideasthatwork.org/find-center-or-grant/find-a-center
13 See for example, New York state's RTI technical assistance centre: www.interventioncentral.org/new-york-state-rti-technical-assistance-center

More recently, the conceptualisation of multi-tiered frameworks has changed again, shifting to a comprehensive framework known as multi-tiered system of supports (MTSS; Pullen et al. 2018). This comprehensive MTSS framework has brought an increased focus on Tier 1 as the foundational intervention that maximises learning for all students. It has also shifted away from a narrow focus on academic instruction or LD diagnosis and now encompasses support for students academically, behaviourally and socially, as well as through a culturally appropriate[14] and/or trauma-informed approach[15]. Within this comprehensive MTSS approach, student academic achievement, behaviour, engagement and social development are inextricably linked.[16]

Multi-tiered frameworks have now been embedded within US federal legislation and state education policies for almost 2 decades. Thus, a substantial body of research exists on implementation and related outcomes for systems, schools and students. This research has, however, largely documented the evidence based on the component parts of RTI and MTSS, and there is little research evaluating the impact of the entire framework. Notably, one large-scale evaluation of RTI in relation to reading for primary school students was conducted (Balu et al. 2015) and this found no improvement in children's reading outcomes and that there were even negative impacts. However, as noted by Gersten et al. (2017), this trial focused only on the improvement for children who fell just below the cut scores used by schools as the eligibility cut-off for receiving Tier 2 services and used the comparator of children who fell just above those cut scores. Therefore, this trial lacked the design features to function as an evaluation of whether reading intervention works for children, ultimately examining only whether the eligibility cut score was appropriate in the study sample schools. There is clearly a need for robust, large-scale evaluations of multi-tiered models such as RTI and MTSS. While their components are based in evidence and documented as effective, the implementation of the framework in its entirety has not been demonstrated in robust research trials.

Another limitation in the evidence base for multi-tiered models is that existing research has predominantly focused on the first 3 years of primary school, with relatively less attention to older underachieving students (Williams et al. 2018). There are important differences between secondary schools and primary schools: for example, their scheduling and structure. As well, secondary teachers' preparation is focused on content-area expertise and they typically receive little preparation in instruction for basic skills. Another important difference is that underachieving students arriving in secondary schools typically have a range of well-established patterns of underachievement. This means that the emphasis on prevention of basic skill deficits through Tier 1, and close monitoring of their progress through the curriculum prior to considering whether intervention is needed, are less pertinent.

For students who have not mastered basic skills by the time they arrive in secondary school, the emphasis should move from prevention to timely intervention. Although these students still require high-quality Tier 1 instruction in the skills being taught within the secondary school curriculum, the research presented in our report focuses on how secondary schools should provide intervention to address these basic skill deficits at the higher tiers of support, rather than at Tier 1. We note, though, that high-quality Tier 1 instruction in the secondary school curriculum remains paramount for these students.

Tiers of instruction and intervention

As a coordinated system of support, RTI and MTSS are generally conceptualised as 3-tiered frameworks, although 4-tiered versions, in which the middle tier is split into 2, are also used. For the purposes of this report, we will be referring to 3-tiered models. Within Tier 1, core instruction is provided universally to all students within general education settings. This should entail using an evidence-based curriculum (for example, based in the 5 elements of reading) and evidence-based practices for instruction and assessment (Hughes and Dexter 2011). Central to the conventional understanding of RTI and MTSS is that Tier 1 is designed to be proactive and preventative. In order to minimise the risk of underachievement, the highest-quality teaching should be universally provided to

14 See for example here: https://www.mtss4els.org/
15 For more information, see here: https://mtss4success.org/special-topics/trauma-informed-care
16 For more information, see here: https://mtss4success.org/

a diverse and heterogeneous student cohort within Tier 1. In a typical school, high-quality Tier 1 instruction should be sufficient for about 80% of students to make adequate progress. Within Tier 1, it has been recommended that students are screened regularly (typically 3 times per year) to monitor their progress. Simple indicators of academic competence on general outcome measures, such as curriculum-based assessment,[17] should be adopted (Burns 2012). The data gathered will enable educators and data teams to identify those students who are not meeting expectations. If students are receiving quality instruction at Tier 1, yet their data suggest that they are underachieving, teaching is typically adjusted to meet their needs and they are monitored more closely, typically weekly or biweekly for about 8 to 10 weeks.

It is important to note that there are very few 'average' schools, with most schools having higher or lower proportions of students experiencing disadvantage, such as poverty. The figure of 80% is an ideal target, but it is not absolute. In schools with a highly advantaged or disadvantaged student cohort, this figure may well be higher or lower, even if Tier 1 is optimal. When the proportion of students needing supplemental support at higher tiers clearly exceeds 20% of the population, this has the potential to exhaust school resources or leave some students on wait lists for intervention for too long. Thus, when a substantially higher proportion of students are underachieving, Tier 1 should be adjusted to build targeted support into the general education classroom, for example by increasing the duration and frequency of explicit teaching of the foundational skills that is required by so many students in that setting.

A student has traditionally been non-responsive to Tier 1 instruction if their achievement is consistently ranked below the 20th percentile on a norm-referenced assessment[18] such as a curriculum-based measure across repeated assessment time points (Hughes and Dexter 2011). The student is then deemed at risk of substantial underachievement. For students who are non-responsive to Tier 1 instruction, school teams engage in collaborative decision-making to determine the most appropriate support to offer at Tier 2. This should be based on the Tier 1 screening data.

However, as noted above, this approach to screening is less applicable for secondary school students underachieving in reading, writing and mathematics. Basic skills screening has been identified as a key gap in the research for RTI and MTSS for secondary school students. One recommendation is that for older students, this non-responsiveness could be established through general measures of reading, writing and mathematics (Espin et al. 2018) without the need to do extensive screening across multiple time points. Screening measures within RTI and MTSS are typically proximal[19] measures of what is being taught in the curriculum and conducted close in time to the point where instruction has occurred. On the other hand, general measures are distal measures of whether students can apply such skills later on within that curriculum domain. For example, universal screening of reading in primary schools would utilise proximal curriculum-based measures of students' phonemic awareness, decoding and fluency. A distal measure would be a norm-referenced assessment of students' accumulated knowledge and skill, and their capacity to generalise what they have learnt and apply it, such as a reading comprehension assessment that would show whether students meet the expectations for their age and year level. Examples of standardised assessments include NAPLAN reading assessments and Progressive Achievement Test for reading comprehension (PAT; Australian Council for Educational Research [ACER] 2022).

Every effort should be made to minimise the impact of any support and intervention offered at a higher tier on students' access to Tier 1 instruction in the academic curriculum and within the general education classroom. The scheduling and staffing of how this should be achieved, and how support at higher tiers should be offered within secondary schools, have been noted as research gaps (for example, Shinn et al. 2016).

17 Curriculum-based measures are simple assessment methods that are used to monitor students' progress in learning basic skills such as reading, spelling, mathematics and written expression. They are well-researched with hundreds of studies documenting their reliability and validity. There are many published curriculum-based measurement tools available with documented reliability and validity.

18 Norm-referenced assessments evaluate and grade the learning of students by judging and ranking them against the performance of their peers. A norm-referenced test is a type of standardised test (such as, a test that is identical for every test-taker). After the items on a norm-referenced test are scored, the scores are compared to those of a comparison group, or norming group.

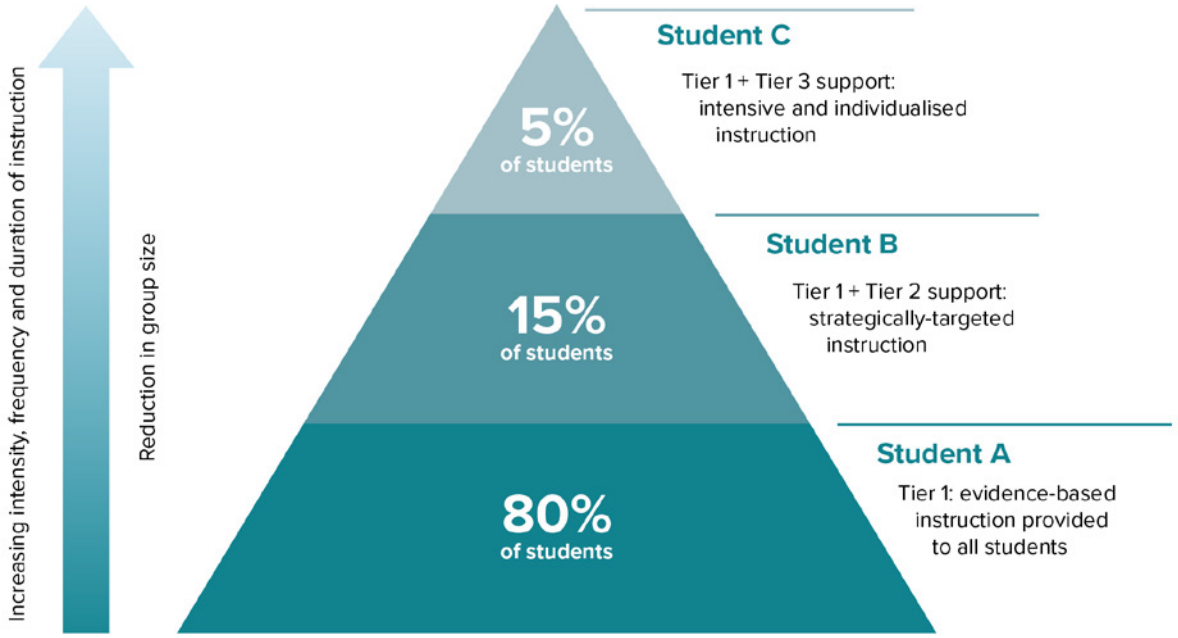
19 Proximal assessments are those that are closely aligned to the knowledge and skills taught to students and offered close in time to when active teaching of that content occurred.

While traditionally underachieving students would have been placed into separate special education classrooms or schools, as described above, instruction designed to address their underachievement and facilitate access to the general education curriculum are generally not provided in these settings. Poorer outcomes on rates of progress and school completion are frequently found. It is important that students are offered interventions that are designed using a problem-solving model to determine the specific skills they need to learn, and that instruction on these skills is offered through evidence-based practices, rather than offering the students a different curriculum in a different location. This is in keeping with a core purpose of the original advent of RTI: to reduce referrals for special education placements. A key focus of our research is to identify which practices are indeed evidence-based for this cohort within the setting of regular schools.

Students referred for Tier 2 support are expected to be provided with more intensive, research-based interventions with close monitoring of progress, in addition to the core instruction received by all students. Tier 2 instruction commonly involves targeted, small-group intervention with ongoing monitoring of progress (Barrio et al. 2015; Berkeley et al. 2009). This intervention is considered sufficient for closing gaps for about 15% of the student population if implemented with fidelity. Tier 2 is time-limited, has clear goals and entry and exit criteria that indicate when students will no longer need support. Monitoring student progress under the intervention is necessary and the intervention provision is adjusted as required. For students who fail to succeed or make sufficient progress, additional work should be undertaken to ensure that Tier 2 intervention appropriately targets skill deficits that underpin poor progress, and that these deficits have been correctly identified and addressed with the most suitable evidence-based practices.

Within RTI and MTSS, instruction across the tiers should be aligned so that Tier 2 supplements and complements Tier 1, but does not replace it (Harn et al. 2011). Rather, support at higher tiers is layered on top of Tier 1, which must be high-quality, efficient and evidence-based instruction designed for a diverse cohort of students. Where quality Tier 1 is in place, Tier 2 should supplement this with more frequent or intensive instruction but not provide multiple or competing instructional approaches; this should be avoided for vulnerable students (Chard 2013). Rather, the logic of RTI and MTSS is that evidence-based instruction at higher tiers should be an *intensified* version of Tier 1 practice, achieved by increasing the frequency and duration of instruction and reducing the group size (Harlacher et al. 2010; Lemons et al. 2014; Powell and Stecker 2014). That is, students access a higher 'dosage' of quality instruction. This is represented graphically in Figure 2.

Figure 2 How tiers of support work in a multi-tiered system of support



Source: de Bruin and Stocker (2021)

For example, in the teaching of reading, the 5 components of phonemic awareness, decoding, fluency, vocabulary and comprehension strategies should be explicitly taught with frequent review, opportunities to rehearse and respond, and with corrective feedback (Hempenstall 2016). Students who do not make sufficient progress in relation to any of these reading components should have increased access to these same high-quality instructional practices through intensification of instruction at Tier 2. This process of intensifying instruction through frequency, duration and group size provides students with greater access to the teaching and engagement in active learning time, thus increasing opportunities to learn, respond and review skills (Van Camp et al. 2020). Again, given that there is no longer any instruction in basic skills for secondary school students who struggle in reading, writing and mathematics, careful consideration needs to be given to how this alignment is achieved. We note, however, that these students should continue to access Tier 1 instruction for the secondary school curriculum within the general education classroom to develop their knowledge and skills through evidence-based practices such as building their disciplinary literacy and background knowledge.[20]

Only if sufficient endeavours to succeed at Tier 2 have been made should students be provided with support at Tier 3. At Tier 3, additional intensification of instruction should be provided through increasing frequency and/or duration, often for a prolonged period as well as lowering group size, usually to individualise instruction further. The goal of multi-tiered frameworks is that students who receive intervention at higher tiers are ultimately returned to the regular classroom to receive Tier 1 instruction only. Researchers examining the application of RTI and MTSS for secondary school students have highlighted that the distinction between Tiers 2 and 3 for these students is less clear (Williams et al. 2018).

Screening and progress monitoring

A key component of multi-tiered frameworks such as MTSS, and of the implementation of targeted (Tier 2) or intensive (Tier 3) interventions, is using data-driven decisions to inform instruction and intervention. Within a multi-tiered framework, these decisions typically draw on triannual data gathered from the universal screening of Tier 1 students, as well as additional data from student progress monitoring when they are receiving support at higher tiers. As noted above, at Tier 1, screening data are typically collected through curriculum-based measures that are robust (that is, statistically or methodologically defensible) and brief (taking one to 5 minutes to administer) (Espin et al. 2019). Typically, schools screen students' growth by comparing them to peers and a *cut score;* for example, scoring at or below the 20th percentile. Students can be classified as being at-risk/not at-risk using cut scores. Cut scores can be percentile ranks within a sample student cohort and year level, or they can be growth lines over time. For example, Year 1 students scoring below the 25th percentile in phonological awareness or decoding or receiving lower growth points over a series of time points would receive an intervention in which high-quality instruction is intensified in that aspect of reading. The same students may also be classified as at-risk through references to normed cut scores.[21]

20 For a fuller discussion of this we recommend Shinn et al. (2016).
21 See, for example, the cut scores classified under the Dynamic Indicators of Basic Literacy (DIBELS; University of Oregon n.d.).

A range of commercially produced curriculum-based measures has been assessed for reliability, validity and predictive value. These measures are available on the National Center for Intensive Intervention's academic screening tools chart (National Center for Intensive Intervention n.d.). The accuracy of these measures is reported on this site regarding whether they are robust for use as a check that students are achieving as expected. In primary schools, once 80% of the target population is making adequate progress against expected benchmarks, and the quality of Tier 1 teaching cannot justify the underachievement of the remaining 20% of students, underachieving students are assessed further to understand their skill deficits that need addressing through intensified and targeted support at Tier 2. School-based teams undertake this data collection and analysis to determine the targeted interventions required for the students who are making less progress than expected and who are performing below benchmarks. Progress monitoring data are collected regularly for students receiving Tier 2 or Tier 3 support so that intervention type, delivery or intensity can be adjusted accordingly.

For secondary school students who are underachieving in reading, writing or mathematics, triannual Tier 1 screening is not appropriate, as these students' underachievement can already be established through their history of low attainment. It is not necessary to continue to document their non-responsiveness to Tier 1. Instead, intervention should be offered immediately. Yet, as discussed earlier, schools need an effective decision-making process for this based on sound data. While their low attainment is likely to have been established through A-E reporting and national/state assessments, these data do not constitute a robust means of determining eligibility for Tier 2, nor can they precisely pinpoint the specific skills that require intervention (for example, number sense in mathematics). Williams et al. (2018) indicate that alternative brief measures may be required to precisely identify specific academic difficulties (for example, decoding or fluency). A key focus in our research was to determine research-based recommendations for what some alternative measures might be.

Multi-tiered frameworks in Australia

In the US, the shift away from categorical approaches happened well over 2 decades ago. Yet in Australia, changes to service delivery frameworks are only now gradually taking place across the federation of states and territories. Just as change in the US followed reforms to federal legislation, these changes in Australia have been enacted in response to the federal needs-based funding model. Specifically, the changes support reforms within the *Australian Education Act 2013*. A base per-student funding allocation, known as the schooling resource standard (SRS), is used to calculate additional school funding *loadings* for priority equity cohorts identified within the NSRA (Council of Australian Governments 2018). The changes have also been enacted in a number of state-based funding reforms, such as South Australia's Inclusive Education Support Program policy and Victoria's Disability Inclusion policy. Most notably, across these federal and state reforms, there is an emphasis on service delivery being funded and rolled out based on need rather than category of disadvantage (Australian National Audit Office 2021). The reforms also share a tiered approach to funding and support that is conceptualised as existing across multiple tiers representing increasingly intensive and individualised supports. A central focus of the research review that we present in this report, therefore, is on the evidence for using multi-tiered frameworks from the US, and how this can be implemented within the context of the shifts towards tiered approaches to funding underachieving high-school students in Australia.

Methodology

Aims and objectives

This review offers 3 substantive contributions to the field of research in this area. The first relates to service delivery frameworks as we provide a systematic overview of the current state of the evidence around how these are defined and supported at the system level. The second relates to the content and practices that should be used with individual secondary school students for reading, writing and mathematics interventions at Tier 2 and Tier 3. The third relates to how these should be implemented with fidelity to optimise student outcomes. We conducted a preliminary search for systematic reviews or protocols through ERIC, Google Scholar and the Campbell Collaboration. No planned, ongoing, or published systematic reviews on the topic were identified.

The objective of this project was to address this gap and systematically review what systematic reviews reveal about:

- how schools should deliver services to secondary students needing support in literacy and mathematics
- a robust and efficient means of identifying secondary school students who are underachieving in literacy and mathematics in a timely fashion
- the most effective intervention foci and practices for schools to implement to address specific areas of underachievement.

The aim of our umbrella review was to synthesise systematic reviews of literacy and mathematics interventions for underachieving Years 7 to 9 students. Specifically, our objective was to report on the recommended service-delivery framework for interventions, how underachieving secondary school students should be identified, what interventions are appropriate for use with these students, and how interventions should be implemented to achieve optimal outcomes. We also sought to identify any gaps in knowledge relating to these objectives. These objectives were addressed by seeking answers to the following research questions that were provided by AERO.

Research question 1

What are the contemporary supporting frameworks or models that guide the delivery of intensive supports to struggling students (for example, a response to intervention or multi-tiered system of supports model)? Have these frameworks been tested or evaluated?

Research question 2

What are the recommended approaches for identifying students who require intensive support? For example, when and how frequently should students be assessed and reassessed?

Research question 3

What are the recommended approaches for supporting struggling students to improve their literacy and numeracy skills and reduce the gap between them and their peers? What level of intervention (such as duration, intensity and frequency) do these approaches require?

Research question 4

What are the most effective ways to implement these approaches, and what preconditions are necessary to support their implementation?

- Who should be implementing these approaches? For example, primary trained teachers, classroom assistants, speech therapists.
- What resources and supports do schools need to implement the recommended approaches with fidelity to the duration, intensity and frequency that is required? For example, time structures, professional development opportunities and age-appropriate materials.

Research question 5

Are there major gaps in the evidence base that prevent these questions being fully addressed? What research could be undertaken to fill these gaps?

Research methods

An umbrella review of systematic reviews was selected as the most appropriate method by which rigorous and robust research could be systematically examined to answer the research questions above. An umbrella review is a form of aggregative review in which the findings from multiple systematic reviews are compiled. It is often referred to as a 'review of reviews'. In educational research, systematic reviews have become increasingly common to address various questions regarding the effectiveness of instruction and intervention practices. This is because, as secondary research, they are helpful in determining a comprehensive synthesis of primary empirical studies, and the effectiveness of interventions can be better determined. When education systems develop guidelines for practice, it is important that they are based on the findings of systematic reviews so that the guidelines are aligned with the relevant body of research evidence.

Umbrella reviews are conducted using systematic processes to examine systematic reviews. The aim is to synthesise findings at a *tertiary* research level. They are appropriate for selection as a methodology particularly when there is a reasonable volume of existing systematic reviews for consideration (Papatheodorou 2019; Wang and Papageorghiou 2021). The most characteristic features of an umbrella review are the inclusion and synthesis of the highest level of evidence, namely other systematic reviews. As a result, the umbrella review creates confidence that the synthesis sits at the peak of evidence-based hierarchies (Wang and Papageorghiou 2021). Umbrella reviews, therefore, offer the most robust method for determining evidence-based practice and enable policy decision-makers to gain a broad understanding of a field by using the highest level of evidence in response to the research questions posed (Aromataris et al. 2015). The systematic processes involved in an umbrella review include using an a priori protocol detailing clearly formulated questions, inclusion/exclusion criteria, a search strategy, the critical appraisal of the systematic reviews to be included, and a data extraction process (Aromataris et al. 2015).

We deemed an umbrella review to be warranted as our preliminary search of the field of research highlighted a large volume of potentially relevant systematic reviews, particularly in reading. At AERO's request, we undertook an umbrella review as a rapid review (a streamlined systematic review), rather than a more comprehensive approach which takes considerably longer. When conducting rapid reviews, it is important to maintain methodological rigour and transparency so that the quality and trustworthiness of the evidence is maintained, while seeking to complete the review in a short timeframe (Tricco et al. 2022). While the research questions were provided by AERO, we selected the protocols of JBI Global (Aromataris and Munn 2020) to inform our inclusion/exclusion criteria, a search strategy, critical appraisal, and a data extraction process.

In accordance with advice from JBI Global (Aromataris and Munn 2020), our umbrella review was methodologically inclusive and not limited solely to synthesising systematic reviews of randomised controlled trials (RCTs). RCTs are widely viewed as the gold standard for high-quality evaluations of effectiveness. Yet, while research using these designs is increasing in education, they remain relatively rare. Moreover, quasi-experimental studies in which the impact of interventions for specific populations of students is examined are relatively common, particularly in relation to students who are underachieving in reading, writing and mathematics.

In these fields, there is an extensive body of research in which the focus is on intervention outcomes for students, variously termed as 'struggling', having 'learning difficulties' or as 'learning disabled'.[22] Educational research of this kind is often conducted using quasi-experimental designs that are considered to lack some of the quality features of RCTs, such as the random assignment of groups. In line with recommendations from JBI Global, we selected to take a pragmatic and methodologically inclusive approach for our umbrella review and include systematic reviews of both experimental and quasi-experimental research. We also note that for Research question 1, in addition to the quantitative systematic reviews we identified, we also included 2 systematic reviews of a type termed, by JBI Global, as 'systematic reviews of text and opinion'[23] (Aromataris and Munn 2020).

Included in our umbrella review are systematic reviews in which a range of research designs were used. These enabled an examination of the totality of the evidence, and determination of whether insights from different research designs and across different populations are consistent (Aromataris et al. 2020). Using the JBI protocols, we applied the concepts of population, phenomena of interest and context (PICo) to amass a body of research literature sharing these elements (see Stage 3 in the Research stages section for more detail).

Reporting on effectiveness

Several of the research questions required us to report on the effectiveness of interventions examined across the domains of reading, writing and mathematics. However, there was variation across the systematic reviews with how intervention effects were reported. For example, in some studies an associated meta-analysis was conducted, yet these calculated and reported a variety of effect sizes. The authors of systematic reviews with meta-analysis reported various effect sizes such as Cohen's *d*, Hedges' *g,* odds ratios (OR), Tau *U* statistics and standard mean difference (SMD); a small number used different statistics. In some of the systematic reviews included in our umbrella review there was no associated meta-analysis conducted; for example when this was precluded by having too few primary experimental or quasi-experimental studies in the systematic review sample. Typically, when a meta-analysis was not conducted, the authors of the systematic reviews provided a lengthy overview of the findings by reporting the effect sizes included in each of the primary studies reviewed.

Our solution to synthesising across this diversity of effect sizes and mix of systematic reviews with and without meta-analyses was to faithfully report results as presented in the systematic reviews, to provide a clear understanding of the outcomes. This process mitigates the risk of distorting or misinterpreting the evidence presented in the systematic reviews. We integrated findings following Aromataris et al. (2020). Accordingly, in our umbrella review, measure of impact was determined by describing the results through prose and citing the effect sizes reported. This process enabled us to compare, contrast and integrate the data sets to develop a comprehensive analysis of the phenomena of interest.

22 These terms are not used consistently across global contexts and their use and interpretation is subject to external influences such as funding policies governing access to educational services. For an overview, see the support materials on the Australian Government's professional learning site for the Nationally Consistent Collection of Data on School Students with Disability, *Learning differences, learning difficulties, learning disabilities and the NCCD.* https://www.nccd.edu.au/wider-support-materials/learning-differences-learning-difficulties-learning-disabilities-and-nccd

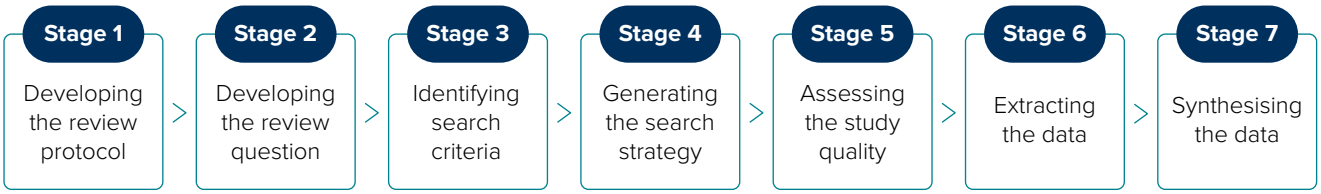
23 JBI defines systematic reviews of text and opinion as those which draw on evidence that does not itself arise from scientific research. Within systematic reviews of text and opinion, the data source might be reports, monographs or even expert opinion. JBI notes that such research has a role to play in determining evidence-based practice by complementing empirical evidence. They recommend that it should not be discounted even though it is not of the same standing in terms of quality of evidence as empirical evidence from research (Aromataris and Munn 2020).

It is important to note that we acknowledge the wide acceptance in the research community that effect sizes are used and should be reported. Effect sizes are powerful statistical measures. However, it is known that different methodological approaches can produce variations in reported effect sizes. The most common standardised effect size measures used included Cohen's *d* and Hedges' *g*, for which conventional interpretations, as defined by Cohen (1969), are the following: 0.2 is considered a small effect size; 0.5 is considered a medium effect size; and 0.8 is considered a large effect size. Recent discourse on the translation of specific effect sizes to benchmark or rank interventions has raised concerns about interpreting effect sizes as a measure of effectiveness through Cohen's definition of small, medium and large (for example, Kraft 2020; Simpson 2019). We acknowledge the concerns raised by these researchers and emphasise that effect sizes are not absolute. It is important to remember that effect sizes are indicators of how an intervention worked under controlled conditions for a certain cohort and context, at a particular moment in time, as measured by a specific set of outcome assessments and methodological choices. It is inappropriate to assume that adopting an intervention reported to have a large effect size is necessarily more likely to translate into the desired educational outcome for every group of students. Moreover, a small effect size can still be educationally meaningful. Implementing an alternative intervention strategy with a smaller effect size may result in a valuable educational outcome for students.

Research stages

The research was conducted using the 7 stages of a systematic review as illustrated in the diagram. Details of the 7 stages are described below.

Figure 3 The 7 stages of a systematic review



Stage 1: developing the review protocol

The review protocol for this study was submitted to AERO as part of the request for proposal process.

Stage 2: developing the review question

AERO initially scoped this systematic review as part of its Tiered interventions project plan. The research questions were developed by AERO and were presented earlier in the report.

Stage 3: identifying the search criteria

We applied 4 essential criteria in searching and screening systematic reviews for inclusion in our review. The first criterion was methodological. The remaining 3 criteria were related to the content and focus of the systematic reviews, determined by their alignment with the research questions and defined by key elements of PICo. Our inclusion criteria involving methodology and PICo were as follows:

- **Methodology:** we included systematic reviews if they constituted a quantitative systematic review with or without meta-analysis, or a systematic review of text and opinion (policy). We excluded primary empirical systematic reviews, qualitative reviews of teacher or student perspectives, scoping reviews and umbrella reviews. As search criteria, we defined the key features of systematic reviews as reviews involving a comprehensive, systematic search strategy across multiple databases, and systematic processes for removing possible bias in the extraction of data.

- **Population:** we included systematic reviews relevant to secondary school students, with a particular emphasis on systematic reviews addressing students in the early years of secondary school (Years 7 to 9) who were underachieving in literacy or mathematics. We included systematic reviews if the sample population was below Years 7 to 9. However, we excluded systematic reviews if fewer than 30% of the sample were within Years 7 to 9. We included systematic reviews if the sample included students with learning difficulties or disabilities, and if the intervention foci or the instructional practices were applied to broader student populations and were also connected to curricular content. We excluded reviews in which the focus was on life skills or instructional design, and if the sample population was students with a specific disability that fell outside of the scope specified by AERO.
- **Phenomena of interest:** our review had a broad focus determined by the research questions and was guided by criteria stipulated by AERO. Accordingly, our criteria for inclusion in the phenomena of interest involved literacy (defined as reading or writing) and numeracy (defined as mathematics). Within the phenomena of interest, reviews focusing on student engagement were not included. The phenomena of interest in our umbrella review were:
  - multi-tiered frameworks such as RTI or MTSS that were used to deliver academic support and intervention for students with literacy and mathematics challenges
  - targeted supplemental (Tier 2) or intensive interventions (Tier 3) that were used to address academic skills in literacy and mathematics
  - elements of intervention implementation, such as assessment data (screening or progress monitoring), intervention agents, intensity (frequency or duration of intervention), resourcing and materials.
- **Context:** we included reviews if they were directly relevant to secondary school settings. Reviews in which the focus was on Years P to 12 or middle schools (defined as Years 5 to 8 or Years 6 to 8) were included, provided that 30% or more of the students were within our target population range of the equivalent of Years 7 to 9. We excluded reviews if the focus was solely on the primary school years. Systematic reviews in which the settings for interventions did not relate to general education (for example, in clinical settings or universities) were also excluded.

Stage 4: generating the search strategy

Based on an initial scoping of the field, the lead author of this review designed the literature search strategy in consultation with an expert educational librarian. First, using Google Scholar and ERIC, a gold set of 9 systematic reviews was identified. The gold set comprised 3 systematic reviews across each of the 3 phenomena of interest described above. Key search terms were identified from the gold set and were used to test the search strategy for retrieving relevant references. Additionally, the gold set was tested to determine how relevant keywords had been indexed within academic databases and to design the search strategy.

Next, a systematic search of the literature was conducted to identify relevant systematic reviews in 5 databases: A+, Scopus, ERIC, ScienceDirect, and Taylor and Francis. These databases were selected as they constitute the key scholarly databases in the field of education, and widely encompass national and international research. Google Scholar and the Monash Library Multi-Database Search were also employed to maximise search results and to include relevant grey literature. The combination of searches provided a comprehensive search strategy and minimised the risk of publication bias. No date limits were placed on the search to allow all relevant literature to be located and considered for inclusion.

The literature search focused on 3 main phenomena of interest: multi-tiered frameworks, literacy and mathematics interventions, and intervention implementation. Accordingly, we began by conducting separate systematic searches using keywords for each phenomenon of interest and appropriate synonyms, abbreviations and phrase combinations. The keywords included ‘multi-tiered systems of support’ OR MTSS OR ‘response to intervention’ OR RTI OR ‘intervention’ and were used in conjunction with ‘literacy’ AND ‘read*’ AND ‘writ*’ for the literacy searches,

and 'math*' OR 'numeracy' for the mathematics searches. The search terms 'systematic review' OR 'meta-analysis' OR 'research synthesis' OR 'review' were used in searching for systematic reviews using the relevant terms related to the research methodologies that would meet our inclusion criteria.

Using the features of the different databases, we then added limiters such as 'literature review', 'systematic review', 'research review', 'meta-analysis' to the systematic search results for each set of searches (literacy and mathematics) to ensure that we were locating those relevant to the methodology. The search also incorporated terms relevant to the population and the contexts of interest (as described above). The search terms that were used included 'student' OR 'adolescent' OR 'high school' OR 'middle school' OR 'secondary school'. Phrases relevant to the third phenomenon of interest were also included, for example, 'assessment' OR 'data-based decision*' OR 'universal screen*' OR 'progress monitor*'. Ancestral searches of the reference lists of included systematic reviews were also conducted to locate further relevant reviews. Reference lists from seminal research studies were scanned to ensure that the maximum number of potential systematic reviews was identified.

After removing duplicates and irrelevant systematic reviews, the lead researcher imported the references of the systematic reviews into Endnote and into the JBI online systematic reviewing platform, SUMARI (JBI Global n.d.). Each research team member was then provided with a training session for the title and abstract screening processes by the lead researcher. The title and abstract screening processes were modelled and team members then independently coded 3 systematic reviews to check for consistency and confirm that the training had been successful.

All records retrieved for the title and abstract screening were double coded by team members with 81% agreement. The discussion resolved any conflicts or discrepancies. The conflicts in initial decisions to include/exclude systematic reviews were primarily related to those reviews in which the proportion of students representing the population of interest was unclear; data mining was required to confirm whether 30% or more of the total numbers of students whose data were analysed in the review were in the age category of interest. Following this, the lead researcher ran a subsequent training session in full-text screening to model the process. All team members then coded 5 systematic reviews to check for consistency and to confirm that the training had been successful, and we reached 100% agreement for coding. Finally, 2 team members screened the full text of all retrieved systematic reviews.

Stage 5: assessing the study quality

Two standardised tools were used to maximise accuracy in critical appraisal.

Study quality of systematic reviews of evidence were critically appraised using the JBI Critical Appraisal instrument for umbrella reviews (Appendix A.1). The lead researcher ran training sessions on using the instrument and modelled the process for critical appraisal. Ten systematic reviews were appraised in triplicate to check for consistency in decision-making; 95% agreement was achieved, and all differences were resolved via discussion. Following this, the remaining systematic reviews were coded individually by team members. In total, 35% of the systematic reviews were coded in duplicate. Each item in the coding sheet was scored 2 points for 'yes', one point for 'unclear' and zero points for 'no'. Systematic reviews had to receive a 'pass' score of 11/20 to be considered of adequate quality.

Two systematic reviews were identified to be systematic reviews of policy documents. The JBI Critical Appraisal instrument for Text and Opinion was employed to critically appraise these 2 systematic reviews (see Appendix A.2). Each item in the coding sheet scored 2 points for 'yes', one point for 'unclear' and zero points for 'no'. A systematic review had to receive a 'pass' score of 3/6 to be considered of adequate quality and acceptable for inclusion in the umbrella review.

The critical appraisal outcomes of all studies we examined are available in Appendices A.6 and A.7.

Stage 6: extracting the data

A standardised tool was used to minimise bias in data extraction of the main findings across the included systematic reviews. Data extraction was completed using the JBI Data Extraction instrument for umbrella reviews (see Appendix A.3). The lead researcher again ran training sessions in using the instrument and modelled the process for extracting data. Twelve systematic reviews were appraised in triplicate to check for consistency in decision-making (achieving 95% agreement). Following this, team members individually extracted data from the remaining systematic reviews. A random selection of 35% of the systematic reviews was extracted in duplicate to check for ongoing consistency in extraction.

Stage 7: synthesising the data

The data extracted in Stage 6 were exported from SUMARI into a Google Sheet and classified according to a set of characteristics. For example, the types of primary studies that were included in the systematic reviews, the type of review methodology, the PICo dimensions as well as the main findings were extracted and recorded (see Appendix A.8). The data were also synthesised using NVivo Software (QSR 2022) using assigned classifications. The classifications mirrored the phenomena and outcomes of interest extracted in Stage 6, and the research questions identified in Stage 2. The NVivo classifications were as follows:

- service delivery framework
- assessment
- reading
- writing
- mathematics
- moderating factors
- implementation fidelity
- research gaps.

Next, a set of a priori codes and sub-codes (adult and child codes) was specified within NVivo based on the research questions (see Appendix A.4). The coding was replicated from the data extraction in SUMARI. The text coded for each adult and child code was then summarised according to its relevance to the research questions in our review. The summaries are presented in the results section that follows.

Results

Study inclusion

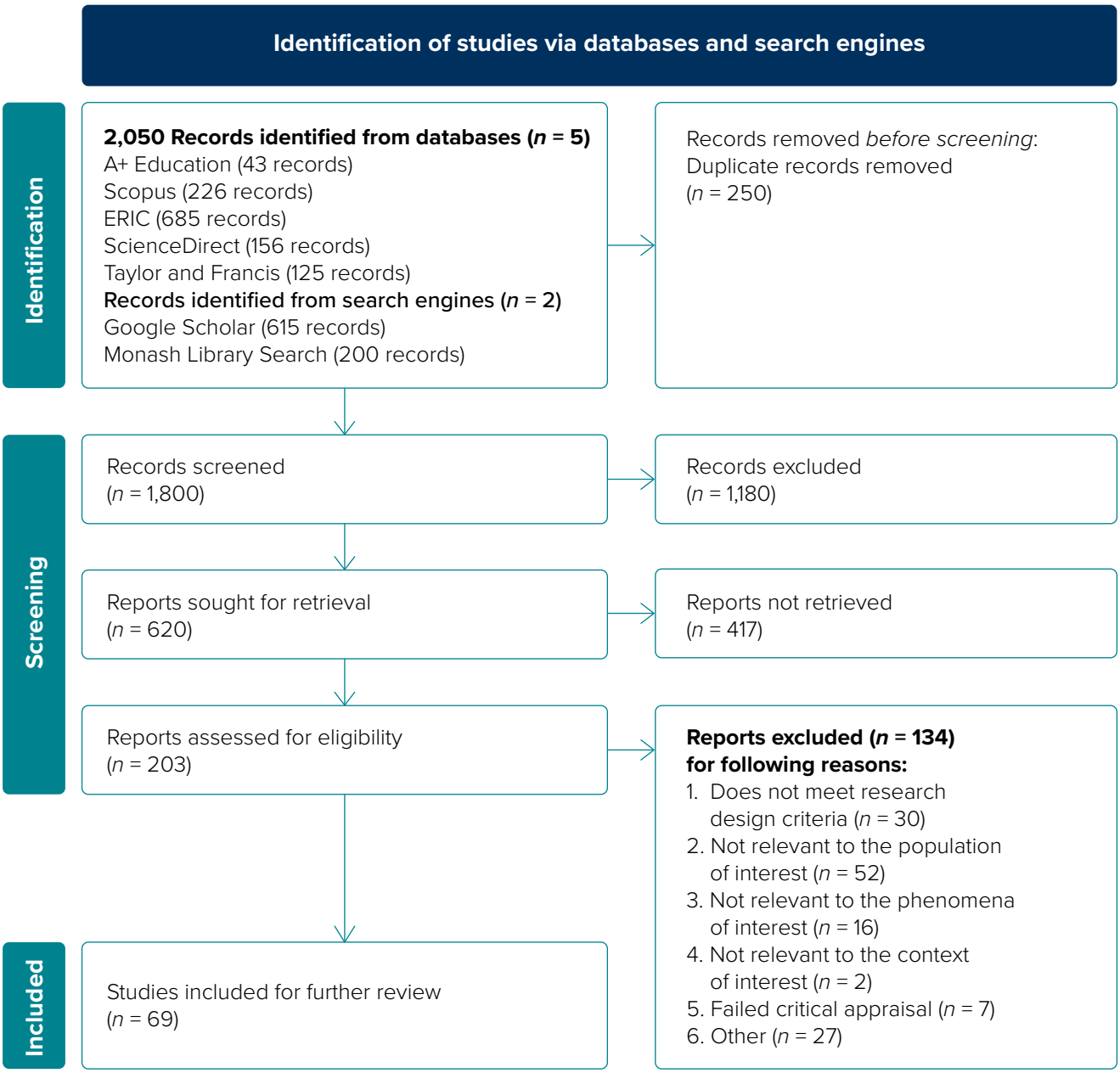
A total of 2,050 records were identified using the search strategy outlined above. The lead author extracted the search results and conducted an initial screen to remove those irrelevant to education. For example, the search term 'RTI' produced several systematic reviews where this abbreviation was from the health field and referred to 'respiratory tract infection' instead of 'response to intervention'. A total of 250 duplicate records were removed through Endnote and 620 reviews were screened at the title and abstract stage. A further 417 reviews were eliminated from the review for misalignment with the PICo, in addition to the methodological criteria for inclusion in our umbrella review. The remaining 203 articles were screened at the full-text stage.

A total of 138 systematic reviews were excluded on methodological grounds, due to misalignment with the PICo, or for failing to demonstrate adequate standards under critical appraisal. Eight systematic reviews did not demonstrate adequate research quality when we critically appraised them and were excluded. There were 69 systematic reviews included in the final sample.

Overview of the quality of the systematic reviews

Of those systematic reviews included in the final sample, 5 notable areas of weakness were identified within their own critical appraisal processes. One flawed component of their critical appraisal processes was the degree to which systematic searches were conducted comprehensively and exhaustively across multiple databases and were inclusive of grey literature. There were also common flaws in the systematic review critical appraisal process outlined in the systematic reviews that we appraised, with this step being missed, incomplete, inadequately described or using processes that did not adequately eliminate bias. The full results from the critical appraisal process are found in Appendices A.6 and A.7.

Figure 4 Preferred reporting items for systematic reviews and meta-analyses (PRISMA) diagram



(Page et al. 2021).

Characteristics of included systematic reviews

Descriptive statistics

We began our analysis by examining descriptive statistics to understand the characteristics of the sample. The descriptive statistics of the 69 systematic reviews that met our inclusion criteria are presented in Table 1.

A total of 69 systematic reviews spanning 23 years (1999 to 2022) were included in the final sample, with most (83%) published in the last 10 years. The 69 systematic reviews in the umbrella review captured over 60 years of research (1960 to 2021), totalling 1,984 studies. Most of the systematic reviews were published in peer-reviewed journals and included a total of 87,246 reported participants.

The methods adopted in the 69 systematic reviews were primarily systematic reviews with a meta-analysis (SRMA; 59%) and included systematic reviews with group design research (SRMA-G; 22%), with single case and group research (SRMA-M; 22%) and systematic review with a meta-analysis for single case studies (SRMA-SC; 10%). We coded reviews that did not fit into any of the research designs (as groups/single/mixed) as 'other' (SRMA-O). These comprised 6% of the total included systematic reviews. The remaining systematic reviews were conducted without meta-analyses (SR; 41%) and included studies with single and mixed designs (SR-M; 17%), group designs (SR-G; 10%), single case designs (SR-SC; 6%), and descriptive reviews that were neither single, mixed, nor group in design (SR-O; 9%).

Year and school levels

A total of 57 systematic reviews (81%) reported year levels while a smaller percentage reported only school levels (*n* = 12; 17%).

The majority of systematic reviews (81%) included student populations of Years 7 and above. Of the secondary school year levels reported, 53 systematic reviews considered populations in Years 7 to 9 (*n* = 54; 76%) and notwithstanding overlapping of populations across the reviews, students were represented in the cohorts of Year 7 (*n* = 57; 81%), Year 8 (*n* = 55; 79%) and Year 9 (*n* = 47; 67%).

Table 1 Descriptive statistics on the 69 included systematic reviews

Descriptive statistics	Number	Percentage
Systematic reviews conducted between 1999 and 2011	12	17%
Systematic reviews conducted between 2012 and 2022	57	83%
Span of studies 1960–2022	1,984	
Peer review journals	69	99%
Total reported effect sizes	1,395	
Total number of participants	87,246	
Systematic reviews without meta-analysis		
Systematic review without meta-analysis group design (SR-G)	4	6%
Systematic review without meta-analysis mixed group and single case research (SR-M)	12	17%
Systematic review without meta-analysis – other (SR-O)	8	12%
Systematic review without meta-analysis single case research (SR-SC)	4	6%
Total number of systematic reviews without meta-analysis	28	

Descriptive statistics	Number	Percentage
Systematic reviews with meta-analysis		
Meta-analysis group design (SRMA-G)	15	22%
Meta-analysis mixed group and single case research (SRMA-M)	15	22%
Meta-analysis other (SRMA-O)	4	6%
Meta-analysis single case research (SRMA-SC)	7	10%
Total number of systematic reviews with meta-analysis	41	
Literacy	52	75%
Reading	38	55%
Writing	14	20%
Mathematics	14	20%
Assessment	7	10%
Service delivery	5	7%
Implementation	4	6%
Reported school levels only	13	18%
Schoolwide (elementary and high school)	5	
Middle school	1	
Secondary/high school	7	
Reported year levels	57	81%
Kindergarten, Year 1 to Year 12	16	23%
Years 7 to 12	57	81%
Kindergarten to Year 6	54	77%
Years 7 to 9	53	76%
Year K	16	23%
Year 1	23	33%
Year 2	27	39%
Year 3	30	42%
Year 4	37	53%
Year 5	40	57%
Year 6	54	77%
Year 7	57	81%
Year 8	55	79%
Year 9	47	67%
Year 10	43	61%
Year 11	38	5%
Year 12	35	50%

Note: An overlap of data will affect the total percentage count. These included more than one classification loaded in systematic reviews and overlaps of years and school levels.

Research methodology

Most systematic reviews included in our umbrella review were conducted as systematic reviews with meta-analyses (n = 41; 59%). The remainder were systematic reviews that were conducted without meta-analyses (n = 28; 41%), which also included systematic reviews of text and opinion (policy documentation).

Systematic reviews that applied statistical methods to summarise their results indicated rigorous processes, more precise data estimates and objective evaluations of research findings.

The 41 systematic reviews with meta-analysis (SRMA) reported a sample of 67,171 students from 1,400 studies (*M* = 1,638 students, *M* = 34 studies) with a total of 1,275 effect sizes. The systematic reviews in which no meta-analyses were conducted had a total sample of 22,075 students from 584 studies (*M* = 717 students, *M* = 21 studies). Only 3 systematic reviews with no meta-analyses reported effect sizes (Berkeley et al. 2009; Oxley and de Cat 2021; Reed and Cummings 2014). In addition, several systematic reviews investigated and reported on more than one domain, thus overlapping classification exists. For example, Jung et al. (2018) classifications included mathematics, assessment and implementation.

A breakdown of the systematic reviews relative to the methods and phenomena of interest that were used to classify the studies for data extraction and coding are shown in Table 2.

Table 2 Breakdown of systematic review methodologies by phenomenon of interest

Classification	Systematic reviews with meta-analysis (SRMA), N = 41 (%)	Systematic reviews without meta-analysis (SR), N = 28 (%)
Assessment	4 (10%)	3 (11%)
Implementation	2 (5%)	2 (7%)
Service delivery model	2 (5%)	3 (11%)
Mathematics	11 (27%)	3 (11%)
Literacy	32 (7%)	20 (71%)
Writing	9 (22%)	5 (18%)
Reading	23 (56%)	15 (54%)

Note: Some systematic reviews were assigned more than one classification where they had more than one phenomenon of interest.

Review results

The results of the umbrella review are provided in this section of the report. The findings related to each of the 5 research questions addressed in the umbrella review are presented in turn.

Research question 1

What are the contemporary supporting frameworks or models that guide the delivery of intensive supports to struggling students (for example, response to intervention or multi-tiered system of supports model)? Have these frameworks been tested or evaluated?

Five systematic reviews were identified as relevant to answering this research question, representing a very small proportion (7%) of the total sample of systematic reviews. Two of these were qualitative systematic reviews of State Department of Education policy and guidance documentation. These reviews addressed the first part of this research question, examining contemporary supporting frameworks that guide the delivery of intensive supports to underachieving students (Barrio et al. 2015; Berkeley et al. 2009; Berkeley et al. 2020).

Another was a qualitative systematic review of the literature relating to the preparation of teachers for implementing multi-tiered frameworks (Barrio et al. 2015). The remaining 2 systematic reviews were quantitative systematic reviews of RTI and addressed the second component of this question regarding whether multi-tiered frameworks have been evaluated (Burns et al. 2005; Burns and Symington 2002). The impact of the implementation of RTI on schools and students was evaluated in these systematic reviews.

None of the systematic reviews focused exclusively, or disaggregated findings, in relation to students in the secondary school years. We outline the findings and results extracted from these reviews below. Each of the subcomponents of the research question is addressed separately.

What are the contemporary supporting frameworks or models that guide the delivery of intensive supports to struggling students?

As noted in the introduction of this report, RTI was introduced into US federal legislation in response to issues identified with the implementation and outcomes of the categorical framework of service delivery. Importantly, within this legislation the states had considerable autonomy as to whether they moved away from the IQ–achievement discrepancy model and adopted RTI for determining whether a student should be labelled as having an LD. The 2 'snapshots' from Berkeley et al. (2009; 2020) provide insights into the degree to which RTI was taken up in policy, how the states enacted this framework for delivering intensive supports to underachieving students, and how these 2 phenomena changed over time by shifting towards a more comprehensive MTSS framework.

Early multi-tiered models in the US

The IDEA legislative reforms went into effect in the US in 2006. In the following years, Berkeley et al. (2009) conducted a systematic review of how the State Education Departments had responded to federal mandates and communicated information to guide districts and schools. The researchers reported that following the reforms within federal legislation, RTI was the only officially named multi-tiered framework in place within the policies of 15 states. The majority of these 15 states blended the standard protocol and the problem-solving model, with a small number adhering to only one of these 2 approaches. The states taking the blended approach developed their own modified framework for implementation. Most states incorporated behaviour within RTI, although 4 states had reading-only RTI models.

In addition to official policy at the state level, there was a range of other ways in which RTI was embraced at the state and district level, such as the provision of guidance documentation. Most of the 51 states acknowledged the potential benefits of adopting RTI through such documentation. However, there was notable variability in the development of official policies in the 15 states that *had officially* adopted RTI, as well as in the implementation guidelines and resources that were also offered in other states that had not. Regardless of whether states had officially adopted RTI, there were efforts to provide resources to facilitate the uptake and implementation of RTI. Some states formed relationships with universities, such as the IRIS Center at Vanderbilt University and the University of Florida, and the universities provided online training. Within the education regulations of most states, it was mentioned that the use of RTI to replace the IQ–achievement discrepancy model was permitted; it was mandated in only 2 states.

In their analysis, Berkeley et al. (2009) noted variability in how RTI was conceptualised within state policy and guidance documentation. These sources of variability included where responsibility lay for implementing support at the higher tiers, how the tiers functioned, and a lack of clarity regarding when the special education process began within RTI frameworks. For example, in some states responsibility for Tier 2 rested with personnel such as classroom or special education teachers or allied health specialists, yet in other states, any member of staff with training could oversee Tier 2.

Additionally, in states specifying that a problem-solving approach to RTI was to be used, Tier 2 intervention responses were developed based on individual students' needs. Yet in the states in which the standard protocol was used, the approach was to group students by need and provide intervention by drawing from a predetermined list. Further examples of variation also existed in how Tier 3 was understood. In some states, small-group interventions were provided at Tier 3; in other states, only individualised interventions were provided. In some other states, both formats were combined (Berkeley et al. 2009). All states conceptualised special education as a process that was separate from RTI and occurring after RTI interventions had been exhausted. In some states, special education was initiated after Tier 2; in others it was initiated after Tier 3; in other states, it was permitted to initiate special education at any stage. A small number of states had 4-tiered models in which special education was considered the fourth tier. There was also variation in how the tiers were conceptualised through intensification. In some states, it was specified that intensification was increased through frequency, duration and group size, but this varied across jurisdictions. The researchers noted that some states were 'early adopters', having embraced RTI prior to legislation; Kansas was a notable example, having been involved in early research that informed the reforms.

Berkeley et al. (2009) noted a significant gap in policy and guidance documentation. One identified gap was that although the importance of using evidence-based practices at each tier was emphasised across states, there was little clarity of what the practices were, or how evidence-based practices should be interpreted and enacted. Another significant gap was that the omission of this information was particularly problematic at the secondary school level. A third gap was that although Tier 1 was emphasised in the policy and guidance information across states, the information was generally only located on the states' special education websites; in practical terms, this meant that there was less emphasis on a high-quality Tier 1 for all students.

Contemporary multi-tiered models in the US

In a second snapshot, Berkeley et al. (2020) found that there were significant changes across the states during the second decade following the legislative reforms of 2004. Perhaps the most significant change was that most US states (n = 39) adopted a multi-tiered model, with only eight states relying on guidance documentation alone; 4 states did not have clear information on their websites making it impossible to determine their policy. Most states with multi-tiered frameworks had also adopted a blend of the standard protocol and the problem-solving approach, instead of adhering to a single approach.

In the 2 decades, there was a shift away from using RTI as the descriptor for state-based frameworks. MTSS frameworks (n = 21) outnumbered RTI frameworks (n = 17), with an additional 5 states using the terms RTI and MTSS interchangeably. Berkeley et al. (2020) identified some common themes across the state policy reforms in the shift away from RTI towards MTSS. For example, while early implementation of multi-tiered frameworks that were examined in the first snapshot (Berkeley et al. 2009) focused largely on academic instruction and intervention, most states (93.3%) had also incorporated behavioural intervention in their MTSS or RTI frameworks.

What was relatively common was that the shift to MTSS was achieved by integrating the Positive Behavioural Interventions and Supports framework with the RTI framework. Several states had also defined broader approaches for how RTI or MTSS were to be used to determine whether a student could be identified as LD. These newer and broader approaches articulated a contemporary approach for determining whether a child had specific learning difficulty or disability in basic reading skills, reading comprehension, reading fluency, mathematics calculation, mathematics problem-solving, or written expression (Berkeley et al. 2020).

However, there was still considerable variability in terminology and definitions. There was considerable disparity in the terminology used for RTI and MTSS across states, with the terms used in ways that were often quite inconsistent. Some states specified that the terms were distinct, while others specified that they were interchangeable. Some specified that MTSS was a subset of RTI, while others specified that RTI was a subset of MTSS. The lack of consistency in definitions and practices was noted as potentially limiting the scalability of RTI and MTSS (Berkeley et al. 2020). There was also considerable disparity between states regarding how MTSS, RTI

and special education fitted together. The documentation and policy in Kansas clearly distinguished the MTSS model from special education, noting that Tier 3 did not constitute special education and that the MTSS process should not delay any diagnostic evaluation when it would be needed for students to be accommodated under US discrimination law (the equivalent of reasonable adjustments under Australian legislation). In contrast, in New Mexico, students were identified in Tier 3 as being in special or gifted education programs (Berkeley et al. 2020).

Some states were cited as being exemplary in their creation and dissemination of MTSS documentation and supporting materials, with the state of Kansas singled out for mention as it had also been in the Berkeley et al. (2009) snapshot. Just as Kansas had been noted as an early adopter of RTI in the first snapshot, Kansas was interestingly the first state to move toward MTSS as a single framework system with the availability of multiple supports for all students. Kansas was also used as a benchmark of statewide approaches in professional development materials and its material was often cited in the materials of other states (Berkeley et al. 2020). For example, the MTSS manual in Kansas carefully provided critical components of the model, such as detailed descriptions of the MTSS model, approaches to identify, analyse, implement and evaluate intervention, and the role of leadership at the district and school levels.

An essential component of contemporary multi-tiered frameworks, such as RTI and MTSS, is the knowledge and preparedness of education stakeholders on the ground. Berkeley et al. (2020) emphasised that this is best undertaken by using the process of implementation science, and they drew heavily on the work of Fixsen et al. (2005) in their discussion. Implementation science involves the identification of a series of stages for implementing reforms such as RTI, including: (a) exploration and adoption, (b) program installation, (c) initial implementation, (d) full operation, (e) innovation and (f) sustainability. Following the introduction of RTI, the Federal Office of Special Education Programs developed task-oriented technical assistance centres to sustain the implementation of MTSS at a national level (OSEP n.d.). In the US, the uptake and scaled implementation of RTI and MTSS have been supported and facilitated by these technical assistance centres.

In several of the systematic reviews included in our umbrella review, the guidance and the materials provided by the federal government and state education authorities were noted as crucial in supporting the uptake of RTI. Also mentioned was implementation with consistency across the system, including through professional learning such as that provided by the university partnerships (Berkeley et al. 2009; 2020). The technical assistance centres have also facilitated cross-state learning opportunities and the creation of state resources on RTI implementation across schools in state systems (Berkeley et al. 2020).

While the reviews discussed above highlighted the rich professional learning and support materials available to educators and schools through US state and local education authorities, and through technical assistance centres, the systematic review by Barrio et al. (2015) highlighted that there has been relatively less attention paid to upskilling pre-service teachers and preparing them for the implementation of multi-tiered interventions. In particular, the authors noted that pre-service preparation was targeted more towards special educators. The inference is that while general educators are responsible for quality Tier 1 instruction, the bedrock of the multi-tiered model, they are not being placed at the heart of teacher preparation for RTI or MTSS. Barrio et al. (2015) recommended that teacher preparation be improved in this respect, and provided particular recommendations for incorporating evidence-based practices and progress monitoring into the higher education curriculum for pre-service teacher education. They also recommended that pre-service teachers, who will be the initial service providers within multi-tiered frameworks, be afforded placements within schools in which RTI and MTSS are used, thus giving them practical experiences (for example, in using evidence-based practices and progress monitoring) to enhance their preparation for implementation of multi-tiered support.

Have these frameworks been tested or evaluated?

The impact of multi-tiered frameworks was examined in 2 meta-analyses (Burns and Symington 2002; Burns et al. 2005). Both reported that multi-tiered frameworks and approaches implemented in schools led to demonstrably improved outcomes for students, as well as the schools themselves.

Burns et al. (2002) conducted a meta-analysis of 9 articles, and 57 effect sizes examining the impact of multidisciplinary problem-solving teams, known as pre-referral teams (PITs), on students and school systems. PITs were school-based problem-solving teams who developed interventions for underachieving students. Student outcomes of interest were time on task, task completion, scores on behaviour rating scales, and observations of target behaviour. Systemic outcomes of interest were the number of referrals for special education assessment, new placements in special education, percentage of student referrals who are diagnosed with a disability, number of students retained in a grade (repeating a year of school) and increases in consultative or counselling activity by school psychologists. Burns et al. (2002) noted that engaging in pre-referral intervention and action prior to referring students for educational assessment or special educational placement/funding was beneficial for both students and systems. They found strong and significant impacts on students (d = 1.15) and school systems (d = 0.9) that, on average, were substantially higher for studies involving random assignment within research designs (d = 1.43) and conducted through university trials (d = 1.32). Additional positive outcomes supported the cost-effectiveness of pre-referral teams (Burns and Symington 2002).

Burns et al. (2005) meta-analysed 21 studies and extracted 25 effect sizes. They reported outcomes arising from schools implementing RTI in the years just prior to IDEA. In the meta-analysis, they examined the impact of RTI on outcomes for students, such as assessments of academic skills, estimates of growth in particular skills, as well as observations of the time spent on task and rates of task completion; all outcomes were related to academic interventions. Burns et al. (2005) also examined the impact of RTI on schools, including outcomes such as the percentage of students referred to and/or placed in special education, student time spent receiving special education services and the number of students retained in (repeating) a grade. Cohen's d effect sizes were calculated and, when possible, converted to an unbiased estimate of effect (UEE) to weight and standardise the effect sizes and make them slightly more conservative than d values. It was found that 'early adopter' RTI schools demonstrated substantial and positive impacts from RTI for both students and the schools themselves.

Strong and significant effects for both schools (UEE = 1.54) and students (UEE = 1.02) were reported. Burns et al. (2005) found that the number of students non-responsive to Tier 1 instruction was 19.8%, on average. An average of 5.98% of students were referred for targeted support and only 2% referred for special education placements. In the systematic reviews in which field-based RTI frameworks were examined, less than 2% of the student population was identified as LD, a statistically lower result than previous estimates of 5.7% (Burns et al. 2005).

Research question 2

What are the recommended approaches for identifying students who require intensive support?

Only a small set of systematic reviews (n = 7) examined assessments for identifying and monitoring students needing support, constituting a small proportion (10%) of the total sample included in the umbrella review. Three were quantitative systematic reviews without meta-analysis (Ardoin et al. 2013; Newell et al. 2020; Reed and Cummings 2014) and 4 were systematic reviews with meta-analyses (Filderman et al. 2022; Jung et al. 2018; Kilgus et al. 2014; Shin and McMaster 2019). Participant numbers were not reported in all the systematic reviews; however, where they were, a total of 13,103 participants were represented. Of the systematic reviews focused on the screening of academic skills for older students, only one considered progress monitoring (Ardoin et al. 2013).

In most of the systematic reviews in our umbrella review, screening through CBM[24] was examined. CBM is a standardised measurement system for indexing student proficiency in basic skills, including reading, spelling, mathematics and written expression (Fuchs and Fuchs 1991). CBM are quick and efficient and can be used often (for example, weekly) to monitor student skill growth. To be fit for this purpose, they need to be relatively reliable and valid (Clemens et al. 2016).

Oral fluency and curriculum-based measures for reading fluency (CBM-R)

Three systematic reviews focused on CBM assessments for reading fluency (CBM-R; Ardoin et al. 2013; Kilgus et al. 2014; Shin and McMaster 2019). CBM-R assessments are proximal assessments administered close in time to teaching activities and drawn directly from the curriculum itself. They are assessments of skills being taught but not of whether students can apply and generalise them. CBM-R assessments are commonly used to screen for underachievement risk in monolingual English speakers, and students in the younger years of school (for example, Years 1 to 6). They can be oral reading fluency (ORF) assessments, for which students are required to read a passage aloud for one to 3 minutes and scored for the number of words correctly read per minute. They can also be maze assessments, for which students read silently and identify missing words and are scored for their correct responses.

CBM-R can be used to screen for achievement by comparing students' scores to norms and they can be used to monitor individual students' progress in Tier 1 by plotting growth over time (Shin and McMaster 2019). Research has shown that oral reading performance is indicative of reading skills, influenced by the retention of previously learned skills (for example, phonemic awareness) and the acquisition of new skills (vocabulary, oral fluency and comprehension) (Kilgus et al. 2014). This makes it a relatively sound proxy for general reading achievement.

Using CBM-R assessments for screening

Shin et al. (2019) examined the criterion validity of the maze and the ORF measures by determining their correlation with standardised reading assessments, such as state reading assessments. Such correlations are typically used to determine how robust a screening instrument is in terms of its predictive value, and how capable the instrument is in identifying students at risk of failing to meet benchmarks with reasonable accuracy. Shin et al. (2019) found that while both the ORF and the maze assessments were correlated with state tests, the oral reading assessment was more highly correlated for older students ($r = 0.64$). By contrast, the maze correlation declined in correlation for these students (Shin and McMaster 2019). This challenged previous work which had hypothesised that the maze would perform better.

Determining cut scores for CBM-R assessments

Kilgus et al. (2014) examined CBM-R to understand the diagnostic accuracy of the ORF assessment in relation to cut scores or thresholds to make dichotomous decisions about students who are/are not at risk of underachievement. There are 2 approaches to determining such cut scores. One is to take a normative approach and rank students. A typical way that schools might apply such a normative approach is to determine how many students for whom they have the resources to provide intervention and use that number to determine the percentile rank of the cut score. Kilgus et al. (2014) gave the example that if a school only has the resources to provide intervention for 25% of their students, then the 25th percentile would become the cut score.

The alternative is to take a standards-based approach in terms of the benchmarks. This is a more objective and data-driven approach for determining cut scores. The rationale provided for the standards-based approach is that it determines cut scores with the sensitivity and specificity to identify 'true' rates of at-risk/not at-risk students, thus providing high predictive value for both of these dichotomous categories and hence diagnostic accuracy in predicting reading difficulties. Kilgus et al. (2014) found that CBM-R had sound diagnostic accuracy in differentiating between students with and without reading difficulties. However, they found that standardised

24 CBM is a standardised measurement system for indexing student proficiency in basic skills, including reading, spelling, math and written expression. It was intended to be used as a general outcome measure to reflect overall proficiency in an academic domain, rather than mastery of discrete skills (Fuchs and Fuchs 1991:7; Fuchs and Deno 1991).

cut scores were variable across time and year levels. Rather than a rigid, standards-based approach, Kilgus et al. (2014) recommended that local standardised cut scores should be developed based on diagnostic accuracy within the local context.

Using CBM-R for progress monitoring

While CBM-R assessments of ORF are well-established as reliable universal screeners for fluency, Ardoin et al. (2013) examined the evidentiary basis for published decision rules for using ORF to monitor progress within Tier 2 or 3 intervention. Specifically, this systematic review investigated the decision rules regarding whether a particular student's progress within intervention is discrepant in relation to a set goal for them within the intervention. Ardoin et al. (2013) found that while the ORF CBM-R are robust for screening and benchmarking where a student is compared to others, they lack a sufficient evidence base for progress monitoring where a student is compared against their own previous scores. They noted the need for a sound measure to be developed for this purpose.

CBM-R and English language learners

Newell et al. (2020) examined the utility of the CBM-R for predicting poor reading outcomes in English learners (ELs). The key findings of this review revealed that CBM-R do not perform consistently as a predictor of poor reading comprehension across languages and language competencies. Specifically, they were found to be variable according to the language proficiency of ELs in their first language and of the language background of ELs.

Although CBM-R were correlated with reading outcome measures, the diagnostic accuracy of CBM-R to identify ELs as at-risk of poor reading outcomes did not meet criteria. The authors concluded that when CBM-R are used, in accordance with established protocols and decision rules for native English speakers, to identify categories of risk, CBM-R were likely to over-identify ELs as being at risk. The authors recommended that a *gating* process be researched to understand its potential for 2 reasons. The first was to introduce flexibility into the screening process and the second was to ensure that the decision rules for determining risk of poor outcomes initially include consideration of students' first language, and their competency in that first language.

Using data-based instruction to intensify support

DBI, also known as data-based individualisation, is a process that can be used to intensify instruction within a multi-tiered system of support. The collection of data from students and its use by teachers were examined in 3 systematic reviews (Filderman et al. 2021; Jung et al. 2018; Reed and Cummings 2014).

Jung et al. (2018) conducted a meta-analysis on teachers' use of student data to intensify instruction for students needing intensive Tier 3 support. They examined the impact of DBI on the reading, mathematics and spelling/writing outcomes of students needing Tier 3 support, as well as the influence of the type and frequency of coaching and support provided to teachers to use the data effectively. The authors consistently found a small-to-medium effect size of $g = 0.37$ across all academic areas. However, the effect size was considerably higher ($g = 0.47$) when teachers were provided with more frequent consultation in using DBI.

Filderman et al. (2021) examined this element of training teachers to work with data in more detail. The focus of their systematic review was on the effectiveness of training in data literacy for teachers that specifically targeted their knowledge, skills and beliefs for using data to individualise instruction. These reviewers found that effective training had a strong effect on teachers' knowledge and skills outcomes ($g = 6.7$), with a substantial but lower impact on their beliefs ($g = 0.48$). Collaborative training of the teachers produced the highest impact on outcomes.

These findings are noteworthy, particularly considering the finding from Reed and Cummings (2014). They found that in research trials, the fidelity of implementing assessments was insufficiently reported for screening or pre-post measures. They also identified that the expertise of training of assessors was rarely reported, and that assessors rarely considered the impact of any potential threats to the quality of data they collected, for example, noisy environments during testing or scoring errors. This highlights the importance of being professionally trained in quality processes for collecting and using data.

Research question 3

What are the recommended approaches for supporting struggling students to improve their literacy and numeracy skills and reduce the gap between them and their peers? What level of intervention do these approaches require?

In this section of the report, we outline the findings from the systematic reviews that examined the impact of interventions provided in reading, writing and mathematics. The systematic reviews examining effective interventions for secondary school students in reading were the largest set in this umbrella review. Of the 69 systematic reviews, 52 (75%) examined literacy interventions, and of these, 38 (55%) focused on reading, with 14 (20%) systematic reviews focused on writing. A total of 14 reviews (20%) assessed numeracy interventions.

The results extracted from these sets of systematic reviews are detailed separately in the following sections. Each section is structured to provide a breakdown of the set of systematic reviews by the following subsections:

- description of systematic reviews, including methodology, participant numbers, and characteristics
- focus and outcomes of intervention in terms of the target skill for improvement
- moderators of outcomes (setting, duration and frequency)
- generalisation and maintenance of outcomes.

Reading

Description of systematic reviews

The systematic reviews focusing on reading constituted the most significant set in our umbrella review. There were 38 reviews identified, constituting over half (55%) of all the included systematic reviews. Of the 38 reviews on reading interventions, 15 systematic reviews did not include a meta-analysis (SR), while 23 systematic reviews included meta-analysis (SRMA). Intervention results in reading were reported for over 58,155 students in the 38 reviews. It should be noted that in several large reviews, individual participant numbers were not reported, meaning that the actual number of participants was substantially higher. See Appendix A.5 for more detail.

In the systematic reviews involving reading interventions, findings were reported on the effects on reading skills in general, or on specific aspects of reading: phonemic awareness, decoding, vocabulary, fluency and comprehension. The findings are outlined in the following sections. It should be noted that in some systematic reviews, an overall effect size for the reviewed studies was calculated; these are reported below as a single value. In other systematic reviews, only effect sizes for individual studies were reported. These effect sizes are presented below as a range of values only when the data could be reliably determined; readers are directed to the original articles for further information.

Several effective strategies for teaching reading skills were also identified. The strategies are collated in Appendix A.9.

Focus and outcomes of reading interventions on general reading skills

The impacts of reading interventions on general reading outcomes were reported in some reviews. Overall, there was evidence from several systematic reviews to show that interventions can, and do, positively impact older underachieving readers. In 2 meta-analyses (Scammacca et al. 2007; Scammacca et al. 2015), the estimated mean effect size across 82 effects from all studies included in the meta-analyses was found to be $g = 0.49$. This indicated a moderate, positive effect of the intervention of nearly half a standard deviation on students' reading outcomes. The mean effect for the earlier (1980–2004) group of studies meta-analysed (Scammacca et al. 2007) was considerably more significant at $g = 0.96$ than the mean effect for all studies, whereas the mean effect for the 2005–2011 set of studies meta-analysed (Scammacca et al. 2015) was smaller at $g = 0.23$. The difference in effect sizes was attributed to a difference over time in the typical measures used in the reviewed research studies. In recent systematic reviews, there is a trend towards using more standardised, distal measures.[25]

25 Distal measures typically referred to assessments examining whether students could generalise any skills learned; proximal measures, such as those developed by researchers, tended to evaluate whether specific skills were learned when taught directly.

School-based interventions (including instructional methods comprising small-group instruction, progress monitoring and peer-assisted instruction) had positive and statistically significant average effect sizes (g = 0.38, 0.19 and 0.19, respectively). In this systematic review, standardised tests were used to measure the effects of interventions for middle and high-school students with, or at risk of, academic difficulties (Dietrichson et al. 2020). In another systematic review, a smaller effect (g = 0.20) was found for small-group reading interventions for secondary students (Hall and Burns 2018). Similarly, DBI produced a positive and statistically significant small effect on reading (g = 0.28) for students with disabilities (Jung et al. 2018).

Other interventions were also found to have a positive impact on general reading outcomes for older students. Notably, these interventions were all multicomponent. Large effects on reading measures were found for students with LD following multicomponent interventions that included explicit training in cognitive strategies that followed the SRSD instructional sequence comprising self-questioning, comprehension strategy training, self-regulated comprehension and explicit writing frameworks (Kang et al. 2015). Similarly, for adolescents, multicomponent programs in which basic reading skills were taught (for example, MultiLit, Corrective Reading Decoding) yielded an overall medium-to-large effect across all reading achievement outcomes; the mean effect size was d = 0.83, the median effect size was d = 0.68 and the modal effect size was d = 0.55 (SD = .82; Joseph and Schisler 2009).

There were some differences in the impact from general reading interventions depending on the age or year level of students. This evidence was mixed. In some general reading systematic reviews, only limited effects (low effect sizes) were found for students in the upper elementary and middle school years (Flynn et al. 2012; Wanzek et al. 2013). Evidence from other meta-analyses indicated that there were no significant differences between mean effect sizes for studies that included only students in Years 4 and 5, Years 6 to 8 and Years 9 to 12 (Scammacca et al. 2015). Importantly, a close inspection of the systematic reviews on general reading interventions revealed that they targeted and measured different interventions. This means that it is important to consider the intervention itself when comparing or generalising findings.

Focus and outcomes of reading interventions on specific reading skills

Most of the systematic reviews of reading focused on the impact of interventions for distinct elements of reading, that is, phonemic awareness, decoding, vocabulary, fluency and comprehension. The results are summarised below, with subheadings for each of the 5 reading subskills defined in the introduction to the report:

- phonemic awareness
- decoding (phonics)
- fluency
- vocabulary
- comprehension.

It should be noted that comprehension studies comprised the largest set of systematic reviews; thus, the reading comprehension section below is significantly more extensive than the others. We also note that, unlike each of the other 4 reading subskills, comprehension is not itself a 'skill' but is the outcome of proficiency in the other 4 subskills as well as specific comprehension skills (for example, finding the main idea) and strategies (for example, summarisation strategies) that have been taught. The comprehension section is therefore broken up into subsections according to the comprehension strategies that were identified in the systematic reviews, that is, explicit instruction[26] and strategy instruction.

26 Explicit instruction is a structured and systematic approach to teaching academic skills. Archer and Hughes (2011:1) explain that it is "characterized by a series of supports or scaffolds, whereby students are guided through the learning process with clear statements about the purpose and rationale for learning the new skill, clear explanations and demonstrations of the instructional target and supported practice with feedback until independent mastery has been achieved". In addition, the authors emphasise proceeding in small steps, checking for understanding and achieving active and successful participation by all children.

Phonemic awareness

There was only one systematic review in which data pertaining to phonemic awareness were reported. This systematic review (Goodwin and Ahn 2013) examined the impact of morphological instruction on literacy outcomes. Morphological instruction generally involves 4 instructional components: (1) improving awareness of the morphological structure of words; (2) increasing knowledge of meanings of affixes and roots; (3) supporting morphological problem solving, and (4) developing hypotheses about meanings of unfamiliar words through morphemic analysis. A mean effect size of *d* = 0.48 was reported for phonological awareness outcomes following morphological instruction for underachieving students and typical achievers in Kindergarten to Year 9 (Goodwin and Ahn 2013).

Decoding

There were limited data pertaining to the effectiveness of reading interventions to improve the decoding skills of older students. Instruction on sight words, phonics and a combination of both was shown to improve word recognition acquisition skills for adolescent underachieving readers (Joseph and Schisler 2009). Morphological instruction[27] produced statistically significant mean effects (d = 0.59) on decoding measures for underachieving readers in middle and high school (Goodwin and Ahn 2013).

Vocabulary

A paucity of vocabulary instruction-based research for secondary-age students with LD was noted (Kuder 2017). However, some evidence to support the effectiveness of interventions targeting vocabulary skills was found. It was reported that rather than stand-alone interventions, vocabulary instruction was included in two-thirds of multiple-component interventions. It is possible that this was due to decisions made by researchers in response to the large effect sizes reported in previous meta-analyses (for example, Scammacca et al. 2015). School-based interventions for improving the reading of older students with, or at risk of, academic difficulties, yielded a small effect size for vocabulary outcomes (*g* = 0.137) that was, nonetheless, statistically significant (Dietrichson et al. 2020).

Results from meta-analyses (Scammacca et al. 2015; Scammacca et al. 2007) showed that vocabulary interventions had a significantly larger mean effect size (*g* = 1.58) than all other types of reading interventions for all outcome measures (Scammacca et al. 2015; Scammacca et al. 2007). There was also evidence to support the effectiveness of vocabulary instruction for specific populations including ELs (Oxley and de Cat 2021) and students with significant cognitive disabilities (Browder et al. 2006).

In one systematic review, it was found that 4 strategies had large effect sizes on vocabulary learning for secondary-age students with LD (Kuder 2017). These strategies were: mnemonic instruction (using strategies such as acronyms, rhymes or pictures to assist students with memory retrieval of taught content; *d* = 1.40 to 2.743); learning strategies that utilised morphemic analysis (involves students being taught to predict the meaning of words; *d* = 4.226 to 4.264); and direct instruction[28], a form of explicit instruction in which students were shown a target word on a flash card, a definition, a sentence that used the word and 2 probing questions that were designed to help students link the word to information in their textbook, (*ES* [effect size type was not specified] = 1.12).

For the fourth strategy, multimedia instruction that was delivered through podcasts, a large effect size (*d* = 1.09 to 1.97) was also reported. Similar effective interventions – direct instruction, supply of mnemonics, or semantic mapping[29] (such as, highlighting the relationship between the words and content) – were found to improve reading comprehension of science texts for middle and secondary school students with LD (Kaldenberg et al. 2015).

27 See section on phonemic awareness.

28 Direct teaching/instruction is a step-by-step, lesson-by-lesson approach to teaching which is teacher directed and typically aided by a script. Lessons are designed according to content and follow a pre-determined skill acquisition sequence using the 'I do, we do, you do' format with individualisation provided through differing entry points, reinforcement, error correction, the amount of practice and the amount or type of feedback (Hempenstall 2020).

29 The evidence for the effectiveness of semantic mapping was derived from only one study within one systematic review, so is not included in the Strategies Table.

The effect sizes in this systematic review were compared between the vocabulary studies in which vocabulary was explicitly taught (*g* = 1.25) and non-vocabulary studies, in which students were taught using a broad and generalisable instructional strategy (including using graphic organisers; *g* = 0.64). The comparison revealed that teaching students the meaning of vocabulary words found in science texts had a greater impact on the comprehension of those texts than other non-vocabulary interventions (Kaldenberg et al. 2015).

Morphological instruction[30] produced significant moderate (*d* = 0.34, Goodwin and Ahn 2013) to large effects (*d* = 4.226 to 4.264; Kuder 2017) on vocabulary. The findings suggested that instruction in analysing the morphological structure of words helped students to use the meaning of a morpheme to determine the meaning of a morphologically related word. It also appeared that knowledge of vocabulary-based structures such as roots and affixes provided students with knowledge to access the meaning of words.

Fluency

In several syntheses, the effects of intervention on reading fluency outcomes were reported. The highest effect sizes (*g* = 0.40 to 0.64) for fluency outcomes were reported in a review of multicomponent reading interventions for adolescent ELs (Huddle et al. 2017).

One strategy for improving reading fluency was highly beneficial. This was the use of repeated readings, a well-established strategy to target and boost fluency in younger readers, which yielded an overall large effect size (*d* = 1.31) on the reading fluency performances of underachieving adolescent readers (Joseph and Schisler 2009). In another review, improved reading fluency following repeated reading interventions was also reported for underachieving readers in Years 6 to 12 (Steinle et al. 2022). However, there was inconsistency between the results of single-case research designs and group designs with respect to positive effects on fluency outcomes. It was concluded that the inconsistent results of all studies indicated a lack of strong consensus that repeated readings is linked to improved fluency for underachieving secondary readers (Steinle et al. 2022).

In several reviews, combining repeated readings with other intervention components was found to produce the largest effect sizes. Explicit reading instruction incorporating repeated reading lessons led to significant gains (*d* = 1.52) for adolescent students with disabilities (Joseph and Schisler 2009). Repeated reading interventions, together with listening passage previews, produced significantly higher effects (*g* = 1.95) than repeated reading interventions without listening passage previews (*g* = 0.94) for students with, or at risk of RD (Lee and Yoon 2017). Similarly, Wexler et al. (2008) found that secondary students who received repeated readings together with a model of fluent reading by an adult or more competent peer, demonstrated more gains in fluency rate than students who received repeated readings without a model.

Findings regarding effectiveness from other intervention strategies were mixed. Some of the reading interventions showed negative or small effects for reading fluency. For example, morphological instruction yielded no statistically significant intervention effects on reading fluency (*d* = -0.05, Goodwin and Ahn 2013). In other studies, there was evidence that even extensive interventions (≥ 75 sessions) yielded only small effects for fluency (*g* = 0.16) for adolescents with reading difficulties (Wanzek et al. 2013). Importantly, Wanzek et al. (2013) did not explicitly address reading fluency but tested to see whether fluency improved following general multicomponent interventions, many of which were directed at comprehension.

Comprehension

Most reading intervention systematic reviews addressed reading comprehension. As noted in the introduction, reading comprehension is not itself a skill but is the outcome of proficiency in phonemic awareness, decoding, vocabulary and fluency, and is teachable through particular strategies. Following a discussion of general findings for comprehension interventions, findings related to the various comprehension strategies that were examined in the systematic reviews are presented under sub-headings for each of the identified strategies.

30 See section on phonemic awareness.

Interventions that specifically targeted reading comprehension were shown to be effective. Evidence from meta-analyses of studies in which comprehension interventions, including outcomes for underachieving older readers, were investigated, indicated positive and moderate effects on reading comprehension skills (*g* = 0.59, Filderman et al. 2022; *g* = 0.45, Scammacca et al. 2015; Scammacca et al. 2007). Further evidence for the effectiveness of comprehension interventions for both criterion-referenced measures[31] (*g* = 0.70) and norm-referenced tests (*g* = 0.52) was found in a synthesis of studies by Berkeley et al. (2010). The highest effect sizes found by these authors were associated with fundamental reading skills instruction (*g* = 1.04), followed by questioning/strategy instruction (*g* = 0.75) and text enhancements (*g* = 0.62). In another systematic review, school-based interventions targeting comprehension yielded a small, but statistically significant, effect size (*g* = 0.155, Dietrichson et al. 2020).

To understand these variations, it is important to consider the instructional components and content foci of the interventions. There was some evidence that interventions targeting decoding or word study produced small to moderate effects on reading comprehension measures for adolescent (*d* = 0.34, Edmonds et al. 2009; *d* = 0.55, Joseph and Schisler 2009; *g* = 0.40, Scammacca et al. 2007) and upper elementary and middle-school underachieving readers (*g* = 0.73, Flynn et al. 2012). By contrast, interventions involving morphological instruction showed no significant effects on reading comprehension outcomes for underachieving students and typical achievers in Kindergarten to Year 9 (*d* = 0.09, Goodwin and Ahn, 2013). Findings regarding the impact of fluency interventions on comprehension outcomes showed that while repeated reading interventions for secondary readers generally improved overall reading rate and accuracy, they did not improve comprehension outcomes (Steinle et al. 2022; Wexler et al. 2008). These findings were confirmed in other systematic reviews in which it was reported that interventions targeting reading fluency yielded negative effect sizes on comprehension outcomes for underachieving adolescent readers (*d* = -0.03, Edmonds et al. 2009; *g* = -0.07, Scammacca et al. 2007).

Strategies for teaching reading comprehension

Particular strategies to teach reading comprehension were examined in some systematic reviews. Explicit instruction and strategy instruction were the 2 main strategies identified. The findings are presented below.

Explicit instruction

Interventions that utilised explicit instruction were found to be effective for older students (Hall-Mills and Marante 2022; Kaldenberg et al. 2015). Explicit text structure instruction led to improvements in reading comprehension of expository text (such as, intended to examine a topic and convey ideas, concepts and information; Valasa et al. 2014) for adolescent students with, or at risk of, LD (Hall-Mills and Marante 2022).

Most of the effect sizes were large, although effects ranged from small to large (d = 0.31 to 2.17). In 6 of the included studies from one systematic review (Hall-Mills and Marante 2022), the gains in expository text comprehension after explicit instruction interventions were also statistically significant. Similarly, improvements in the comprehension of science texts for students with LD in Years 5 to 11 were found following explicit teaching of the definitions of vocabulary words, using a direct instruction approach or cognitive strategy instruction[32] (g = 2.07; Kaldenberg et al. 2015).

Results from a meta-analysis also indicated that middle and secondary students with reading difficulties and disabilities improved their reading comprehension when provided with targeted, explicit instruction in comprehension strategies, such as modelling and thinking aloud about how to self-question and reflect during and after reading *(d* = 1.23; Edmonds et al. 2009).

31 Criteria-referenced assessment is the process of evaluating (and grading) the learning of students against a set of pre-specified criteria.
32 Cognitive strategy instruction involves teaching with a strong focus on explaining the general process to use and why to use the strategy within particular learning tasks.

Strategy instruction

Various types of strategy instruction led to positive gains in reading comprehension (Donker et al. 2014; Filderman et al. 2022; Solis et al. 2012; Stevens et al. 2019). A substantial mean effect for comprehensive reading (g = 0.36, based on 95 interventions and 180 effect sizes) was reported in a meta-analysis of learning strategy instruction interventions aimed at improving students' cognitive, metacognitive and management strategy skills, as well as motivational aspects and metacognitive knowledge (Donker et al. 2014).

Explicit instruction of comprehension strategies also yielded positive effects. Strategy instruction with and without background knowledge was found to be effective for underachieving readers in Years 3 to 12. Of all strategies, main idea strategy instruction (such as instructional practices for reading text and identifying the most critical information, that is, the main idea) yielded the strongest effect of g = 0.72, with the effects of inferencing, retell and prediction ranging from g = 0.56 to g = 0.60.

Strategies to teach text structure and metacognitive strategy skills yielded smaller effect sizes of g = 0.47 and 0.49 respectively (Filderman et al. 2022). Similarly, findings from one meta-analysis indicated a positive effect of almost one standard deviation (g = 0.97) for main idea and summarising interventions on underachieving readers' comprehension, albeit on primarily unstandardised measures (Stevens et al. 2019). Further evidence was also found to support using summarisation or main idea strategy instruction to improve understanding of text for students with LD (g = 1.41, Kim et al. 2012; g = 0.71 to 6.66 and *PND*[33] = 33 to 85, Solis et al. 2012) and for middle and high-school students underachieving in reading (g = 0.55 to 2.65, Daniel et al. 2021).

There was some evidence to support the effectiveness of instructional enhancements (graphic organisers and technology) for improving the reading comprehension skills of older students (Filderman et al. 2022; Kim et al. 2004; Kim et al. 2012; Solis et al. 2012). In one systematic review, it was found that using graphic organisers consistently related to large effect sizes (d = 0.80 to 1.08) for students with learning disabilities (Kim et al. 2004). However, in the studies included in this review that yielded significant effect sizes, it was reported that the mean post-test scores ranged from 48% to 83% correct, with 15 of the 29 post-tests yielding means below 70%. This indicated that using graphic organisers was not sufficient for ensuring that these students obtained acceptable levels of understanding (Kim et al. 2004).

Similarly, the findings from another systematic review (Kim et al. 2012) suggested that instructional modifications were helpful in improving reading comprehension skills for middle school students with LD, but their effects varied. For example, high effect sizes were reported in some studies in which graphic organisers were used. Yet, in other studies little or no differences between intervention and control groups were found (g = 0.07 to 1.77, Kim et al. 2012). Other strategies that were found to be effective for middle school students with LD included mnemonics use (g = 0.33 to 1.41; Solis et al. 2012), mapping (10% - 100% correct post-test responses, Solis et al. 2012) and questioning (g = 0.052 to 2.55; Solis et al. 2012).

There was evidence to suggest that specific student populations derived particular benefits from reading comprehension interventions. Findings suggested that interventions that specifically targeted students with learning disorders were associated with the highest gains in reading comprehension (Edmonds et al. 2009). Moderate average effects were found for samples of underachieving readers (d = 0.45), or both underachieving readers and students with disabilities (d = 0.68), but a large effect (d = 1.50) was found for systematic reviews with samples of only students with disabilities (Edmonds et al. 2009). By contrast, only a small positive effect (g = 0.10) of intervention on reading comprehension for students with disabilities was reported in another systematic review (Wanzek et al. 2013).

Evidence also suggested that interventions using practices such as strategy instruction, student grouping practices, and explicit instruction, implemented with students with reading difficulties and LD (Edmonds et al. 2009; Scammacca et al. 2007; Solis et al. 2012), also enhanced reading comprehension for students with autism spectrum disorders (ASD; El Zein et al. 2014), as well as for middle and secondary students with, or at

33 PND: Percentage of non-overlapping data.

risk of, emotional and behavioural disorders (EBD; Tau *U* = 0.59, Burke et al. 2015). There was less evidence of reading comprehension gains for ELs (Oxley and de Cat 2021), with a considerable amount of variability in effect sizes (*g* = -0.20 to 0.71) reported on comprehension measures for adolescent ELs following reading interventions (Huddle et al. 2017).

Findings regarding the benefits from reading comprehension interventions for students of particular ages were mixed. Evidence from most systematic reviews in which age or year levels of participants were reported suggested that these factors did not moderate the effects of intervention (Filderman et al. 2022; Goodwin and Ahn 2013; Hall and Burns 2018). By contrast, other findings from Berkeley et al. (2010) showed significantly larger mean treatment effects on criterion-referenced measures for middle and high-school students (*g* = 0.80) than were the mean effects for elementary school students (*g* = 0.52).

Moderators of reading outcomes (setting, duration and frequency)

Findings from the systemic reviews in relation to the moderating variables of setting, duration and frequency are presented below.

Findings about intervention settings as moderators of reading outcomes pertained to intervention group size; these findings were mixed. General academic interventions delivered in small groups were shown to be effective on reading outcomes (*g* = 0.38; Dietrichson et al. 2020). There was some evidence of a negative relationship between group size and reading intervention effects (Hall and Burns 2018; Kim et al. 2012; Stevens et al. 2019). Only one systematic review reported a correlation value of *r* = −0.21 for this negative relationship (Hall and Burns 2018). In another systematic review, overall larger effects sizes were found for individual instruction (*g* = 1.20 to 2.57; Kim et al. 2012) than for larger groups (*g* = 0.07 to 5.23; Kim et al. 2012), although there was variability between individual studies.

One group of reviewers (Stevens et al. 2019), noted that while group size was not a statistically significant moderator, smaller groups (one to 4 students) showed a mean effect of almost 1.5 standard deviations, compared to a mean effect of one standard deviation for interventions provided in larger groups (≥ 5 students). By contrast, in another systematic review it was reported that larger groups (≥ 11 students) were associated with larger effect sizes (*SMD*[34] = 1.16) than smaller groups (*SMD* = 0.87, Swanson et al. 2014). Other reviewers found that group size did not significantly moderate differences in reading outcomes (Berkeley et al. 2010; Goodwin and Ahn 2013; Stevens et al. 2019; Wanzek et al. 2013).

Findings regarding intervention duration as a moderator of effects of reading interventions were limited to comprehension outcomes. The evidence was mixed. The results of morphological instruction studies for underachieving students and typical achievers in Kindergarten to Year 9 showed that interventions that were more than 20 hours long produced moderate effects (*d* = 0.40) on reading comprehension (Goodwin and Ahn 2013). However, there were no clear findings from other systematic reviews regarding the duration of interventions. In one systematic review (Berkeley et al. 2010), it was found that for criterion-referenced measures, mean weighted treatment effect sizes were significantly higher for interventions of medium duration (> 1 week but less than 1 month, *g* = 0.84) than for interventions of shorter duration (< 1 week, *g* = 0.48), or for longer durations (> 1 month, *g* = 0.50); however, differences on norm-referenced tests by intervention duration were not statistically significant.

Generalisation and maintenance of reading outcomes

There were limited findings pertaining to generalisation and maintenance of reading outcomes, as few studies included distal measures. The evidence that was reported is presented below for reading in general and for reading comprehension.

34 SMD: Standard mean difference.

Limited data were reported from studies of general reading interventions to show whether reading outcomes for underachieving readers in middle and high school were maintained over time. In one systematic review (Daniel et al. 2021), it was reported that across all studies, the effect of treatment was $g = 0.78$ at post-test and $g = 0.27$ at follow-up.

Findings regarding the generalisation and maintenance of reading comprehension effects were mixed. There was some evidence that gains were maintained ($g = 0.36$, Filderman et al. 2022; $d = 0.34$ to 0.52, Hall-Mills and Marante 2022; $d = 0.60$, Kim et al. 2004).

There were also findings suggesting that explicit text structure instruction transfers across content areas and that gains are maintained over longer periods of time for adolescents with LD (ηp^2 [35]$= 0.05$ and $d = 1.57$, Hall-Mills and Marante 2022). Similarly, a large effect size ($d = 0.60$) was reported on follow-up tests of reading comprehension for school students with LD, following interventions using graphic organisers (Kim et al. 2004). In contrast, there were indications from some systematic reviews that reading comprehension improvements were not generalised (Edmonds et al. 2009; Filderman et al. 2022; Kim et al. 2004; Stevens et al. 2019) nor maintained (Kim et al. 2012).

Standardised tests, often norm-referenced, are considered to be an indication of the generalisability of intervention gains, as they are measures independent of the researcher or intervention (Hanushek 2021). Accordingly, such findings are also reported here. Evidence from standardised tests from the systematic reviews examining reading interventions was varied. A few reviewers found positive and statistically significant results in standardised tests of reading ($g = 0.38$, Dietrichson et al. 2020). Other reviewers reported positive effects ($g = 0.28$, Filderman et al. 2022; $g = 0.42$, Scammacca et al. 2007) or positive small effects ($g = 0.21$, Scammacca et al. 2015; *ES* not reported, Stevens et al. 2019) on standardised or other independent tests ($g = 0.45$, Donker et al. 2014) that were significantly lower than on researcher developed measures ($g = 0.78$, Donker et al. 2014; $g = 0.73$, Filderman et al. 2022). In one synthesis, no significant effects on standardised tests were found (Kim et al. 2004).

Writing

Description of systematic reviews

Fourteen systematic reviews relevant to writing interventions for older, underachieving students were identified; this constituted 20% of the total sample included in our review. Collectively, these systematic reviews reported on the outcomes from a range of writing interventions for at least 12,101 participants with and without disability. It should be noted that numbers of participants were not reported in some reviews; this means that the total was higher than the reported 12,101. The numbers of participants and other details are summarised in Appendix A.5.

In the systematic reviews involving writing interventions, the effects on writing skills in general, or specific aspects of writing – writing fluency, sentence writing, text writing and spelling – were reported. The findings are outlined in the following sections. It should be noted that in some systematic reviews, an overall effect size for the reviewed studies was calculated. These are reported below as a single value. In other systematic reviews, only effect sizes for individual studies were reported. These effect sizes are presented below as a range of values only when the data could be reliably determined; readers are directed to the original articles for further information.

Several effective strategies for teaching writing skills were also identified and have been collated in Appendix A.9.

Focus and outcomes of writing interventions on general writing skills

All researchers that reviewed or analysed general writing interventions reported overall improvements for students underachieving in writing. Indeed, Datchuk et al. (2020) concluded that all students, regardless of disability type, gender or race, may benefit equally from writing interventions. In another systematic review (Kaldenberg et al. 2016), it was found that a variety of instructional strategies embedded within interventions including sequencing, drill-repetition-practice, segmentation of teaching instruction, directed question and responses, controlling for

35 Partial eta squared – an effect size measure of different variables in ANOVA models. Interpretation of values for partial eta squared: 0.01: small effect size; 0.06: medium effect size; 0.14 or higher: large effect size. Reference: https://www.statology.org/partial-eta-squared/

task difficulty, technology and strategy cues, improved the writing of students with LD (mean *phi* = 0.71 to 0.83). Other researchers reported moderating variables that influenced the effectiveness of intervention outcomes, including type of writing interventions and student characteristics. These findings are outlined below.

There were many types of writing interventions included in the included systematic reviews. Effective interventions for improving writing for students with LD (Gillespie and Graham 2014; Kang et al. 2015; Peterson et al. 2020) and high incidence disabilities (Valasa et al. 2014) included those that featured explicit (*g* = 1.199; Peterson et al. 2020) or systematic instruction, teacher support and scaffolding (*g* = 0.74; Gillespie and Graham 2014), explicit training in cognitive strategies (*g* = 1.15 to 4.16; Kang et al. 2015) and strategy instruction, which involved modelling of strategies for planning, writing, revising and/or editing text, and student practice of the strategies (Gillespie and Graham 2014). Some writing treatments, that is, strategy instruction (*g* = 1.09), dictation (*g* = 0.55), goal setting (*g* = 0.57) and process writing (*g* = 0.43) had statistically significant effects on the writing quality of students with LD (Gillespie and Graham 2014).

Learning strategy instruction that focused on improving self-regulated learning including cognitive, metacognitive and management strategy skills, as well as motivational aspects and metacognitive knowledge, resulted in substantial effects for writing for students with and without disabilities and learning difficulties (*g* = 1.25; Donker et al. 2014). Interventions that combined (general) metacognitive knowledge and evaluation (a subcategory of metacognitive strategies) were shown by the regression coefficient (β) to have higher effect sizes than other interventions that didn't include metacognitive knowledge and evaluation (β = 0.75 and 0.57, respectively; Donker et al. 2014).

Additionally, several systematic reviews examined the impact of strategy instruction through SRSD for students in particular categories of disability. This was found to be a promising approach for writing intervention for students with ASD (*ES* value not reported, *p* = 0.69; Asaro-Saddler et al. 2021) and students identified with LD (*g* = 1.33, Gillespie and Graham 2014; mean *phi* = 0.83, Kaldenberg et al. 2016; *g* = 1.15, Kang et al. 2015; *d* = 0.42 to 8.97 and *PND* = 58% to 100%, Valasa et al. 2014). It was noted that more research is needed to conclusively recommend SRSD (Asaro-Saddler et al. 2021) and strategy instruction (see below) without explicit instruction in self-regulation, as best-practice narrative essay intervention techniques for adolescents with disabilities (Valasa et al. 2014).

Interestingly, positive impact was reported for writing outcomes arising from interventions that did not explicitly focus on writing as the phenomenon of interest. There was some evidence that reading interventions enhanced performances on multiple writing measures (*g* = 0.57) for students underachieving in literacy and typically developing students. This supports the assumption that there is a bidirectional relationship between reading and writing (Graham et al. 2018). For example, benefits were found to arise for students' writing achievement due to reading interventions, specifically interventions focused on phonological awareness (*g* = 0.69), phonics (*g* = 0.69), reading comprehension instruction (*g* = 0.66) and increasing students' interaction with text (*g* = 0.35: Graham et al. 2018). DBI, aimed at improving academic performances for students with intensive learning needs, was shown to be more effective for spelling/writing (*g* = 0.47) than other academic areas such as reading (Jung et al. 2018).

Focus and outcomes of writing interventions on specific writing skills

Some writing studies focused on the impact of interventions for distinct elements of writing, that is, writing fluency, sentence writing, text writing and spelling. These results are summarised below with subheadings for each of these writing subskills.

Writing fluency

Positive effects on writing fluency were found for students with learning and other disabilities (Datchuk et al. 2022; Datchuk et al. 2020). Findings regarding the most effective writing interventions for writing fluency were mixed.

There was some evidence that different interventions produced similar effects on writing fluency. Gains in total words written per minute (TWWpM) per intervention session were achieved following acquisition interventions (such as targeted to an area of low proficiency and reliant on modelling, prompting, initial practice and error correction; TWWpM = 0.16; Datchuk et al. 2022) and fluency interventions that featured timed practice paired with one or more components of initial instruction (for example, explicit instruction or SRSD; see below), performance feedback, goal setting and graphing of performance (TWWpM = 0.10; Datchuk et al. 2022). By contrast, other evidence suggested that different writing interventions such as direct instruction and SRSD had similar effects on correct writing sequences (written) per minute (CWSpM = 0.39; Datchuk et al. 2020) and per intervention session. It was noted that, on average, intervention produced a gradual, rather than an immediate, positive change in the trend of writing fluency (Datchuk et al. 2020).

Sentence writing

There was limited evidence of the effects of interventions on sentence writing. For students with disability, a higher rate of correct writing sequences per minute (CWSpM = 1.3) was seen on sentence tasks compared to discourse tasks of narratives and essays, following intervention (Datchuk et al. 2020).

Text writing

Several intervention components were found to have a positive benefit for text writing outcomes. One that arose from several systematic reviews was using graphic organisers. Findings indicated that using graphic organisers improved the writing skills of students on several dimensions such as writing quality, genre elements, text productivity, syntax and vocabulary (Boon et al. 2018; Datchuk et al. 2022). Computer-based graphic organiser interventions yielded a large overall effect size (Tau *U* = 0.65; Boon et al. 2018), whereas the overall effect size for paper-and-pencil graphic organiser interventions was moderate (Tau *U* = 0.37; Boon et al. 2018). Teacher versus student-generated graphic organisers were equally effective with both yielding moderate effect sizes (Tau *U* = 0.48 and 0.47, respectively; Boon et al. 2018).

It was noted that using graphic organisers in the writing process should be considered not in isolation, but as a supplementary tool of effective writing instruction for students with LD (Boon et al. 2018; Ciullo and Reutebuch 2013). It was reported in one systematic review (Ciullo and Reutebuch 2013) that studies with the highest effect sizes (*g* = 0.80 to 2.41 and *PND* = 67%) incorporated principles of effective instruction for students with LD (for example, explicit instruction, guided practice; Vaughn et al. 2000) and effective literacy recommendations from the Institute of Educational Sciences (for example, explicit strategy instruction, using practices like technology to increase motivation, and intensive support for underachieving readers by trained professionals; Kamil et al. 2008).

Moderation of effect sizes by type of text was found for students with LD and other disabilities. Mostly large and very large effect sizes were reported for expository writing studies (overall ES: Tau *U* = 0.72); in narrative writing studies, a wide range of effect sizes was reported with an overall effect size of Tau *U* = 0.42 (Boon et al. 2018).

Spelling

School-based interventions for students in Years 7 to 12 with, or at risk of, academic difficulties yielded average statistically significant effect sizes (*g* = 0.167) on measures of spelling. Small-group instruction was associated with the largest effect sizes (*g* = 0.38; Dietrichson et al. 2020). Overall, significant effects (*d* = 0.30) were found for morphological instruction on spelling proximal outcomes (Goodwin and Ahn 2013). In this systematic review, statistically significant intervention effects were reported for students in upper elementary (*d* = 0.29) and middle school (*d* = 0.34) but not for high-school students (*d* = 0.45). ELs showed the largest intervention effect (*d* = 0.54), followed by children with LD (*d* = 0.37), then poor readers/spellers (*d* = 0.35) and typical achievers (*d* = 0.29).

There was some evidence that secondary students with disabilities were generally able to increase spelling accuracy for words directly taught in the interventions, or for words written in compositions. However, there was some variability for individual participants, and some of the interventions could be interpreted as having minimal practical significance despite improved performance from baseline to intervention phases. Findings suggested that the most effective spelling interventions for older students are: (1) systematic study procedures that emphasise repeated practice with words and self-correction of misspelled words (Tau *U* = 0.12 to 1.13; Williams et al. 2018), and (2) technological assistance interventions where students with disabilities learn to use word prediction software to improve spellings in written compositions (Tau *U* = 0.93; Williams et al. 2018).

A small positive effect of *g* = 0.15 on spelling outcomes following interventions to improve reading outcomes for middle school students with disabilities was also reported (Wanzek et al. 2013).

Moderators of writing intervention outcomes (setting, duration and frequency)

No evidence for the moderating effect of frequency of intervention sessions on writing outcomes was found.

Evidence for effective writing intervention durations was scarce. There was limited evidence that interventions of longer duration resulted in improved writing outcomes (such as, the number of words was estimated to increase by an average of 3.53 words per session; Asaro-Saddler et al. 2021).

No evidence for the moderating effect of frequency of intervention sessions on writing outcomes was found.

Generalisation and maintenance of writing intervention outcomes

There was no clear evidence to suggest that strategies taught during intervention generalised across writing domains (Peterson et al. 2020) or that improvements in writing were maintained (Graham et al. 2018; Williams et al. 2018). It was noted that SRSD holds the most promise as an effective writing intervention that may also promote the maintenance of writing skills (Valasa et al. 2014).

Mathematics

Description of systematic reviews

There were 14 systematic reviews constituting 20% of the total set of systematic reviewed classified as relevant to the category of mathematics interventions. While this indicates a relatively lower proportion of total intervention reviews for older students pertaining to mathematics than to literacy, they collectively represent the intervention impact evaluation for over 16,000 students including those with and without disability (*n* = 16,307). Approximately two-thirds of the systematic reviews in mathematics were quantitative syntheses. Details of the systematic reviews related to mathematics are summarised in Appendix A.5.

The systematic reviews involving mathematics interventions addressed the effects on mathematics skills in general or on specific mathematics topics: word problem solving, algebra, fractions, basic number facts and arithmetic skills. The findings are outlined in the following sections. It should be noted that in some systematic reviews, an overall effect size for the reviewed studies was calculated; these are reported below as a single value. In other systematic reviews, only effect sizes for individual studies were reported. These effect sizes are presented below as a range of values only when the data could be reliably determined; readers are directed to the original articles for further information.

Several effective strategies for teaching mathematical skills and concepts were also identified and are collated in Appendix A.9.

Focus and outcomes of mathematics interventions on general mathematics skills

Interventions designed to target general academic skills for secondary students with, or at risk of, learning problems produced notable improvements in mathematics (*g* = 0.34, Dietrichson et al. 2020; *g* = 0.66, Donker et al. 2014).

The principles of effective mathematical intervention for elementary students, for example, modelling, including visual models, guided practice, independent practice, monitoring student performance and/or corrective feedback, were shown as also effective for middle and secondary school students (overall *g* = 0.37; Jitendra et al. 2018; *d* = 0.43 to 3.87; Maccini et al. 2007). Similarly, the concrete-representational-abstract approach[36] (CRA) was shown to be also effective for older students (Bouck et al. 2018). It was noted in the latter systematic review that few of the studies (10 out of 29 studies) reported effect sizes; the effect sizes for the 10 studies were described as strong, but the values were not reported.

Strong intervention effects were found when combining visual strategies (for example, fraction circles or squares, manipulative objects, diagrams, mental representations) with other strategies (for example, priming the underlying problem structure, cognitive strategy, contextualised instruction such as problems using real-world scenarios; Jitendra et al. 2018). In this systematic review, it was reported that using visuals combined with other strategies yielded an average effect size of *g* = 0.52, compared to an average effect size of *g* = 0.28 for using visuals only, but the difference was not statistically significant. Other effective strategies identified for improving general mathematics skills included: (1) mnemonic strategy instruction (range of *d* = 1.68 to 2.22; Maccini et al. 2007), (2) graduated instructional approach, including CRA (range of *d* = 0.43 to 3.87; Maccini et al. 2007), (3) explicit teaching of cognitive strategies (see above), especially planning (range of *d* = −0.2 to 1.4; Maccini et al. 2007) and elaboration, that is, explicitly linking new content to already existing knowledge (*g* = 0.66; Donker et al. 2014) and (4) contextualised videodisc[37] instruction (range of *d* = 0.56 to 0.81; Maccini et al. 2007).

Interventions that included small-group instruction, peer-assisted instruction, progress monitoring, computer assisted instruction[38] (CAI) and coaching of personnel had positive and significant average effect sizes (*g* = 0.38, 0.19, 0.19, 0.17 and 0.10, respectively; Dietrichson et al. 2020). Similarly, interventions utilising DBI yielded an average effect size of *g* = 0.37 on mathematics outcomes (Jung et al. 2018). Interventions that provided incentives for students did not have a significant average, effect size (g = 0.05; Dietrichson et al. 2020). It is noted that studies included in most of the systematic reviews used only proximal measures, especially research-developed measures (see Section 3.4.3.3.d, below).

Focus and outcomes of mathematics interventions on specific mathematics skills

Most of the systematic reviews focused on the impact of interventions for a specific mathematics topic, that is, word problem solving, algebra, fractions, basic number facts and arithmetic skills. The results are presented below for each of the mathematic topics in turn.

Word problem solving

In several systematic reviews on interventions for secondary students in mathematics, the focus was on word problem solving. Evidence arising from these reviews indicated that word problem solving interventions improved the performances of underachieving middle and high-school students, notably those identified with LD, or for those who were identified as being at risk of developing learning difficulties in mathematics (Lein et al. 2020; Shin et al. 2021; Xin et al. 1999). Average effect sizes were mostly moderate (*g* = 0.56, Lein et al. 2020) to large (between-case standardised mean difference *[BC-SMD]* = 4.52, Shin et al. 2021; *d* = 0.89, Xin et al. 1999).

36 Also referred to as the concrete-representational-abstract sequence of instruction.
37 The video program was used as an 'anchor' or situation for creating a realistic context to make learning motivating, meaningful and useful.
38 Computer-assisted instruction refers to instruction or remediation presented on a computer. Many educational computer programs are available online and from computer stores and textbook companies. They enhance teacher instruction in several ways. For further reading, refer to https://www.readingrockets.org/article/computer-assisted-instruction-and-reading

Notably, in these systematic reviews it was revealed that performance improvements and effect sizes were moderated by intervention type. The most effective interventions were multicomponent, incorporating explicit teaching of strategies in the identification of underlying structures and relationships among key components within word problems, also known as schema-based instruction (SBI; *g* = 0.40 to 1.06, Lein et al. 2020; *BC-SMD* = 6.21, Shin et al. 2021) and with student practice involving visual models and strategic use of appropriate tools (*BC-SMD* = 3.98 to 0.21, Shin et al. 2021). CAI was also shown to be effective (*d* = 1.80, Xin et al. 1999).

Algebra

Two systematic reviews examined interventions that focused on algebra (Lee et al. 2020; Watt et al. 2016). Overall, the effects of interventions targeting algebra were moderate (*g* = 0.48, Watt et al. 2016) to large or very large (*g* = 0.37 to 1.38 and Tau *U* = 0.33 to 1.00, Lee et al. 2020). Variability in effect size magnitudes was noted between studies (Watt et al. 2016) and between individual participants (Lee et al. 2020). Beneficial instructional strategies for teaching algebra were incorporated in the effective interventions. These included: CRA sequencing, explicit instruction (including modelling the steps for solving problems, sufficient practice opportunities and ongoing feedback), multiple representations (for example, concrete and virtual manipulatives, pictures) and a logical sequence of a range examples (for example, addition, subtraction, multiplication and division).

Fractions

In 3 systematic reviews, the impact of interventions for fractions learning was reported. In one of the systematic reviews, fractions interventions utilising the features of explicit and systematic instruction – including explicit modelling of step-by-step strategies, sufficient practice, and cumulative reviews with feedback[39] in combination with concrete and visual representations – yielded large and/or significant effects (*g* = 1.05 to 2.45 and Tau-*U* = 0.74 to 1.00, Shin and Bryant 2015). Similarly, in another review (Maccini et al. 2007), studies with significant treatment effects for fractions learning incorporated features of effective instruction (see Arithmetic skills below). In the third review, the reported average weighted effect size for fractions interventions was moderate (*g* = 0.41; Stevens et al. 2018).

There was also evidence for the effectiveness of other multicomponent fraction interventions, for example, those that combined student verbalisations of mathematical thinking combined with virtual manipulatives (Tau *U* = 0.74 to 1.00; Shin and Bryant 2015).

Basic number facts

There was a paucity of systematic reviews targeting improvements in basic number facts for older students. Evidence for the effectiveness of interventions involving the CRA approach with manipulatives for improving basic number facts (described as basic operations) were reported in one systematic review (Bouck et al. 2018; see above).

Arithmetic skills

Two systematic reviews included studies of interventions that focused on improvements in arithmetic or computational skills. In one systematic review, it was reported that interventions incorporating CRA with manipulatives were shown to be beneficial for improving arithmetic skills (Bouck et al. 2018; see above).

In another systematic review (Maccini et al. 2007), computations involving decimals, fractions, geometry, integers, solving linear equations and problem-solving tasks for secondary students with LD also incorporated features of effective instruction[40] including modelling, guided practice, independent practice, monitoring student performance and/or corrective feedback methods to measure student outcomes including percentage of correct computations between baseline and treatment.

39 Refer to National Mathematics Advisory Panel, 2008 for further information.
40 See also Rosenshine and Stevens (1986).

Number sense

There were no systematic reviews in which number sense interventions for students in Years 7 to 9 underachieving in mathematics were examined.

Moderators of outcomes (setting, duration and frequency)

Findings about intervention settings were rarely reported as moderators of intervention outcomes in the systematic reviews of mathematics interventions. The setting moderators that were reported pertained to intervention group size. These were noted in 3 of the systematic reviews (Jitendra et al. 2018; Lein et al. 2020; Stevens et al. 2018). In each of these systematic reviews, it was found that group size did not significantly moderate differences in mathematical outcomes.

There was some evidence that the duration of mathematical interventions significantly mediated effect sizes, that is, intensifying instruction by increasing instructional time was beneficial ($p < .05$; Jitendra et al. 2018; $p = 0.013$; Stevens et al. 2018; $p < .02$; Xin and Jitendra 1999). The duration of more effective interventions varied from more than 10 hours (Jitendra et al. 2018) to more than 15 hours (Stevens et al. 2018) and more than one month (Xin and Jitendra 1999). Interestingly, in one meta-analysis, it was reported that short-term (≤ 7 sessions) interventions were at least as effective as intermediate-term (> 7 sessions but ≤ 1 month) interventions but less effective than long-term (> 1 month) interventions (Xin and Jitendra 1999).

No data were found in the systematic reviews for the moderating effect of the frequency of interventions on students' mathematics outcomes.

Generalisation and maintenance of outcomes

In the systematic reviews, the evaluation of whether general mathematics interventions lead to sustainable and transferable (that is, to other contexts) performance gains were rare. Most studies examined within the systematic reviews did not include distal measures.

Standardised tests measures indicate the generalisability of intervention gains, as they are measures independent of the researcher or intervention (Lein et al. 2020; Peterson et al. 2020; Shin and McMaster 2019). Standardised tests are often norm-referenced (Hanushek 2021). As such, the following from studies using standardised or norm-referenced tests is presented as evidence of the generalisability of improvements in mathematical skills. In one systematic review (Dietrichson et al. 2020), performances on standardised tests as the only mathematics outcome measure was examined. In this systematic review, the average effect size for mathematics was low to moderate ($g = 0.34$, Dietrichson et al. 2020). Similarly, in another review, it was reported that norm-referenced measures yielded very small effect sizes ($g = 0.09$), whereas researcher-developed measures yielded larger effects ($g = 0.68$, Lein et al. 2020). In only one systematic review was it reported that generalisation measures yielded large effect sizes ($d = 0.84$, Xin et al. 1999).

For the few studies in the systematic reviews that included follow-up outcomes, the average effect sizes varied. In one systematic review, the average effect for outcomes measured more than 3 months after the intervention was small and not statistically significant ($g = 0.05$, Dietrichson et al. 2020), although substantial variation was noted. By contrast, in another systematic review (Xin et al. 1999), the unbiased mean effect size on maintenance was moderate to large ($d = 0.78$) for word problem solving skills.

Research question 4

What are the most effective ways to implement these approaches, and what pre-conditions are necessary to support their implementation?

- Who should be implementing these approaches? (For example, primary trained teachers, classroom assistants, speech therapists.)
- What resources and supports do schools need to implement the recommended approaches with fidelity to the duration, intensity and frequency that is required? (For example, time structures, professional development opportunities and age-appropriate materials.)

Pre-conditions necessary to support implementation

Two meta-analyses offered findings about the pre-conditions that support the implementation of tiered interventions (Burns and Symington 2002; Burns et al. 2005). The findings from both meta-analyses were discussed in relation to Research question 1 and the effectiveness of multi-tiered frameworks in schools; however, it is important to note their relevance here. Both studies examined the impact of multi-tiered frameworks that were implemented where a collaborative group problem-solving model underpinned the selection and implementation of tiered interventions. These problem-solving teams incorporated interprofessional collaboration between school psychologists and educators incorporating consultation, observation and conferencing (Burns and Symington 2002). They also incorporated the selection and implementation of research-based interventions and close monitoring of student progress (Burns et al. 2005). The findings showed that multidisciplinary teams using this problem-solving approach were associated with strong improvements in student academic and behavioural outcomes, by both Burns and Symington (2002; d = 1.15) and Burns et al. (2005; UEE = 1.02). On this basis, team-based problem-solving can be considered as a pre-condition for intensive interventions.

Effective ways to implement interventions

Findings relevant to how schools should implement interventions to achieve optimal student outcomes were reported in 4 systematic reviews (Fallon et al. 2015; Filderman et al. 2021; Jung et al. 2018; Scammacca et al. 2007). This set of 4 reviews was the smallest subset of systematic reviews included in our umbrella review, highlighting the relative lack of attention to this important issue within the studies focusing on intervention effectiveness. The participants involved in the studies totalled 7,156 (See Appendix A.5 for more detail). Several moderating factors were found to influence the outcome of interventions.

Who should implement interventions?

There were no explicit and consistent findings regarding the most effective implementer of intervention. No review included the question of who should implement interventions investigated directly. However, the relative impact of teachers implementing interventions compared to researchers implementing interventions was considered in some of the systematic reviews. There was also evidence regarding the fidelity of intervention implementation. The fidelity of implementation is closely related to the question of who should implement interventions. The concept of implementation fidelity refers to the degree to which an intervention is consistently delivered and requires adherence to the protocols of the program. The protocols must incorporate processes of delivery that include implementer and participants, as well as the evaluation of the program, to ensure that the intervention has been delivered as intended. The findings from the systematic reviews regarding the implementers of intervention and the fidelity of interventions are outlined below.

There was mixed evidence regarding whether literacy interventions delivered by researchers were more effective than those delivered by teachers. Some studies indicated that this was the case. For example, in one systematic review (Kim et al. 2012), it was found that the effects of interventions were consistently stronger when delivered by a researcher (g = 0.86 to 3.14), relative to instruction delivered by a teacher (g = 0.07 to 1.14). Similarly, another group of reviewers of a large set of reading studies (Scammacca et al. 2007) reported that for all outcome measures, researcher-led implementation of interventions resulted in higher effect sizes (g = 1.70) than teacher-led

implementation (g = 0.63). Jitendra et al. (2018) also found that researcher-led implementation of mathematics interventions yielded a larger effect size (g = 0.70) than interventions delivered by school personnel (g = 0.35).

Conversely, other reviewers reported minimal differences in literacy intervention effects delivered by researchers or teachers. For example, in one systematic review, it was found that the differential mean of the teacher effect was not significantly different from the researcher effect (d = 0.37 and d = 0.29 respectively; Goodwin and Ahn 2013). Similarly, other reviewers (Berkeley et al. 2010) reported no significant difference between reading intervention effects delivered by researchers (g = 0.83) or teachers (g = 0.56). Other reviewers (Kim et al. 2004) found that it did not matter who implemented the intervention using graphic organisers, the intervention itself generated large effect sizes. Inconsistent results were found for the effects of mathematics intervention implementers. However, in one systematic review, greater intervention effectiveness was noted when the intervention was jointly implemented by teachers and researchers (Xin and Jitendra 1999).

In general, it was found that for studies in which there was adherence to quality indicators, such as fidelity of implementation, instruction and assessment, as well as professional development and maintenance, the effect sizes of the interventions were robust (Scammacca et al. 2007). Reference to fidelity was made in the study designs of 50 systematic reviews of interventions (72%), while in only 44 systematic reviews (64%) were fidelity data provided. In 15 systematic reviews of interventions, neither fidelity nor fidelity data were reported. In 5 of these 15 reviews reference was made to relevant fidelity or quality indicator guidelines, such as What Works Clearinghouse (Institute of Education Sciences). However, most of the authors of the systematic reviews who cited fidelity in the intervention process emphasised the need for future researchers to ensure fidelity of implementation at the forefront of their study designs (Ciullo and Reutebuch 2013; Collins et al. 2018; Datchuk et al. 2022; Edmonds et al. 2009; El Zein et al. 2014; Filderman et al. 2022; Garwood et al. 2014; Gillespie and Graham 2014; Graham et al. 2018; Hall-Mills and Marante 2022; Kaldenberg et al. 2015; Lee et al. 2020; Peterson et al. 2020; Reed et al. 2014; Scammacca et al. 2007; Shin et al. 2021; Solis et al. 2012; Steinle et al. 2022; Stevens et al. 2019; Stewart and Austin 2020; Swanson et al. 2014; Wanzek et al. 2013; Wexler et al. 2008; Xin and Jitendra 1999).

While there were no clear findings on who is in the best position to implement interventions, there were indications that people with content expertise are in a better position to do so. Some reviewers identified that specific teacher training formats may influence the effects of interventions. In 2 systematic reviews, it was found that frequent support for teachers positively influenced teacher expertise outcomes. For example, in one review (Filderman et al. 2021), it was reported that training teachers in data literacy had significant positive effects on both teacher knowledge and skills. Training through collaborative or active learning was more effective than individual learning (β = 1.31). In contrast, other forms of learning, such as coaching or content-based learning, had no effect (β < .00 and β = .01, respectively). In the other review, (Jung et al. 2018), it was found that supporting the professional learning of teachers in DBI to implement Tier 3 intervention resulted in significant effects on student outcomes (g = 0.37). These findings were confirmed in another systematic review (Fallon et al. 2015), in which it was reported that providing individual, in-person performance feedback to teachers conducting interventions improved academic outcomes for students receiving intervention at higher tiers of support; it was noted that both daily and monthly feedback sessions were impactful in this way.

Resources

Evidence from reviews of DBI revealed that effect sizes were obtained from researcher-developed CBM, teacher-generated CBM and commercial CBM (Filderman et al. 2021; Jung et al. 2018). Through moderator analysis, it was found that 13% of the effect sizes from studies that had been meta-analysed showed that teacher-generated CBM yielded larger effect sizes than researcher-developed CBM (g = 0.80 and g = 0.39, respectively). However, Jung et al. (2018) concluded that this did not constitute an argument in favour of teacher-generated CBM, and noted the strong limitation posed by the lack of technical adequacy of the teacher-generated measures. These reviewers also noted that there was no evidence that teacher generated CBM led to stronger DBI effects.

An additional finding from another systematic review (Scammacca et al. 2007) offered insights about resources for interventions. Evidence was found that interventions for older underachieving readers are most effective when provided as early as possible, suggesting that resources are best devoted to intervening early.

Research question 5

Are there major gaps in the evidence base that prevent these questions being fully addressed? What research could be undertaken to fill these gaps?

Our umbrella review represents a large set of systematic reviews involving the data from many thousands of students. Yet there are some clear gaps in the research base, and aspects of the research questions within our own study that we could not completely answer. These are outlined below.

Multi-tiered systems of support

We located 2 high-quality systematic reviews that applied meta-analyses that pertained to multi-tiered frameworks (Burns et al. 2005; Burns and Symington 2002). These systematic reviews offered important findings about the impact of multi-tiered frameworks when implemented in schools who were early adopters of this approach to service delivery in the years prior to IDEA. Yet, there has been little research to replicate this in the years since, aside from the evaluation of reading in primary schools by Balu et al. (2015) which was discussed in the introduction. Accordingly, there is a clear research gap in updating these findings to understand the impact of multi-tiered models for schools and students now that they have been scaled across the US. Additionally, we identified that there is insufficient research examining multi-tiered support frameworks in secondary schools, which was also noted in the systematic reviews that we examined (for example, Berkeley et al. 2009). All of the systematic reviews that we located examined multi-tiered systems of support within whole schools with no disaggregation for older students, meaning that it remains unclear whether the findings of impact that we reported were moderated by school or year level. Finally, there is a research-to-practice gap in terms of fidelity of on-the-ground implementation of multi-tiered frameworks (Berkeley et al. 2020) and the preparation of teachers to implement multi-tiered supports to students (Barrio et al. 2015).

Universal screening and progress monitoring

We identified several research gaps regarding the question of how schools should identify and monitor progress for secondary school students needing intensive support. One important gap is that, as noted in the results section for Research question 2, we only identified systematic reviews relevant to screening for reading. We located no systematic reviews offering clear evidence for how to screen students for writing or mathematics, meaning that it remains unclear how to screen students in secondary school for their achievement in these foundational areas. An additional gap is that substantial research is necessary to guide progress monitoring implementation if it is to be established as an evidence-based practice. Ardoin et al. (2013) noted that CBM-R reading fluency screening tools, while often used for progress monitoring, lack an evidence base to support this practice. We located no systematic reviews for monitoring progress in writing or mathematics interventions. The gaps in knowledge about progress monitoring in reading, writing and mathematics are multifaceted as there is a lack of evidence about the measures that should be used, as well as the optimal frequency of progress monitoring.

Reading

Several gaps in the reading research literature were identified. In particular, the need for more research of high quality and rigour (such as, fidelity of implementation, use of standardised measures and the nature of the counterfactual's instruction) is required (Stevens et al. 2019; Wexler et al. 2008).

There was a lack of research that investigated the impact of interventions on phonemic awareness and decoding skills for adolescent underachieving readers (for example, Browder et al. 2006). Also, although positive effects, such as increasing reading rate, resulted from a variety of fluency interventions, more research is warranted to determine the most effective way of increasing the fluency and ultimately, reading comprehension of secondary underachieving readers (Wexler et al. 2008). This research should include investigation of the nature of reading disabilities and difficulties for adolescents, to refine the approach to remediating fluency deficits for these students (Steinle et al. 2022). More research to examine the impacts of vocabulary interventions is also needed (Kuder 2017).

An additional gap is that many of the reading intervention studies were multicomponent, meaning that the relative contributing factors of each of the component parts of the interventions was unclear (for example, Kim et al. 2004). The need for investigations of the following components was specifically identified:

- research that includes only one type of strategy instruction in isolation (for example, main idea) rather than multiple, related strategies (for example, main idea and retell) to clearly demonstrate which strategies are most effective (Filderman et al. 2022)
- the effects of basic reading instruction on understanding particular text types such as expository text and text in other content areas such as science, social studies or history (Joseph and Schisler 2009)
- using graphic organisers compared to other comprehension strategies for students with LD (Kim et al. 2004).

Using both standardised and researcher developed outcome measures in studies is needed to determine the extent to which intervention methods lead to generalisation of learning beyond the immediate intervention context, for example, to new vocabulary items (Kuder et al. 2017).

There is also a need for further research to establish evidence-based practices for groups of underachieving adolescent readers who can be significantly underserved in both research and education. These include students with EBD (Garwood et al. 2014), significant cognitive disabilities (Roberts et al. 2013) and ELs (Huddle et al. 2017; Oxley and de Cat 2021; Wexler et al. 2008). Specific knowledge gaps include how to teach phonemic awareness and phonics to individuals with significant cognitive disabilities (Browder et al. 2006; Roberts et al. 2013) and those with EBD (Garwood et al. 2014).

Writing

The need for more high-quality writing research studies was identified (Asaro-Saddler et al. 2021; Gillespie and Graham 2014), including the need for RCTs with larger numbers of participants (Ciullo and Reutebuch 2013). Studies that evaluate long-term and generalisation measures are also required (Boon et al. 2018; Graham et al. 2018; Valsa et al. 2014).

It was noted that future research should evaluate comprehensive writing programs and multicomponent interventions that involve teaching a wide range of writing skills to students with LD (Gillespie and Graham 2014), adolescents with disabilities (Valasa et al. 2014) and more diverse samples of students (Datchuk et al. 2022). Data pertaining to fidelity of implementation were also lacking (Ciullo and Reutebuch 2013).

The impact on writing of various reading treatments, such as reading fluency instruction, vocabulary instruction, increased reading time, teaching students about the functions and structure of text, reading and analysing text, and observing readers' interaction with text, needs further investigation (Graham et al. 2018).

Mathematics

Overall, a need for high-quality mathematics intervention research employing rigorous designs, larger sample sizes, and the use of standardised mathematics assessments with better reliability and validity, was identified (Stevens et al. 2018).

More specifically, the need for evidence of the following was noted:

1. ideal components of mathematics interventions; that is, a multicomponent mathematics intervention, targeting multiple yet related domains, or an intervention that builds in complexity to address gaps in students' prior understanding (Stevens et al. 2018)
2. computer-based mathematics interventions for older students with learning disorders in mathematics (Stevens et al. 2018)
3. the acceptability by teachers and students of approaches implemented by researchers in settings outside of the classroom (Xin and Jitendra 1999)

4. maintenance of post-intervention gains (Xin and Jitendra 1999)
5. using manipulatives in mathematics devoid of the CRA instructional sequence (Bouck and Park 2018)
6. effective interventions targeting topics with which students with LD typically find difficult at the secondary school level such as ratios, proportions and percentages (Shin et al. 2021); application of the understanding of numbers to the rational number set, as well as defining and using functions to model relationships between quantities (Lee et al. 2020); and multiplicative reasoning and the inverse relationship with division of fractions, including making a connection to algebraic problem solving (Shin and Bryant 2015).

Maintenance and generalisation

Evidence for the maintenance and generalisation of both literacy and mathematics intervention effects was lacking. As such, more research about the long-term effects of interventions is required. Many reviewers identified a lack of data to assess the maintenance (Berkeley et al. 2010; Daniel et al. 2020; Dietrichson et al. 2020; Graham et al. 2018; Filderman et al. 2022; Jung et al. 2018; Peterson et al. 2020; Williams et al. 2018) and generalisation (Berkeley et al. 2010; Peterson et al. 2020; Valasa et al. 2014) of intervention gains.

Fidelity of implementation

Fidelity of intervention implementation was identified as a potential difficulty for both research and practice (Burns et al. 2002; Graham et al. 2018; Peterson et al. 2020; Shin et al. 2021). Further research is needed to determine the factors that lead to consistent implementation of field-based interventions (Burns et al. 2005; Filderman et al. 2022; Shin et al. 2021). More controlled studies of systemic outcomes in general appear warranted (Burns et al. 2005). Specifically, it was identified that RCTs are needed to examine the effect of the implementation of multi-tiered models on referrals to, and placements in, special education, student time in special education services and the number of students retained in a year level (Burns et al. 2005).

Study quality

A lack of high-quality research studies was noted across systematic reviews of both literacy (Filderman et al. 2022; Gillespie and Graham 2014; Kang et al. 2015; Stevens et al. 2018; Valasa et al. 2014) and mathematics (Bouck et al. 2018). The need for more experimental studies (Peterson et al. 2020) utilising large RCTs (Goodwin and Ahn 2013) was emphasised.

Resources

An additional gap was noted regarding the resources and supports needed for successful implementation of interventions using MTSS in secondary schools. The only resources that were considered in the systematic reviews included in our umbrella review were those relating to identifying underachieving students and monitoring their progress if they receive intervention (Ardoin et al. 2013). In particular, CBM assessments developed variously by researchers, teachers and commercial entities were examined; however, no clear findings emerged from these systematic reviews (Filderman et al. 2021; Jung et al. 2018). We did not identify any systematic reviews that examined age-appropriate materials or other resources. This is a clear research gap given that the materials often used for teaching foundational skills in reading, writing and mathematics may well be unsuitable for older students; however, this remains unknown.

Discussion

This umbrella review responded to the call to investigate how schools can best provide educational support and intervention to students who underachieve in reading, writing and mathematics. In undertaking our review, we examined the evidence for equitable and efficient service-delivery frameworks to provide support to underachieving students. We also examined the evidence for how these students should be identified by schools, which interventions are empirically supported for use in supporting them and how progress should be monitored.

A real strength of this umbrella review is the size of the dataset, meaning that readers can have increased confidence in the results. Our results arise from a total of 69 systematic reviews over 6 domain classifications over the past 23 years (1999 to 2022). The systematic reviews that we included in our umbrella review encompassed a very large number of studies (n = 1,984) and as well as a large number of participants that included students and teachers (n = 87,246); 90% of which were students in secondary school (Years 7 to 12). A total of 1,395 effect sizes were examined in the systematic reviews. Most of the systematic reviews that we examined were high in quality, although we note that there is a demand for more systematic reviews that adopt a robust and high-quality methodology and that maintain integrity under the scrutiny of critical appraisals.

We located a large number of systematic reviews deemed potentially relevant to our investigation, yet many failed to meet the inclusion criteria as determined by JBI for quality, integrity and replicability (JBI Global n.d.). We see these systematic reviews as opportunities missed, and we encourage future researchers to undertake high-quality systematic reviews and contribute to improving educational outcomes for underachieving students.

Multi-tiered support frameworks

In our umbrella review, the set of systematic reviews examining multi-tiered models was small, yet it represented a substantial number of primary studies. There were 2 high-quality meta-analyses of the effectiveness of multi-tiered frameworks (Burns and Symington 2002; Burns et al. 2005), complemented by systematic reviews of documentation relating to policy and the preparation of general education teachers to implement these frameworks in practice (Barrio et al. 2015; Berkeley et al. 2009; Berkeley et al. 2020). Collectively, these systematic reviews offer evidence of the impact of multi-tiered frameworks on schools and students, as well as how these frameworks can be scaled and their implementation optimised. There are a number of implications that arise from the shared findings.

One implication that arises from the meta-analysis of outcomes at the student and system level is that implementing multi-tiered supports and problem-solving teams in elementary and middle schools (students in Kindergarten to Year 8) is an effective and efficient whole-school approach to supporting the learning and progress of all students. In particular, Burns and Symington (2002) focused on the effectiveness of school-based collaboration involving school psychologists using a problem-solving model to develop interventions appropriate to students' support needs. This study showed that delivering educational support services through this approach was highly beneficial to both students themselves as well as the schools, resulting in improvements on both behavioural and academic outcomes for students with minimised separation from their peers, and reduced referrals for special education.

The results from Burns et al. (2005) supported these earlier findings. These researchers found that a multi-tiered framework, in which students' responsiveness to Tier 1 instruction was closely monitored and Tier 2 and 3 interventions were provided through a problem-solving approach, was effective for improving students' academic progress and behavioural outcomes and reducing special education placements. Burns et al. (2005) found that approximately 80% of students met benchmarks through quality instruction in the classroom, leaving approximately 20% needing additional support beyond (equivalent to Tier 2) and approximately 6% requiring intensive services (equivalent to Tier 3). The results from this second meta-analysis provide empirical support for the expected proportions of students needing support across the 3 tiers of RTI/MTSS (see Figure 1).

Another implication from the Burns et al. (2005) meta-analysis was that the average proportion of students referred beyond multi-tiered supports for special educational services in schools implementing RTI was very small. Burns et al. (2005) noted that this was, on average, 1.68% as compared to the national average of 5.7%. This finding suggests that schools using multi-tiered support frameworks achieved an average reduction of approximately 70% of special education placements when compared to schools using the traditional categorical service delivery model (Burns et al. 2005). It also indicates that under categorical frameworks, the likely number of students being inappropriately placed into special education is high.

The finding that multi-tiered approaches are efficient for schools and school systems, and effective for student academic progress, has implications for Australian schools and systems in terms of how we might go about addressing the large proportions of students who underachieve in reading, writing and mathematics. The proportion of secondary school students in Australia with such gaps is unacceptably high. For example, Australian students who are meeting the NMS in NAPLAN for Year 7 are working at the expected level of a Year 5 student. Thus, any student below that NMS is working at the expected level of a student at lower to mid-primary school level. In 2021, 16.4% of Year 7 students were at or below the NMS in reading, 24.8% were at or below the NMS in writing and 16.8% were at or below the NMS for numeracy (ACARA 2021). This gap in students' basic skills proficiency makes it very challenging for them to access learning across the secondary school curriculum.

Such large gaps can only be addressed through extensive and individualised basic skills intervention. The findings from the 2 meta-analyses discussed above point to the potential of multi-tiered frameworks as an effective way to minimise these gaps in the basic reading, writing and mathematics skills of secondary school students by preventing them from widening in the first place. The results from both Burns and Symington (2002) and Burns et al. (2005) suggest that these figures could be substantially lowered in Australia if Tier 1 instruction in primary schools were optimal and evidence-based and that learning gaps in 80% of students would be prevented. This preventative approach would require appropriately targeted intervention to be provided through a through a multi-tiered framework involving team-based problem-solving to the 20% of students needing targeted intervention to supplement Tier 1, and intensive intervention to the few with more persistent learning gaps. Scaling multi-tiered systems of support consistently across primary schools, and ensuring that Tier 1 and 2 instruction is implemented with fidelity, could see 95% of students making adequate progress and potentially mitigate the number of students who commence secondary school requiring intervention at the rates currently indicated by NAPLAN. Both meta-analyses indicate that this could be a low as 5% for secondary school students (Burns and Symington 2002; Burns et al. 2005).

A further finding is that schools do not need to wait for legislative changes to embrace multi-tiered frameworks. The positive impact of multi-tiered frameworks implemented in early-adopter schools prior to IDEA was examined in both meta-analyses (Burns and Symington 2002; Burns et al. 2005). These findings are highly relevant to Australia, where no such legislative reform has yet occurred, but where there is uptake of RTI/MTSS in individual schools[41] as well as school regions.[42] Across the Australian public education system, all states and territories are at various stages of readiness to move away from categorical frameworks to align with the needs-based requirements and funding provisions within the *Australian Education Act 2013*.

Multi-tiered frameworks were purpose-built to replace these outdated service delivery frameworks and are ideal to close the equity gaps entrenched in Australian school systems. The 2 US meta-analyses clearly show the potential benefits of implementing multi-tiered frameworks to ensure that support is provided where and when it is needed, without the requirement for students to have a diagnosis or categorical funding which, as noted in the introduction, creates barriers and delays in accessing support. This could be done in a considered manner, driven by research. Suggestions for future research include examining the implementation advice, and outcomes achieved in Australian early-adopter schools and school regions, to provide a contextualised and much-needed update to the meta-analyses from the US.

41 See for example, Bentleigh West Primary School. https://youtu.be/RbcgktNpBq4
42 See for example, the Melbourne Archdiocese of Catholic Schools. https://mtss.education/

An additional set of implications arising from these systematic reviews points to how Australia might achieve the implementation of MTSS with scale and consistency across school systems. This set of implications arose from the 2 systematic reviews in which implementation and scaling of MTSS through policy guidance was examined (Berkeley et al. 2009; Berkeley et al. 2020) and the systematic review focusing on teacher education for MTSS (Barrio et al. 2015). The 2 snapshots of policy implementation over a decade by Berkeley et al. (2009, 2020) found that legislating for multi-tiered models to reform educational service delivery for equity and inclusion is impactful and leads to a scaling up of their implementation across jurisdictions. While some states implemented a multi-tiered framework prior to legislation, such as Kansas, having legislation as well as policy guidance facilitated the further scaling and refining of the multi-tiered frameworks in jurisdictions across the country in the decade following IDEA. These systematic reviews also found that definitional and conceptual variation in MTSS models across jurisdictions resulted in national inconsistency and contributed to confusion. This suggests that in Australia, valuable work might be undertaken in providing conceptual and definitional clarity as schools and school systems begin to implement MTSS.

Two additional facilitators of implementation with scale and consistency stand out from the second snapshot of MTSS implementation (Berkeley et al. 2020). One standout facilitator was the role played by technical assistance provided through national and state-based centres. When the IDEA reforms were enacted, these technical assistance centres were small in number and many began as university partnerships. For example, university faculties in which there were researchers with expertise in multi-tiered models created and provided online professional learning modules and guidance for implementation. Several of these university collaborations have since become national technical assistance centres, such as the IRIS Center at Vanderbilt. Another standout facilitating factor was the leading role played by exemplary US states that had been early adopters, with Kansas repeatedly providing leadership in multi-tiered framework implementation and their documentation used as a blueprint by other states. These point to the potential benefits for Australia to invest in such technical assistance, collaborate with universities and learn from early adopters. The evidence suggests that this would offer valuable support to schools and school systems in implementing a scalable and consistent multi-tiered approach to school-based collaboration and tiered interventions.

An important implication arising from our review that offers suggestions for implementation in Australia is that MTSS has become the preferred contemporary framework for multi-tiered service delivery in the US in preference to RTI. As noted by Berkeley et al. (2009; 2020) there has been an important shift that took place in the conceptualisation of multi-tiered frameworks in the decade following IDEA. This shift was a move away from RTI, that had a narrow focus on special education and academic intervention, towards a comprehensive MTSS approach that emphasised prevention at Tier 1 and that encompassed student learning, behaviour and social development. Indeed, MTSS is now preferred as the framework for the majority of US states (Bailey 2017) and the original, federally-funded technical assistance centre, the National Center on RTI, has been replaced by the National Center on MTSS (2022).

RTI began with its roots in the special education paradigm. It was developed to address the over-identification and over-servicing of students identified with LD, and to reduce placement within special education and the limitations that this placed on students' access to the academic curriculum. Yet, while it was intended that this be achieved largely through prevention by having a strong foundation at Tier 1, in essence the early implementation of RTI was focused more on targeted and intensive interventions (Tiers 2 and 3) than on foundational instruction; this meant that it remained grounded in a special education paradigm of remediation. This was also noted by both Berkeley et al. (2009) and Barrio et al. (2015) who found that professional learning and pre-service preparation literature remained oriented towards special education, rather than on general education. These findings offer important lessons for Australia, where multi-tiered frameworks are just starting to be seen in the language of guidelines provided to schools about the funding of support.[43] Rather than following the US path of beginning with multi-tiered approaches articulated as RTI, and as a special education initiative, the evidence presented here suggests that preference should be given to a comprehensive MTSS approach from the outset.

43 For example, in Victoria https://www2.education.vic.gov.au/pal/tutor-learning-initiative/policy and NSW https://education.nsw.gov.au/student-wellbeing/contact-us/positive-behaviour-for-learning/school-wide-systems-of-support

The systematic reviews discussed here also offer some potentially valuable advice about pitfalls to avoid in early implementation. The first is that without clear mandates and definitions built into legislation and policy guidance, inconsistencies proliferate across jurisdictions. IDEA provided autonomy to the US states regarding the use of RTI, particularly in relation to the identification of students with LD. This autonomy led to divergent models at the state level with many divergent norms for advice and policy about how RTI should be implemented, with inconsistencies in how Tiers 2 and 3 were understood, staffed and resourced, as well as in how the models aligned with educational provisions for students with disability.

As Australia begins to tentatively adopt MTSS, within guidance to schools, one potentially valuable approach would be to learn from the US experience; that is national inconsistency can be avoided through clear and authoritative advice to schools across jurisdictions and sectors. This would entail avoiding the implementation of separated frameworks of RTI and PBIS, and to promote a braided and comprehensive MTSS framework to be used from the outset. A further pitfall from the US approach that slowed consistent scaling is that Australia might benefit from avoiding implementing a multi-tiered approach using the remediation logic of special education. The 2 meta-analyses (Burns and Symington 2002; Burns et al. 2005) suggest MTSS works when effective, inclusive and high-quality practice is provided for all students at Tier 1. This mitigates the number of students needing support at Tiers 2 and 3, as well as forming the basis on which intervention works. Stemming the tide of underachievement is Australia could therefore potentially be accomplished by getting this foundation right from the start.

How should schools identify and monitor underachieving students?

Seven systematic reviews were identified in our umbrella review that were relevant to screening and progress monitoring of underachieving students in secondary schools. All but one focused specifically on CBM and 5 were focused specifically on reading. Collectively they offer important findings about screening and progress monitoring and point to some practical implications for secondary schools.

There was a set of clear findings about how students should be screened for poor reading that have implications for secondary schools. One compelling finding relates to the CBM-R assessment for oral reading fluency. Previous research had suggested that maze assessments would be the most robust screening instrument and a better predictor of general reading achievement for older students. However, the meta-analyses in our study showed that oral reading fluency was favoured for use with older students (g = 0.61, Shin and McMaster 2019). It was affirmed by Shin and McMaster (2019) that the ORF is a quick, reliable and valid screening instrument to use with older students, taking only a few minutes to implement with each student. While this focus on ORF could be interpreted as only examining one component of reading (fluency), it is important to emphasise that this is not the case. Fluency is a robust and valid means of screening for general reading achievement in older students. The assessment of fluency is indicative of the cumulative acquisition of lower-order reading skills, such as phonemic awareness and decoding, letter knowledge, knowledge of the alphabetic principles and concepts of print (Tunmer and Hoover 2019), as well as vocabulary, all of which underpin fluency which, in turn, underpins comprehension.

Using the ORF could be a highly efficient way for schools to identify those students who are the furthest behind in reading. Undertaking this assessment with all Year 7 students upon entry, for example, would enable schools to quickly see the spread of achievement in reading. School clusters, such as regions within Departments of Education, could also use the data to develop local norms, as suggested by Kilgus et al. (2014). This systematic review offered important findings about how this assessment could be used to determine decision rules around who requires Tier 2 reading intervention. Kilgus et al. (2014) observed that this could be achieved by developing locally derived cut scores, such as a student's percentile rank amongst their cohort of peers, or by working with school districts to determine local cut scores with diagnostic accuracy.

Using that information, schools could then select which students may require further time-intensive diagnostic assessments of proficiency in the lower-order skills of phonemic awareness, decoding and vocabulary. Such assessments include the Phonological Awareness Screening Test (PAST, Kilpatrick 2003), the Neale Analysis of Reading Ability (Neale 1999) or the Test of Word Reading Efficiency (Torgersen et al. 1999). Using the ORF to identify poor readers, in conjunction with reliable and valid diagnostic assessments, would yield clear data to drive decisions about the implementation of appropriate interventions and close achievement gaps for this vulnerable cohort. It would also make further assessments to inform a problem-solving approach more feasible by minimising the number of students likely to need them. Using this approach can be used to pinpoint who needs support, so that it can be offered quickly.

While fluency is (or should be) actively taught in primary schools, secondary school students are generally assumed to have mastered these skills and instruction is through reading rather than in reading. However, given that NAPLAN data indicate that there are a significant number of secondary students who do not have this firm foundation in place, the utility of quick curriculum-based screens early, on their arrival at the school, is clear. The information gleaned can inform the allocation of students for further assessment and intervention, without delay.

CBM across reading, writing and mathematics were found to be highly effective for intensifying instruction for Tier 3 students, using data-based individualisation. Jung et al. (2018) found that when teachers were provided with regular professional learning to use CBM data in this way, it had a substantial and important impact on reading, writing and mathematics outcomes for students. This impact was even more marked when teachers, in collaborative groups, were provided with regular professional learning designed to improve their knowledge, beliefs and skills in using data to individualise instruction at Tier 3 (Filderman et al. 2021). This points to the importance of schools having access to CBM and professional learning in using these assessments both in terms of data literacy as well as to individualise instruction for the students who are the furthest behind their peers.

How should schools intervene in reading, writing and mathematics?

The set of systematic reviews identified as relevant to academic interventions in reading, writing and mathematics was the largest set of reviews in our umbrella review. Collectively, they point to valuable practices that secondary schools can use to reduce achievement gaps.

Reading interventions

The systematic reviews of reading interventions constituted over half of the full set of reviews included in the umbrella review. As a set, they provided evidence that offering reading interventions to older underachieving readers is moderately effective, with a general estimate of effect being 0.49 (Scammacca et al. 2015). In some of the systematic reviews it was noted that younger students make more gains in intervention than older students (Flynn et al. 2012; Wanzek et al. 2013), but there are 2 important points to be made in relation to this. One is that older students who are not underachieving in reading also make lower annual gains in reading; that is, progress slows for all students as they get older (Scammacca et al. 2007). This means that the appropriate comparator for underachieving students receiving intervention should be the rate of progress for older students who are not receiving intervention, rather than younger students who are. The second point is that even small gains can make an important and valuable difference in reading outcomes for older students.

Not all reading interventions were found to be equally effective for older students underachieving in reading. With regard to improvements in general, particular instructional strategies were identified as highly impactful when used in Tier 2 or 3 interventions. Notably, these interventions were all multicomponent and included explicit training in cognitive strategies such as SRSD, as well as commercial programmes, such as *MultiLit* or *Corrective Reading,*[44] that offer multicomponent interventions across a range of basic reading skills.

44 Further reading: Joseph and Schisler 2009.

Within the systematic reviews of reading interventions, various instructional strategies that were associated with positive reading intervention outcomes were differentially spread across the 5 components of reading: phonemic awareness, phonics, vocabulary, fluency and comprehension. In very few systematic reviews was the focus on intervening to improve the lower-order reading processes (basic skills) of phonemic awareness (and decoding or phonics) for older students, although morphological instruction was indicated as being significantly impactful in improving phonemic awareness and decoding for older underachieving readers (Goodwin and Ahn 2013). The systematic reviews included in our umbrella review indicated that morphological instruction for students to provide them with knowledge about the roots and affixes of words to analyse the morphological structure of unknown words supported their acquisition of these basic skills.

While stand-alone vocabulary interventions were scarce (Kuder 2017), it was noted that most multicomponent reading interventions included vocabulary instruction (Scammacca et al. 2015). Instructional strategies shown to positively impact on students' vocabulary skills are listed below.

- Explicit instruction. This involved the structured and systematic breaking down of the meaning of specific vocabulary into small chunks or steps, providing scaffolds, modelling and opportunities to practise, and frequently checking for understanding (Kaldenberg et al. 2015; Kuder 2017). Explicit or direct instruction was also used in the instructional practices below.
- Morphemic analysis. This involved explicitly teaching students to predict the meaning of words based on their structural components (Kuder 2017), often in 4 steps: breaking words into their morphemic parts (such as prefix, suffix, root), attaching meaning to each word part, making a prediction about the meaning of the unknown word based on the meaning of the parts, and checking the dictionary for the definition.
- Mnemonic instruction. Here, strategies such as acronyms, rhymes or pictures were explicitly taught to assist students with memory retrieval of vocabulary content (Kaldenberg et al. 2015; Kuder 2017).
- Multimedia augmented instruction. This involved the creation of content acquisition podcasts using the universal design of learning (UDL)[45] principles to teach vocabulary.

A finding from 2 systematic reviews of reading intervention studies spanning 3 decades (1980 to 2011) was that of all types of reading interventions offered to older underachieving readers, vocabulary interventions achieved the highest effect sizes on all reading outcome measures (Scammacca et al. 2007; Scammacca et al. 2015). This finding was consistent across specific subjects, such as science, and for diverse populations of students, including ELs and students with cognitive disabilities (Browder et al. 2006; Kaldenberg et al. 2015; Oxley and de Cat 2021). It points to an effective use of resources within high schools to support students underachieving in reading, in terms of returns in improved outcomes for the investment of time and funds.

Several of the systematic reviews included findings about the impact of interventions on reading fluency outcomes. What proved to be effective was intervening in very targeted and precise ways that addressed reading fluency directly. There was support for the use of repeated reading interventions, a well-established strategy to target and boost fluency in younger readers (Joseph et al. 2009). Although there was a lack of clear evidence for the effectiveness of stand-alone, repeated reading interventions for improving reading fluency for secondary underachieving readers (Steinle et al. 2022), repeated reading interventions were more effective when combined with additional intervention components such as explicit instruction and modelling (Joseph et al. 2009; Lee and Yoon 2017; Steinle et al. 2022; Wexler et al. 2008).

There were no clear findings regarding whether setting a specified number of times to read repeatedly, or reaching a certain criterion, was more beneficial. However, it seemed that re-reading text 3 or 4 times may have as much benefit as re-reading 7 times, to increase students' reading rate on practised passages (Wexler et al. 2008). Interestingly, positive reading fluency outcomes were the largest for adolescent ELs. This finding suggests that this group of students may continue to benefit from fluency instruction even after their English-only counterparts reach the ceiling effect (Huddle et al. 2017).

45 For more information about Content Acquisition Podcasts, see https://tedcec.org/sites/default/files/2020-12/Content_Acquisition_Podcasts.pdf

Interestingly, morphological instruction was found to be ineffective for improving reading fluency. A proposed explanation was that these interventions, many of which involve teaching word study strategies apart from text, make it more difficult for participants to apply word-level supports within connected text (Goodwin and Ahn 2013). Similarly, general multicomponent reading interventions using broad and indirect approaches to improve reading without specifically targeting fluency, even those that were conducted for an extended period, (≥ 75 sessions), were not effective for improving reading fluency for older underachieving readers (Flynn et al. 2012; Wanzek et al. 2013). This highlights the importance of teaching being intentionally explicit and targeted.

The finding that fluency can be significantly improved in older students underachieving in reading by using precise and targeted strategies is important, particularly considering the findings about oral fluency being a key predictor on state tests, as discussed in the introduction to this review and in the results examining universal screening and progress monitoring through CBM. This finding, in conjunction with the finding that vocabulary can also be significantly improved for older underachieving readers, indicates that, provided students have mastered the lower-order skills of decoding and phonemic awareness, with effective intervention they can achieve significant progress in the middle to higher-order key components of reading that ultimately support skilled comprehension.

The largest group of systematic reviews within the reading group of reviews focused on reading comprehension. Interventions that specifically targeted reading comprehension skills variously produced moderate to strong effects (Berkeley et al. 2010; Filderman et al. 2020; Scammacca et al. 2007; Scammacca et al. 2015). Reading comprehension is often described as an outcome of acquiring particular skills for comprehending; it is not a skill in and of itself. In our review, we identified several teachable skills and strategies that support comprehension for older underachieving readers, such as those discussed above. There was some evidence to suggest that decoding or word study interventions can improve reading comprehension skills for older underachieving readers (Edmonds et al. 2009; Flynn et al. 2012; Joseph et al. 2009; Scammacca et al. 2007). Interestingly, fluency interventions were not found to improve reading comprehension directly (Edmonds et al. 2009; Scammacca et al. 2007; Steinle et al. 2022; Wexler et al. 2008). These findings suggest that for students who lack word reading skills, it is necessary to build these word-level skills while teaching other comprehension skills (for example, vocabulary), so that access to increasingly difficult levels of print is available to them (Edmonds et al. 2019).

Several instructional strategies were shown to have positive impacts on students' reading comprehension skills. The 2 most impactful instructional practices were explicit instruction and strategy instruction. Explicit instruction involved precision teaching of fundamental reading skills (Berkeley et al. 2010; Flynn et al. 2012; Joseph et al. 2009), the linguistic structure and organisation of expository text (Hall-Mills et al. 2022), the definitions of science vocabulary words (Kaldenberg et al. 2015), and reading comprehension skills and strategies such as previewing, clarifying, generating questions and summarising (Edmonds et al. 2009; El Zein et al. 2014; Burke et al. 2015). Strategy instruction involved the explicit teaching of meta-cognitive (for example, students actively reflecting on their thinking) or cognitive processes, such as those used to solve problems, regulate attention, organise thoughts and materials, and monitor one's own thinking. There was wide variation in the application of strategy instruction across the systematic reviews. A description and some examples of these are provided below.

- **SRSD** instructional sequence involves self-questioning, comprehension strategy training, self-regulated comprehension and explicit writing frameworks (Kang et al. 2015).
- **Main idea or summarisation** is typically explicit teaching of instructional practices for reading text and identifying the most critical information (such as, main idea) and then how to link these main ideas across paragraphs to create summaries (Berkeley et al. 2010; Daniel et al. 2021; Filderman et al. 2022; Kim et al. 2012; Solis et al. 2012; Stevens et al. 2019).
- **Questioning** involves direct questioning of students (with corrective feedback). For example, after each sentence, questions are asked to illicit reflection and clarify meaning, such as: 'Why does that make sense?' Alternatively, it involves modelling the process of thinking aloud, showing how to self-question and reflect during and after reading, for example, 'Who or what is the paragraph about?' and 'What is happening to them?' (Berkeley et al. 2010; Burke et al. 2015; El Zein et al. 2014; Solis et al. 2012).

- **Targeting underlying structures** involves teaching students how to identify structures of different types of passages (main idea, list and order passages) and organise the passages (Kim et al. 2012).
- **Background knowledge** is explicit teaching of information needed to understand the text, including building students' content knowledge (for example, science, social studies), or vocabulary knowledge prior to, or during, reading and strategy instruction (Filderman et al. 2022).
- **Learning strategy instruction** involves teaching students the skills to manage their learning. For example, emphasis on effort, accessing peer and parental support, managing the learning environment (for example, regulating study and break time, dealing with distractions), motivational aspects (such as self-efficacy, task value, goal orientation) and metacognitive knowledge (such as, teachers model different strategies and explain their rationale to the students; Donker et al. 2014)

Other instructional strategies shown to have positive impacts on students' reading comprehension skills, included the below.

- DBI: a systematic data-based approach to intensifying instruction. For students needing Tier 2 or 3 intervention, this entails adjusting instruction to achieve a long-term goal in skill or knowledge acquisition and using data-based rules to determine which instructional changes are needed and when they should be implemented or adjusted. DBI incorporates frequent progress monitoring (Jung et al. 2018; Dietrichson et al. 2020).
- Graphic organisers: simple visual and spatial displays in which lines, arrows and spatial arrangements are used to describe text content, structure and key conceptual relationships. Examples include semantic feature analysis, cognitive maps, story maps and Venn diagrams (Filderman et al. 2022; Kim et al. 2004; Kim et al. 2012; Solis et al. 2012).
- Semantic maps (a specific type of graphic organiser): the relationship between the words and content are highlighted. Semantic maps can be beneficial for supporting students' comprehension of science texts (Kaldenberg et al. 2015; Solis et al. 2012).

There was limited evidence regarding the maintenance and generalisation of gains following reading interventions. It was noted that the effect sizes for some recent reading intervention studies were smaller than those shown in earlier studies (Scammacca et al. 2015). A possible contributing factor to this phenomenon may be that more rigorous measures of reading outcomes have been included in recent research and give a better indication of the extent to which skills gained through the interventions generalise beyond the immediate context of the intervention (Scammacca et al. 2015). Therefore, this may be an encouraging finding in terms of providing evidence for the effectiveness of reading interventions for improving broader academic outcomes for underachieving secondary readers. Likewise, there were some findings to suggest that improvements in reading comprehension skills are maintained beyond the conclusion of the intervention (Daniel et al. 2021; Filderman et al. 2022; Hall-Mills and Marante 2022; Kim et al. 2004). Typically, these findings were derived from measures taken only a few weeks or months after the intervention. No evidence was found to show that gains were maintained for longer periods. This was identified as a gap in the research.

There were very few systematic reviews in our umbrella review in which the settings, the duration or the frequency of interventions were reported as moderating the effects of reading interventions. Findings regarding settings all related to the size of instructional groups. For reading outcomes, the findings were mixed. There was some evidence that intervention in small groups targeting general academic skills improved reading outcomes (Dietrichson et al. 2020), and that there were greater improvements in reading skills for students in smaller instructional groups than for students in larger groups (Hall and Burns 2018; Kim et al. 2012; Stevens et al. 2019). In contrast, evidence from another systematic review showed that interventions delivered to groups of more than 11 students produced higher magnitude effects than smaller groups (Swanson et al. 2014). Other reviewers found that group size did not moderate reading intervention outcomes (Berkeley et al. 2010; Goodwin and Ahn 2013; Stevens et al. 2019; Wanzek et al. 2013).

The few findings regarding the moderating effect of the duration of interventions were all derived from reading comprehension studies. Findings from 2 systematic reviews provided minimal evidence that interventions lasting longer than one week but less than one month (Berkeley et al. 2010), and interventions providing more than 20 hours of instruction (Goodwin and Ahn 2013) were most effective. There were no clear findings about duration as a moderator on reading comprehension outcomes from other systematic reviews. There was no clear evidence to show that the frequency (number of intervention sessions per week) moderated the effects of reading interventions (for example, Stevens et al. 2019). In summary, there were no clear findings from our umbrella review regarding the optimal setting, duration or frequency of sessions for reading interventions.

Across the systematic reviews, reading interventions were found to benefit students with and without disability, indicating that effective instructional strategies offered through Tier 1, and intensified through Tier 2 and 3 interventions, can support all underachieving students. This demonstrates the value of a multi-tiered approach to intervention using intensification as required, rather than using a categorical approach to target students' personal characteristics.

Writing interventions

The 21% of systematic reviews in which the impact of intervention on older, underachieving students' writing outcomes were examined represented a large sample of participants; hence, the findings can be considered robust. In general, there were positive outcomes for older underachieving students receiving writing intervention. The findings from the reviews generally revealed that effective intervention can improve writing achievement outcomes for high-school students. The importance and value in intervening were highlighted in relation to writing skills in general, as well as for particular aspects of writing. Various strategies were identified to be associated with positive intervention outcomes. Improvements in students' performance with writing fluency, the production of texts and spelling were noteworthy.

Writing fluency involves proficiency in the following underpinning skills: (1) transcription (handwriting, typing, spelling); (2) text generation (phrases, sentences and compositions) with ease and automaticity; and (3) executive functions (coordinating transcription and text generation through planning, translating and revising). Only 2 systematic reviews reported outcomes for writing fluency (Datchuk et al. 2022; Datchuk et al. 2020). For students with low proficiencies in these underpinning skills, acquisition interventions were effective. These interventions featured using untimed writing tasks, explicit teaching through modelling, prompting, initial practice and error correction, and using graphic organisers for text generation (Datchuk et al. 2022).

For students who had acquired the underpinning skills of transcription, text generation and executive functions, but who needed support to execute them with ease and automaticity, writing fluency interventions were effective. Instructional practices among these fluency interventions included timed writing practice, performance feedback, goal setting and graphing of student performances (Datchuk et al. 2022). Effective fluency interventions were also combined with explicit instruction or SRSD (Datchuk et al. 2022; Datchuk et al. 2020). It was noted that rather than an immediate positive change (indeed, some initial changes were negative), student improvements in writing fluency were gradual. Evidence for the most effective writing fluency interventions was mixed, with the largest gains reported for acquisition and fluency interventions (Datchuk et al. 2022). Similar gains across writing strategies (direct instruction and SRSD) were reported in another systematic review (Datchuk et al. 2020). Of these latter interventions, direct instruction tended to last for a shorter duration than other interventions (for example, SRSD), indicating a possible advantage in terms of efficiency (Datchuk et al. 2020).

Findings pertaining to effective text writing interventions were found in systematic reviews of studies specifically targeting the production of connected text (Boon et al. 2018; Datchuk et al. 2022; Ciullo and Reutebuch 2013) and also in systematic reviews of general writing or academic interventions (Gillespie and Graham 2014; Kang et al. 2015; Peterson et al. 2020; Valasa et al. 2014). The studies included in these systematic reviews featured many types of writing interventions.

Effective interventions identified across the reviewed studies included the following:

- Explicit or systematic instruction: this involved the modelling of strategies for planning, writing, revising and/or editing text, together with student practice of the strategies. (Gillespie and Graham 2014; Kang et al. 2015; Peterson et al. 2020; Valasa et al. 2014). Explicit instruction, coupled with other instructional strategies, was also included in other interventions (for example, learning strategy instruction and graphic organisers – see below).
- Learning strategy instruction: this involved explicitly teaching students strategies designed to improve self-regulated learning (including cognitive, metacognitive and management strategy skills), motivational aspects and metacognitive knowledge (see Reading discussion above). These strategies were typically taught through SRSD, that is, the explicit teaching of writing strategies and the development of self-regulation skills, as well as the Strategic Instruction Model, that is, explicit or direct instruction of writing strategies (Asaro-Saddler et al. 2021; Donker et al. 2014; Gillespie et al. 2014; Kaldenberg et al. 2016; Kang et al. 2015; Valasa et al. 2014).
- Graphic organisers: this involved using computer-based graphic organisers, paper-and-pencil graphic organisers, and teacher- and student-generated graphic organisers to plan ideas before writing (Datchuk et al. 2022). There was clear evidence to suggest that middle and secondary students with LD who use technology-based graphic organisers or concept maps to comprehend, improve writing or acquire content, must also receive explicit instruction, guided practice and feedback to support learning, and to enable proficient use of the software (Boon et al. 2018; Ciullo and Reutebuch 2013). There was limited evidence to suggest that using graphic organisers may be more effective for writing expository texts than for writing narrative texts (Boon et al. 2018).

Findings from systematic reviews that examined the impact of interventions on spelling outcomes indicated that school-based interventions, especially those delivered in small groups, are effective for older students (Dietrichson et al. 2020). Specific strategies identified as effective included morphological instruction that involved students using morphological features to assist them to deal with the complex spelling patterns and exceptions to grapheme–phoneme correspondences (Goodwin and Ahn 2013) that are often encountered in middle and secondary school texts. There was less evidence to support the effectiveness of systematic study procedures that emphasised repeated practice with words and self-correction of misspelled words, for example, cover-copy-compare (Williams et al. 2018). Similarly, there was limited evidence for the effectiveness of technology-based interventions that involved students learning to use word prediction software (for example, Co:Writer) to reduce the number of misspellings in written compositions (Williams et al. 2018). In a systematic review of the impact of reading interventions on writing outcomes, a small positive effect on spelling outcomes was reported (Wanzek et al. 2013).

A lack of evidence regarding the maintenance and generalisation of gains following writing interventions was reported; this was identified as a gap in the research. There was minimal evidence to show that strategies taught during intervention generalised across writing domains (Peterson et al. 2020) or that improvements in writing were maintained (Graham et al. 2018; Williams et al. 2018). In 2 systematic reviews, it was noted that SRSD holds the most promise as an effective writing intervention and that employing SRSD is also most likely to promote maintenance of writing skills (Valasa et al. 2014). More research is clearly needed to identify whether skills learnt within writing interventions can be generalised to other contexts and maintained over time.

There were very few systematic reviews in our umbrella review in which the settings, the duration or the frequency of interventions were considered as variables or reported as moderators of writing outcomes. Findings regarding settings all related to instructional group sizes. It was reported in one systematic review that school-based interventions delivered in small groups resulted in statistically significant improvements in reading and writing outcomes (Dietrichson et al. 2020). However, in this systematic review, small-groups intervention delivery was not compared with other settings so group size cannot be considered a moderating factor. There was inconclusive evidence that increasing the duration of interventions improved writing performances in terms of increasing the number of words that students wrote each session (Asaro-Saddler et al. 2021). No evidence for the moderating effect of frequency of intervention sessions on writing outcomes was found. In summary, there was no conclusive evidence that intervention setting, intervention duration or the frequency of intervention sessions moderated the outcomes of writing interventions.

The findings from our review of interventions for older students underachieving in writing indicate that appropriately targeted writing interventions can improve writing outcomes for students with and without disabilities. The importance of determining the specific writing skills in which students lack proficiency prior to intervening cannot be overemphasised. While not explicitly stated by the authors of the systematic reviews, it seems that for these underachieving students, the administration of accurate and timely screening measures within a multi-tiered support framework is vital if writing interventions are to be appropriately targeted.

Mathematics interventions

While only a fifth of the systematic reviews included in our umbrella review focused on mathematics interventions for older underachieving students, they collectively represented a large sample of participants. The findings can be considered relatively robust. The findings from the reviews generally revealed that interventions for mathematics for older students can be impactful and worthwhile. The effectiveness of intervening in various aspects of the mathematics curriculum was evident. Improvements in students' performance with fractions, algebra and word problem solving were noteworthy. Various strategies were identified to be associated with positive intervention outcomes.

- Explicit and systematic instruction was identified as beneficial for improving mathematical skills overall, as well as for skills in algebra, fractions and arithmetic (Bouck et al. 2018; Lee et al. 2020; Jitendra et al. 2018; Maccini et al. 2007; Shin and Bryant 2015; Watt et al. 2016). Word problem solving skills were improved through the explicit teaching of strategies for the identification of underpinning mathematical structure and relationships (Lein et al. 2020; Shin et al. 2021; Xin and Jitendra 1999) as well as cognitive strategy instruction (Donker et al. 2014).
- Interventions using visual models or physical representations of mathematical concepts were found to enhance outcomes for mathematics in general, as well as for fractions and word problem solving in particular (Maccini et al. 2007; Shin and Bryant 2015; Shin et al. 2021; Xin and Jitendra 1999). When multiple representations of mathematical concepts were employed in interventions, algebra skills were improved (Jitendra et al. 2018; Lee et al. 2020; Watt et al. 2016).
- Guided and independent practice were beneficial for improving overall mathematics performance, as well as improving skills associated with algebra, word problem solving and arithmetic (Bouck et al. 2018; Maccini et al. 2007; Shin et al. 2021).
- When mathematics examples were presented to students in a logical sequence, algebra skills were found to improve (Lee et al. 2020; Watt et al. 2016). Increased monitoring of students' performance by teachers as well as more frequent provision of corrective feedback were found to improve students' overall mathematics performance, as well as enhancing their skills associated with fractions and arithmetic (Bouck et al. 2018; Maccini et al. 2007; Shin et al. 2015).
- The CRA approach was associated with improvements in mathematics in general, in algebra and arithmetic skills, as well as for knowledge of basic number facts when embedded within interventions (Bouck et al. 2018; Lee et al. 2020; Maccini et al. 2007; Watt et al. 2016).
- Schema-based instruction (SBI) proved beneficial in enhancing skills associated with word problem solving (Maccini et al. 2007; Lein et al. 2020).
- CAI strategies of using tutorial or interactive programs were found beneficial for improving word problem solving skills (Shin et al. 2021; Xin and Jitendra 1999).
- There was limited evidence that mnemonic strategy instruction improved word problem solving skills (Maccini et al. 2007), or that contextualised instruction (using real-world problems) improved overall mathematics and word problem solving skills (Jitendra et al. 2018; Maccini et al. 2007).

A lack of evidence regarding the maintenance and generalisation of gains following mathematics interventions was noted in our umbrella review and we identify this as a major gap in the research. Some evidence for the generalisability of mathematics skills of low to moderate effect sizes was found in studies using standardised or norm-referenced tests (Dietrichson et al. 2020; Lein et al. 2020). In only one systematic review, a large average effect size for generalisation measures was reported (Xin and Jitendra 1999). In the few systematic reviews that included maintenance measures, the average effect sizes varied between small (Dietrichson et al. 2020) and moderate to large (Xin and Jitendra 1999).

There were very few systematic reviews in our umbrella review in which the settings, the duration or the frequency of interventions were considered as variables or reported as moderators of mathematics outcomes. Group size, a setting moderator, was considered in 3 reviews (Jitendra et al. 2018; Lein et al. 2020; Stevens et al. 2008) and was not found to significantly moderate differences in mathematics performance outcomes. There was inconclusive evidence that increasing the duration of interventions improved performance outcomes (Jitendra et al. 2018; Stevens et al. 2018; Xin and Jitendra 1999). Variations in the duration of effective interventions were identified: more than 10 hours (Jitendra et al. 2018), more than 15 hours (Stevens et al. 2018) and more than one month (Xin and Jitendra 1999). In one review (Xin and Jitendra 1999), it was reported that short-term (≤ 7 sessions) interventions were at least as effective as intermediate-term (> 7 sessions but ≤ 1 month) interventions, but less effective than long-term (> 1 month) interventions. The effectiveness of the frequency of mathematics interventions was not found to moderate intervention effects in any of the systematic reviews included in our umbrella review. In summary, only group size was found to be a positive moderator on the outcomes of mathematics interventions.

Collectively, there was evidence that the principles of effective mathematical intervention for elementary students (for example, Gersten et al. 2009) were also effective for middle and secondary school students. Notable examples of interventions were modelling that included visual models (Jitendra et al. 2018), guided practice, independent practice, the monitoring of student performance and/or corrective feedback (Maccini et al. 2007), as well as the adoption of the CRA approach (Bouck et al. 2018). Stevens et al. (2018) suggested that older students may require interventions of greater intensity and duration than younger students to make substantial improvements in mathematics performance.

We did not always find that the relevance to a multi-tiered support framework was explicitly stated by the authors of the systematic reviews. However, the findings from our review of interventions for older students underachieving in mathematics are particularly relevant to the core concept of intensification within multi-tiered frameworks. That is, the strategies identified as beneficial all constitute effective practices for use in core mathematics instruction at Tier 1. Interestingly, the mathematical content of the interventions was anchored in the general mathematics curriculum. An implication of the findings related to strategy and mathematical content is that the same instructional strategies are appropriate for intervention across all tiers of support, with greater intensity and in smaller group settings at Tiers 2 and 3.

It was notable that across the systematic reviews, students with and without disabilities were included. This suggests that all students can learn and achieve if provided with sufficient quality of instruction by appropriately qualified (content knowledge and pedagogical content knowledge) teachers of mathematics employing strategies that should already be within their skillset.

How can schools support intervention implementation to achieve optimal outcomes for students?

Our research found that schools can achieve optimal outcomes for students through interventions by adopting a whole-school approach ensuring that Tiers 2 and 3 are based on a strong foundation at Tier 1, and that there are effective models of collaboration in place to ensure that support at higher tiers is appropriately targeted.

Both are cornerstones of multi-tiered frameworks for service delivery. Burns and Symington (2002) looked at the benefits achieved when schools supported multidisciplinary problem-solving teams to consult, observe and confer in determining which students are not making adequate progress in Tier 1 and what the most appropriate supports would be to address their area/s of need. The meta-analyses found that the benefits of such an approach were many, including improved student outcomes, fewer students needing resource-intensive support at higher tiers and reduced rates of special education placements. Burns and Symington (2002) noted that such an approach required obtaining staff buy-in, ensuring staff are appropriately skilled in working with data, that all team members are provided with time to collaborate built into their job responsibilities, that useful interventions strategies are ready to be rolled out and that all team members have been upskilled in the problem-solving approach. These are all important considerations for schools to build into their multi-tiered frameworks, as interventions are offered within a system of supports that requires a whole-school approach and a skilled team to achieve this. In addition, Burns et al. (2005) found that optimal outcomes for interventions at Tiers 2 and 3 were obtained when a strong foundation of Tier 1 was in place. They found that this was sufficient to meet the needs of approximately 80% of students and reserved more intensive support for the fewest students.

Beyond the meta-analyses of implementing multi-tiered supports to students, and associated collaboration of problem-solving teams to achieve the implementation of these supports, there were some additional but somewhat limited findings regarding supports for the implementation of intervention to achieve optimal outcomes for students. One is that schools can best support intervention implementation through the appointment of personnel who are well-equipped to deliver interventions and through the judicious use of resources supported by evidence. Of note here is that 4 systematic reviews examined moderators for the quality and fidelity of intervention implementation (Fallon et al. 2015; Filderman et al. 2021; Jung et al. 2018; Scammacca et al. 2007). These offer important implications for how schools should support quality implementation.

The first implication relates to the importance of investing in quality professional learning supported by evidence. In schools and school systems shifting towards a multi-tiered framework for delivering interventions, in which data-based instruction is a key component, a critical component for success is the data literacy of teachers. These systematic reviews showed that professional learning, collaboration and individual feedback strongly improved teachers' data literacy skills and knowledge (Jung et al. 2018) as well as improving the fidelity of intervention implementation (Fallon et al. 2015). The impact on teachers' beliefs in the value of using data was more moderate. However, while this suggests that these beliefs may be somewhat less malleable than skills, it does suggest that this can contribute to getting buy-in, which is essential when it comes to long-term improvements in practice (Filderman et al. 2021).

These studies also showed that when teachers were provided with this professional learning, improved academic outcomes arising from interventions were achieved. Schools and school systems moving to a multi-tiered approach should therefore consider investing resources in such supports. Such support systems might be challenging to initiate though, as they require individuals with DBI-related expertise and/or time to act as coaches and to professionally collaborate with colleagues to meet and share data. Ongoing partnerships between schools and researchers may be one promising way to address this challenge and to establish sustainable systems of successful DBI implementation (Jung et al. 2018).

A further important finding pertained to resources to support screening, progress monitoring and data-based individualisation. Specifically, the lack of evidence of technical adequacy for teacher-generated CBM tools in DBI raises questions about their appropriateness for use in data-based decision-making and data-based instruction (Filderman et al. 2021; Jung et al. 2018). Multiple such applications are available and have been rigorously evaluated by the National Center for Intensive Intervention[46] and may enhance the feasibility of administration, scoring and data management (Jung et al. 2018). This suggests that a far better investment by schools would be in acquiring commercial and robust CBM and investing in professional learning, rather than investing teachers' time in creating bespoke but potentially unreliable assessments for collecting data when robust tools are already available.

Research gaps/future research

One research gap that we identified arises from the obvious imbalance of systematic reviews across the domains of reading, writing and mathematics. More than half of the systematic reviews that we included in our umbrella review focused on reading, while there were fewer systematic reviews in which mathematics and writing were investigated. There clearly is a need for more high-quality systematic reviews with a focus on investigating effective interventions across these domains. Particular attention should also be paid to the quality and fidelity of intervention implementation, given that these are essential components of MTSS yet were frequently noted as missing in intervention research by the systematic reviewers (for example, Burns et al. 2005; Shin et al. 2021). Experimental studies involving large RCTs provide the most robust evidence. These research designs were in the minority within the systematic reviews in our sample. More research using these designs that include consideration of implementation fidelity would be important additions to the existing data (Goodwin and Ahn 2013; Peterson et al. 2020).

Within the reading literature, we identified a particular gap regarding effective intervention for older students to teach fundamental lower-order aspects of reading typically taught at primary school, such as decoding skills (for example, Browder et al. 2006) and reading fluency (for example, Wexler et al. 2008). Research is needed to examine reading interventions that target only one type of comprehension strategy instruction in isolation (for example, main idea) rather than multiple related strategies (for example, main idea and retell) to clearly demonstrate which strategies are most effective (Filderman et al. 2022). In addition, further research is also needed to establish evidence-based practices for groups of adolescent students with reading skill deficits (for example, Garwood et al. 2014). This research is especially needed to understand how effective these interventions are for students with complex learning profiles, such as intellectual disabilities, or those with complex communication support requirements, and how evidence-based instruction may need to be adjusted for accessibility and impact (for example, Browder et al. 2006).

Another research gap is in the area of writing. High quality writing research investigations are needed on comprehensive writing programs for larger (Ciullo and Reutebuch 2013) and more diverse (for example, Datchuk et al. 2022) populations of students (including students with disability) and to identify the impact of various reading treatments on writing outcomes (Graham et al. 2018). Given the significant number of students in Australia with poor results in NAPLAN for writing (ACARA 2021), we suggest that more research and systematic reviews within this domain is particularly urgent, although further research across all domains is needed.

High-quality studies are also required to examine aspects of effective mathematics interventions, including multicomponent interventions, and those that target specific skills (Stevens et al. 2018) including technology-based mathematics interventions (Stevens et al. 2018), the use of manipulatives (Bouck et al. 2018) and difficult mathematics topics (Shin et al. 2021) for older students. Another much-needed research focus is on better understanding both the impact and the role played by particular instructional approaches for mathematics interventions.

46 See, for example, Acadience (https://acadiencelearning.org/), aimswebPlus (www.aimsweb.com), easyCBM (www.easycbm.com), DIBELS (https://dibels.uoregon.edu/) and FAST (www.fastbridge.org).

Comparatively few systematic reviews were identified in relation to how underachieving students should be identified, and how their progress should be monitored when they receive intervention. This gap was noted by researchers (Ardoin et al. 2013; Hall-Mills et al. 2022). More research is required on proximal screening measures for which systematic reviews relating to reading fluency only were identified (Kilgus et al. 2014; Shin and McMaster 2019). Further research involving distal measures of student achievement following intervention is also required (for example, Daniel et al. 2020; Dietrichson et al. 2020).

There is a need for more research on the utilisation of measures such as standardised tests (for example, Stevens et al. 2018), as indicators of the generalisation of intervention gains (for example, Peterson et al. 2020; Valasa et al. 2014). Given the lack of evidence identified through systematic reviews, we recommend schools select from the CBM that have been assessed as robust from the Academic Screening Tools Chart[47], and the most appropriate progress monitoring measures from the Academic Progress Monitoring Tools Chart[48] available through the National Center for Intensive Intervention technical assistance centre. These are based on the US curriculum; however the basic skills reading, writing and mathematics are common to both curricula in the US and Australia.

From our umbrella review of the systematic reviews relating to supporting underachieving students in secondary schools, several questions remain unanswered. One such unanswered question is: who should be screening students, implementing interventions or monitoring their progress? Should this be primary-trained teachers, teaching assistants or speech therapists? While researchers from the UK (for example, Sharples et al. 2016) suggest that teaching assistants should be deployed to run structured intervention, no systematic reviews to support this practice were located. Nor did we find any evidence to suggest that interventions for secondary school students be run by speech therapists, pathologists or by primary school teachers. Additionally, we did not identify any systematic reviews to answer questions about the resources, time structures or age-appropriate materials. Based on the research, key actors in the intervention process, the materials and scheduling interventions were not phenomena of interest within the intervention or MTSS literature. Accordingly, we encourage future research on the identification of who is best positioned to undertake screening, progress monitoring and intervention implementation, and what resources and scheduling considerations are required for successful outcomes.

A paucity of studies relevant to secondary schools was highlighted in our review. We focused explicitly on older students in the early years of high school, and we excluded many systematic reviews that focused solely on students in primary school. This gap in research for older students was noted in the systematic reviews that we included, and we observed that more research relevant to older students had been called for consistently over the last decade. This is a well-documented and persistent research gap. The gap in this aspect of research highlights the need for driving a clear research agenda to understand how to close achievement gaps for older students, with implications for both academics to undertake the work, as well as funding bodies to enable the work to be done.

There is a clear need for more evaluations of MTSS impact and implementation in general and in relation to secondary school students. The evaluations that were included (Burns et al. 2005; Burns and Symington 2002) were high in quality but dated, and specific findings relating to secondary schools were not delineated. More systematic reviews on the evaluation of the benefits of MTSS for systems, and especially secondary schools and students, is required. There is also a need for more high-quality systematic reviews focusing on investigating the impact of interventions on outcomes for older students underachieving in literacy and mathematics (for example, Bouck et al. 2018; Stevens et al. 2018). Other research gaps pertaining to specific aspects of literacy and mathematics interventions are outlined above.

47 Academic Screening Tools Chart. https://intensiveintervention.org/resource/academic-screening-tools-chart
48 Academic Progress Monitoring Tools Chart. https://charts.intensiveintervention.org/aprogressmonitoring?_ga=2.125223946.1848310776.1664457303-392614326.1664163453

The shift away from a categorical service delivery framework in Australian policy and legislation signals the need for more research in this area. Reforms to address equity and inclusion need to be based in sound evidence, and rigorous research and analysis methods should be used. The existing evidence is robust, but the research was focussed on addressing implementations at the whole school level, and as noted above, the research examined was somewhat dated (Burns et al. 2005; Burns and Symington 2002). Accordingly, more current evidence is required to validate the findings for contemporary Australian schools. These data are vital to guide teachers and administrators in driving efficient and effective change to multi-tiered frameworks (Barrio et al. 2015). Finally, the systematic reviews that were examined did not account for the quality of Tier 1. This is such a central element in the success of tiered interventions that these practices should be examined in future research.

Given that there are some signs that contemporary, multi-tiered systems of support are being introduced, such as those noted earlier from Victorian and NSW within Education Department guidance documentation, there is a clear need for high-quality, up-to-date research to guide this. While there is high-quality evidence from the US providing important data regarding the impact and uptake of multi-tiered frameworks we have synthesised in our umbrella review, this research is somewhat dated and is grounded in the unique legislative context of the US. Important differences exist in the Australian legislative framework (for example, see de Bruin 2019).

Further research is required to understand how MTSS can be implemented within Australia's national and state-based policy contexts, and on the impact of MTSS on Australian systems, schools and students. We recommend examining the uptake and impact that can be derived from early-adopter schools and systems that are already underway in Australia. We also recommend the creation of research centres in Australian universities. In the US, university research centres were initiated and funded by the federal government. Ultimately, these centres were instrumental in developing and refining the early models of RTI and PBIS frameworks that were scaled up and that have become integrated within a comprehensive MTSS framework that now dominates implementation across the US. There are no such initiatives in Australia, but they are needed. Given that research centres were so successful, with many becoming scaled up over time, and now serve as national technical assistance centres, we recommend that research be funded to establish university partnerships similar to those in the US that catalysed the uptake of MTSS and remain highly valued for nationwide implementation of MTSS.

Limitations

There are several limitations on the umbrella review that should be considered. One limitation is methodological. In line with recommendations for conducting a rapid review, every effort was made to ensure that streamlining the process to meet the tight turnaround time did not compromise its integrity. We made several intentional choices to meet the constraints of conducting a rapid umbrella review, including the scaling back of methodological features of a typical broad-spectrum umbrella review. For example, we did not undertake duplication of 100% of the included set of studies in the critical appraisal and data extraction phases. Rather, we undertook a rigorous training procedure until we reached 100% inter-coder agreement and then we limited duplication to approximately 30% of the studies. This means that approximately 70% of decisions made about the quality and the coding of the systematic reviews located, as well as the extraction of the results, were only undertaken by a single member of the team. This introduced some risk of bias.

Another limitation is the applicability of the findings of the umbrella review to secondary schools. This limitation stems from the degree to which the systematic reviews aligned with the population of interest in the PICo that informed our inclusion criteria. We initially located very few systematic reviews of relevant literature conducted solely (or primarily) in secondary schools. This meant that we had to determine a new threshold for inclusion. Accordingly, we imposed a minimum 30% population threshold that was deemed to be the minimum to show relevance (see methodology section for more information). While no moderating factors identified secondary school settings as a variable for intervention outcomes, they cannot be ruled out. Secondary schools differ from primary schools in important ways including school scheduling, as well as the relative expertise of teachers. This could limit the applicability of our findings to secondary schools.

A further limitation is that we applied the rigorous JBI Global criteria (Aromataris et al. 2020) for the critical appraisal of the systematic reviews for inclusion. This ensured that only high-quality systematic reviews were included and meant the exclusion of systematic reviews that were in all other ways strongly aligned with our PICo, for example, Paul and Clark (2016) and Elleman et al. (2019). We excluded these systematic reviews because they did not meet the threshold of 10/22 scoring within the critical appraisal process (see Appendix A.7), and thus their design/reporting/methodology were deemed too weak to inform the outcomes of our umbrella review. This means some research was excluded that may have been potentially relevant to the questions that we sought to answer.

Conclusion

Australian federal legislation now provides funding to state education departments and school sectors (for example, catholic and independent school sectors) under the needs-based provisions of the *Australian Education Act 2013*. Each state and sector are thus required to disseminate those funds to schools using a needs-based framework and to move away from categorical approaches to funding and support. Accordingly, Education Departments across jurisdictions and sectors are developing new policies and guidance for schools to provide educational supports to students on the basis of need (Australian National Audit Office 2017). There are several implications arising from our umbrella review that are of relevance to how systems and schools can achieve this needs-based approach with consistency and at scale, using the evidence that is summarised in our report.

One implication arising from our umbrella review is that there is evidence to suggest that MTSS can support a needs-based approach to instruction and intervention in Australia. While this evidence is dated, it does suggest that this approach implemented in schools may have value to effectively and efficiently address the achievement gaps that persist and that disproportionately affect students from priority equity cohorts. The evidence that we reviewed shows that MTSS begins with prevention. The conclusion that may be drawn from the studies that we reviewed is that if a multi-tiered approach is embraced and the highest-quality instruction is provided to all students at Tier 1, underachievement is preventable for 80% of students, regardless of their personal circumstances or characteristics.

If timely and targeted Tier 2 intervention is provided in addition to Tier 1 and aligned to the students' needs, achievement gaps can be reduced, meaning that *95% of students can meet standards for reading, writing and mathematics*. Implementing MTSS with fidelity as early as possible, and ensuring quality instruction at Tier 1, could therefore reserve resource-intensive intervention for those students who need more sustained and individualised support. This evidence offers important suggestions for how persistent underachievement, particularly by priority equity cohorts, might be addressed in Australian schools and would be worthy of consideration by policymakers. The evidence also suggests the importance of a consistent national approach to MTSS and developing evidence-based resources for technical assistance and professional learning that is needed to implement MTSS with fidelity at scale across schools, school systems and jurisdictions.

Several implications arise from our findings about implementation and national scaling of MTSS based on how the reforms were introduced in the US. One lesson from the US experience is the value of consistency in achieving reform at scale. In the US, reforms were originally enacted by giving autonomy to the States to opt-in and to develop their own approaches to multi-tiered frameworks. This has resulted in a nationally inconsistent approach with policies across states variously adopting RTI or MTSS and using differing definitions and models. The result is a confusingly varied set of frameworks from one state to the next, inhibiting the scaling and uptake of practices. This also provides evidence that is worthy of consideration by policymakers in Australia.

If Australian school systems pursue MTSS to replace categorical models, this evidence suggests that, from the outset, clear definitions and guidance for MTSS should be used, with consistency by states to inform policy documentation and guidance for schools to enhance a nationally consistent approach. Given that the MTSS approach in Kansas, an early adopter of MTSS, has been used as a benchmark by so many US states for articulating their own policies and developing their own professional learning programs, we recommend that Australia might consider using this exemplary model[49] in MTSS to inform policy and professional learning materials.

49 See Kansas Technical Assistance Centre in Appendix A.10.

A further implication is the importance of driving the introduction of MTSS as a general education initiative that is for *all* students and not a special education initiative that is reserved for some. As found in our umbrella review, RTI was the first approach scaled across the US, and was characterised by a deficit-driven, special-education approach, ultimately emphasising remediation at Tiers 2 and 3 rather than prevention at Tier 1. As a consequence, there has not been sufficient emphasis on evidence-based instruction at Tier 1 which is the key to preventing inappropriate special education referrals and minimising the number of students needing targeted intervention. The narrow focus of RTI on academics alone has proven hard to shift in a consistent manner across jurisdictions as a minority of US states continue to maintain a narrow RTI approach focused solely on academic intervention. Further, the narrow, special-educational approach has also been hard to shift within many states as well as US initial teacher education courses, hindering the preparation of teachers for quality implementation at Tier 1. It has taken a decade for the majority of states to move towards a comprehensive MTSS with a focus on a quality Tier 1 as the best way to educate all students, to prevent achievement gaps and to create a positive learning environment for all students.

As Australian school systems phase out categorical approaches and enact a similar shift away from deficit-driven frameworks and introduce a multi-tiered framework in their place, the lesson from the US is to ensure that a comprehensive MTSS is embraced from the start. We recommend a research agenda be prioritised to accomplish this. Particular attention could be paid to generating reform in teacher preparation and professional learning, building capacity for using data in schools, fostering interprofessional teams in schools and expanding the use of evidence-based practices in instruction and intervention with an emphasis on Tier 1.

The evidence presented in our umbrella review highlights that all students can learn and achieve success with exposure to quality instruction and the right amount of it. MTSS is about determining how much is enough for each student and providing it to them. Our identification of effective instructional strategies for teaching foundational reading, writing and mathematics skills is the key to achieving this. Importantly, the findings of our umbrella review highlight that the instructional strategies that are effective at Tiers 2 and 3 for older underachieving students in secondary school are those that are also effective for use to support all students in Tier 1 (Foorman and Wanzek 2015; Jitendra and Dupuis 2016). Strategies including explicit instruction, strategy instruction, (meta)cognitive instruction, as well as teaching students to use graphic organisers and mnemonics, were identified to be effective for Tier 2 and 3 interventions in reading, writing and mathematics.

Underachieving students do not require different or special strategies. What they do require is longer duration and more frequent instruction in using these strategies. We have provided a summary table of effective and ineffective instructional strategies in relation to all of the domains of learning examined in our review (see Appendix A.9). We recommend that the effective practices identified are prioritised for embedding within initial teacher education and quality professional learning for in-service teachers and school leaders. We have also provided a list of useful resources for schools to access which offer excellent resources and professional learning materials (see Appendix A.10).

For secondary schools, there are several key implications arising from this umbrella review about how staff should be upskilled and prepared for MTSS implementation. One is the importance of targeted professional learning in using the instructional strategies and approaches that we found to be beneficial. Providing high-quality professional learning in using these instructional strategies will support *all* students, including those receiving instruction or intervention across Tiers 1, 2 and 3. This professional upskilling will also ensure that instruction is aligned across the 3 levels of support, providing the most cohesive, unified and effective support for underachieving students, while also benefiting those who are already making expected progress.

A related implication exists regarding professional learning for secondary teachers in the basic skills that underpin the more complex reading, writing and mathematics skills taught in secondary schools. Secondary school teachers may be less familiar with the curriculum content associated with basic skills as these are typically taught in primary schools. However, it is important to recognise that the Australian Curriculum works as a spiral. The more complex knowledge and skills in reading, writing and mathematics that are taught in secondary schools build on those taught with less complexity in the earlier years of schooling. That is, the secondary school curriculum is underpinned by the primary curriculum. This means that there would be value in upskilling secondary school teachers to be more familiar with breaking down these skills from complex to basic. Knowing what those foundational skills are, as well as how to teach and assess them, should enable teachers to conduct high-fidelity Tier 2 and 3 interventions for secondary school students needing the most intensive support. This is achievable with the right professional learning and coaching and our research shows that this improves student outcomes. We recommend this also as a priority for professional learning to implement MTSS in schools.

An additional implication arising from our findings is that teachers benefit from professional learning in using data. This arises from the findings regarding how teachers should be prepared to work in collaborative problem-solving teams to support MTSS. It also arises from the findings regarding teachers' quality use of robust CBM data to individualise instruction. Collaborative problem-solving teams and data-driven decision-making are cornerstones of MTSS and were found in our review to be beneficial for supporting instruction and intervention outcomes for older students. Considering this, we recommend prioritising the development and subsidising of quality professional learning that is run through such a model of prolonged contact, active learning and performance feedback for secondary school teachers in using the evidence-based practices that we identified as beneficial in our review.

Not intervening is not an option. The umbrella review presented here clearly indicates that students can be supported in ways that are highly effective. The current rate of underachievement, particularly by priority equity cohorts, cannot continue. With new models of tiered funding being rolled out into systems across the country (already in place in South Australia, Tasmania and Victoria and planned in other states, such as Queensland), it is important for schools to prepare and ensure they have dedicated and trained staff to work in data teams and offer intervention. It is equally important for school authorities to identify ways to support the implementation of a multi-tiered framework in secondary schools before time runs out for underachieving students and they transition out of the school system and formal learning. The findings presented here, and the roadmap offered by the reforms in the US, offer a potential pathway for Australian schools to address the current inequities and achievement gaps in our education system and make our schools places where all students are supported to learn and succeed.

References

*References marked with an asterisk indicate that studies were included in the umbrella review.

Adams EK, Hancock, KJ and Taylor CL (2020), 'Student achievement against national minimum standards for reading and numeracy in Years 3, 5, 7 and 9: A regression discontinuity analysis', *Australian Journal of Social Issues, 55*(3):275-301. https://doi.org/10.1002/ajs4.124

Archer, AL and Hughes, CA (2010), 'Explicit instruction: Effective and efficient teaching', Guilford Publications.

*Ardoin SP, Christ TJ, Morena LS, Cormier DC, and Klingbeil, DC (2013), 'A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules', *Journal of School Psychology, 51*(1): 1-18. https://doi.org/10.1016/j.jsp.2012.09.004

Aromataris E and Munn Z (Eds.) (2020), *JBI manual for evidence synthesis*. https://doi.org/10.46658/JBIMES-20-01

Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H and Tungpunkom P (2015), 'Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach', *JBI Evidence Implementation, 13*(3):132-140. https://doi.org/10.1097/XEB.0000000000000055

Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H and Tungpunkom, P (2020), 'Umbrella reviews', In Aromataris E and Munn Z (Eds.), *JBI manual for evidence synthesis*. https://doi.org/10.46658/JBIMES-20-11

*Asaro-Saddler K, Moeyaert M, Xu X and Yerden X (2021), 'Multilevel meta-analysis of the effectiveness of self-regulated strategy development in writing for children with ASD', *Exceptionality, 29*(2):150-166. https://doi.org/10.1080/09362835.2020.1850457

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2016), *Why NAP*. https://www.nap.edu.au/about/why-nap

Australian Curriculum, Assessment and Reporting Authority (2019), *National Assessment Program—Literacy and Numeracy achievement in reading, writing, language conventions and numeracy: National report for 2019*. https://nap.edu.au/docs/default-source/default-document-library/2019-naplan-national-report.pdf?Status=Tempandsfvrsn=2

Australian Curriculum, Assessment and Reporting Authority (2020), *Measurement framework for schooling in Australia 2020*. https://www.acara.edu.au/docs/default-source/default-document-library/measurement-framework-2020-for-pubafa92e404c94637ead88ff00003e0139.pdf?sfvrsn=1c3d4d07_0

Australian Curriculum, Assessment and Reporting Authority (2020a). *National literacy learning progressions*. https://www.ofai.edu.au/media/01nixkio/national-literacy-progressions-v3-for-publication.pdf

Australian Curriculum, Assessment and Reporting Authority (2020b). *National numeracy learning progressions*. https://www.ofai.edu.au/media/iiwbecoj/national-numeracy-progression-v3.pdf

Australian Curriculum, Assessment and Reporting Authority (2021). *National Assessment Program— Literacy and Numeracy achievement in reading, writing, language conventions and numeracy: National report for 2021*. https://nap.edu.au/docs/default-source/default-document-library/2021-naplan-national-report.pdf

Australian Curriculum, Assessment and Reporting Authority (2022). *Version 9.0. The Australian Curriculum*. https://v9.australiancurriculum.edu.au/

ACARA (2022), *General capabilities (Version 8.4)*. https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/

Australian Curriculum, Assessment and Reporting Authority (n.d.). *Literacy (Version 8.4)*. https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/literacy/

Australian Education Act (Cth) (2013). https://www.legislation.gov.au/Details/C2020C00142

Australian Education Research Organisation (2022), *Examining equity in NAPLAN achievement. Examining NAPLAN data 2013 to 2019: Methodology*. https://www.edresearch.edu.au/our-research/current-work/equity-education/examining-equity-naplan-achievement

Australian Institute of Health and Welfare (2020), *Australia's children. Cat. no. CWS 69*. https://www.aihw.gov.au/reports/children-youth/australias-children/contents/executive-summary

Australian National Audit Office (2017), *Monitoring the Impact of Australian Government School Funding (Final Report).* https://www.anao.gov.au/work/performance-audit/measuring-impact-australian-government-school-funding

Australian National Audit Office. (2021). *Monitoring the impact of government school funding—Follow-up. Auditor-General Report No.37 of 2020–21.*Department of Education, Skills and Employment, Australian National Audit Office, Commonwealth of Australia 2021. https://www.anao.gov.au/work/performance-audit/monitoring-the-impact-government-school-funding-follow-up

Bailey T (2017). 'Is MTSS the new RTI? Depends on where you live'. *Center on Multi-Tiered Systems of Suppor.* American Institutes for Research. https://mtss4success.org/blog/mtss-new-rti-depends-where-you-live

Balu R, Zhu P, Doolittle F, Schiller E, Jenkins J, and Gersten R (2015). *Evaluation of response to intervention practices for elementary school reading* [NCEE 2016-4000]. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/20164000/pdf/20164000.pdf

*Barrio BL, Lindo EJ, Combes BH, and Hovey KA (2015). 'Ten years of response to intervention: Implications for general education teacher preparation programs'. *Action in Teacher Education, 37*(2):190-204. https://doi.org/10.1080/01626620.2015.1004603

*Berkeley S, Bender WN, Peaster LG, and Saunders L (2009). 'Implementation of response to intervention: A snapshot of progress'. *Journal of Learning Disabilities, 42*(1):85-95. https://doi.org/10.1177/0022219408326214

*Berkeley S, Scanlon D, Bailey TR, Sutton JC, and Sacco DM (2020). 'A snapshot of RTI implementation a decade later: New picture, same story'. *Journal of Learning Disabilities, 53*(5):332-342. https://doi.org/10.1177/0022219420915867

*Berkeley S, Scruggs TE, and Mastropieri MA (2010). 'Reading comprehension instruction for students with learning disabilities', 1995—2006: A meta-analysis. *Remedial and Special Education, 31*(6):423-436. https://doi.org/10.1177/0741932509355988

Biancarosa C, and Snow CE (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York (2nd ed.).*Washington, DC: Alliance for Excellent Education. https://media.carnegie.org/filer_public/b7/5f/b75fba81-16cb-422d-ab59-373a6a07eb74/ccny_report_2004_reading.pdf

*Boon RT, Barbetta PM and Paal M (2018). 'The efficacy of graphic organizers on the writing outcomes of students with learning disabilities: A research synthesis of single-case studies'. *Learning Disabilities: A Multidisciplinary Journal, 23*(2):18-33. https://doi.org/10.18666/ldmj-2018-v23-i2-9042

*Bouck EC and Park J (2018). 'A systematic review of the literature on mathematics manipulatives to support students with disabilities'. *Education and Treatment of Children, 41*(1):65-106. https://doi.org/10.1353/etc.2018.0003

*Browder DM, Wakeman SY, Spooner F, Ahlgrim-Delzell L and Algozzine B (2006). 'Research on reading instruction for individuals with significant cognitive disabilities'. *Exceptional Children, 72*(4):392-408. https://doi.org/10.1177/001440290607200401

*Burke MD, Boon RT, Hatton H and Bowman-Perrott L (2015). 'Reading interventions for middle and secondary students with emotional and behavioral disorders: A quantitative review of single-case studies'. *Behavior Modification, 39*(1):43-68. https://doi.org/10.1177/0145445514547958

Burns MK (2012). 'Small-group academic interventions'. In Riley-Tillman TC, VanDerHeyden AM, and Burns MK (Eds.), *RTI Applications, Volume 1: Academic and Behavioral Interventions* 1:46-59. Guildford Press.

*Burns MK, Appleton, JJ and Stehouwer JD (2005). 'Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models'. *Journal of Psychoeducational Assessment, 23*(4):381-394. https://doi.org/10.1177/073428290502300406

*Burns MK and Symington T (2002). 'A meta-analysis of prereferral intervention teams: Student and systemic outcomes'. *Journal of School Psychology, 40*(5):437-447. https://doi.org/10.1016/S0022-4405(02)00106-1

Chard DJ (2013). 'Systems impact: Issues and trends in improving school outcomes for all learners through multitier instructional models'. *Intervention in School and Clinic, 48*(4):198-202. https://doi.org/10.1177/1053451212462876

*Ciullo S and Reutebuch C. (2013). 'Computer-based graphic organizers for students with LD: A systematic review of literature'. *Learning disabilities research and practice, 28*(4):196-210. https://doi.org/10.1111/ldrp.12017

Clemens NH, Keller-Margulis MA, Scholten T and Yoon M (2016). 'Screening assessment within a multi-tiered system of support: Current practices, advances, and next steps'. In Jimerson S, Burns M and VanDerHeyden A (Eds.), *Handbook of response to intervention (pp. 187-213).* Springer. https://doi.org/10.1007/978-1-4899-7568-3_12

*Collins AA, Lindström ER and Compton DL (2018). 'Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis'. *Journal of Learning Disabilities*, 51(2):108-123. https://doi.org/10.1177/0022219417704636

Commonwealth of Australia [Senate Standing Committee on Education and Employment]. (2016). *Access to real learning: The Impact of policy, funding and culture on students with disability*. Commonwealth of Australia. http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Employment/students_with_disability/Report.

Council of Australian Governments (2018). *National School Reform Agreement*. . Department of Education and Training. https://www.dese.gov.au/quality-schools-package/resources/national-school-reform-agreement

*Daniel J, Capin P and Steinle P (2021). 'A synthesis of the sustainability of remedial reading intervention effects for struggling adolescent readers'. *Journal of Learning Disabilities, 54*(3):170-186. https://doi.org/10.1177/0022219420972184

*Datchuk SM, Rodgers DB, Wagner K, Hier BO and Moore CT (2022). 'Effects of writing interventions on the level and trend of total words written: A meta-analysis'. *Exceptional Children, 88*(2):145-162. https://doi.org/10.1177/00144029211027537

*Datchuk SM, Wagner K and Hier BO (2020). 'Level and trend of writing sequences: A review and meta-analysis of writing interventions for students with disabilities'. *Exceptional Children, 86*(2):174-192. https://doi.org/10.1177/0014402919873311

de Bruin K (2018). *The NCCD and needs-based funding: An evaluation of disability resource allocation in Australian education* Australian Association for Research in Education, https://www.aare.edu.au/publications/aare-conference-papers/show/11762/the-nccd-and-needs-based-fundingan-evaluation-of-disability-resource-allocation-in-australian-education

de Bruin K, Graham LJ and Gallagher J (2020). 'What is the NCCD and what does it mean for my practice?' In Graham LJ (Ed.), *Inclusive education for the 21st century: Theory, policy and practice* (pp. 122-155). Allen and Unwin.

de Bruin K, and Stocker K (2021). 'Multi-Tiered Systems of Support: Comparing implementation in primary and secondary schools'. *Learning Difficulties Australia Bulletin, 52*(3). https://www.ldaustralia.org/publications/bulletin-volume-53-no-3-dec-2021/

de Haan M (2021). 'Supporting struggling adolescent readers through the Response to Intervention (RTI) framework'. *Australian Journal of Learning Difficulties, 26*(1):47-66. https://doi.org/10.1080/19404158.2020.1870512

Dempsey I and Davies M (2013). 'National test performance of young Australian children with additional educational needs'. *Australian Journal of Education, 57*(1):5-18. https://doi.org/10.1177/0004944112468700

Department of Education and Training Victoria (2022). *Student Resource Package Guide*. https://www.education.vic.gov.au/PAL/2022-srp-revised-guide.pdf

*Dietrichson J, Filges T, Klokker RH, Viinholt BCA, Bøg M and Jensen UH (2020). 'Targeted school-based interventions for improving reading and mathematics for students with, or at risk of, academic difficulties in Grades 7–12: A systematic review'. *Campbell Systematic Reviews, 16*(2). https://doi.org/10.1002/cl2.1081

*Donker A S, de Boer H, Kostons D, Dignath van Ewijk CC and van der Werf MPC (2014). 'Effectiveness of learning strategy instruction on academic performance: A meta-analysis'. *Educational Research Review, 11*, 1-26. https://doi.org/10.1016/j.edurev.2013.11.002

*Edmonds MS, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett KK and Schnakenberg JW (2009). 'A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers'. *Review of Educational Research, 79*(1):262-300. https://doi.org/10.3102/0034654308325998

*El Zein F, Solis M, Vaughn S and McCulley L (2014). 'Reading comprehension interventions for students with autism spectrum disorders: A synthesis of research'. *Journal of Autism and Developmental Disorders, 44*(6):1303-1322. https://doi.org/10.1007/s10803-013-1989-2

Elliott S, Davies M and Kettler R (2012). 'Australian students with disabilities accessing NAPLAN: Lessons from a decade of inclusive assessment in the United States'. *International Journal of Disability, Development, and Education, 59*(1):7-19. https://doi.org/10.1080/1034912X.2012.654934

Espin C, Chung S, Foegen A and Campbell H (2018). 'Curriculum-based measurement for secondary-school students'. In P. Pullen and M. Kennedey (Eds.), *Handbook of response to intervention and multi-tiered systems of support* (pp. 291-315). Routledge.

*Fallon LM, Collier-Meek MA, Maggin DM, Sanetti LMH and Johnson AH (2015). 'Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research'. *Exceptional Children, 81*(2):227-246. https://doi.org/10.1177/0014402914551738

*Filderman MJ, Austin CR, Boucher AN, O'Donnell K and Swanson EA (2022). 'A meta-analysis of the effects of reading comprehension interventions on the reading comprehension outcomes of struggling readers in third through 12th grades'. *Exceptional Children, 88*(2):163-184. https://doi.org/10.1177/00144029211050860

*Filderman MJ, Toste JR, Didion L and Peng P (2021). 'Data literacy training for K–12 teachers: A meta-analysis of the effects on teacher outcomes'. *Remedial and Special Education*, Advance Online Publication. https://doi.org/10.1177/07419325211054208

Fixsen DL, Blase KA, Horner R and Sugai G (2009). *Intensive technical assistance. Scaling-up brief. Number 2*. FPG Child Development Institute, University of North Carolina. https://eric.ed.gov/?id=ED507441

Fixsen DL, Naoom SF, Blase KA, Friedman RM and Wallace F (2005). *Implementation research: A synthesis of the literature*. https://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature

*Flynn LJ, Zheng X and Swanson HL (2012). 'Instructing struggling older readers: A selective meta-analysis of intervention research'. *Learning Disabilities Research and Practice, 27*(1):21-32. https://doi.org/10.1111/j.1540-5826.2011.00347.x

Foorman B, Beyler N, Borradaile K, Coyne M, Denton CA, Dimino J, Furgeson J, Hayes L, Henke J, Justice L, Keating B, Lewis W, Sattar S, Streke A, Wagner R and Wissel S (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade (NCEE 2016-4008)*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/practiceGuide/wwc_foundationalreading_040717.pdf

Foorman BR, Francis DJ, Fletcher JM, Schatschneider C and Mehta P (1998). 'The role of instruction in learning to read: Preventing reading failure in at-risk children'. *Journal of Educational Psychology*, *90*(1):37. https://doi.org/10.1037/0022-0663.90.1.37

Foorman B and Wanzec J (2016). 'Classroom reading instruction for all students'. In Jimerson A, Burns M and VanDerHeyden A (Eds.), *Handbook of Response to Intervention* (pp. 235-252). Springer. https://doi.org/10.1007/978-1-4899-7568-3_14

Francis DJ, Shaywitz SE, Stuebing KK, Shaywitz BA, and Fletcher JM (1996). 'Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis'. *Journal of Educational Psychology*, *88*(1):3–17. https://doi.org/10.1037/0022-0663.88.1.3

Fuchs LS and Deno SL (1991). 'Effects of curriculum within curriculum-based measurement'. *Exceptional Children, 58*(3):232-243. https://doi.org/10.1177/001440299105800306

Fuchs LS and Fuchs D (1991). 'Curriculum-based measurements: Current applications and future directions'. Preventing School Failure: *Alternative Education for Children and Youth, 35*(3):6-11. https://doi.org/10.1080/1045988X.1991.10871068

Frye D, Baroody AJ, Burchinal M, Carver, SM, Jordan NC and McDowell J (2013*). Teaching math to young children: A practice guide (NCEE 2014-4005)*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/early_math_pg_111313.pdf

*Garwood JD, Brunsting NC and Fox LC (2014). 'Improving reading comprehension and fluency outcomes for adolescents with emotional-behavioral disorders: Recent research synthesized'. *Remedial and Special Education, 35*(3):181-194. https://doi.org/10.1177/0741932513514856

Gersten R, Compton D, Connor CM, Dimino J, Santoro L, Linan-Thompson S and Tilly WD (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009-4045)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sci ences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_reading_pg_021809.pdf

Gersten R, Beckmann S, Clarke B, Foegen A, Marsh L, Star JR and Witzel B (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf

Gersten R, Chard DJ, Jayanthi M, Baker SK, Morphy P and Flojo J (2009). 'Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components'. *Review of Educational Research, 79*(3):1202-1242. https://doi.org/10.3102/0034654309334431

Gersten R, Jayanthi M and Dimino J (2017). 'Too much, too soon? Unanswered questions from national response to intervention evaluation'. *Exceptional Children, 83*:244–254.

Geiger V, Goos M and Forgasz H (2015). 'A rich interpretation of numeracy for the 21st century: A survey of the state of the field'. *ZDM Mathematics Education*, 47(4):531-548. https://doi.org/10.1007/s11858-015-0708-1

*Gillespie A and Graham S (2014). 'A meta-analysis of writing interventions for students with learning disabilities'. *Exceptional Children, 80*(4):454-473. https://doi.org/10.1177/0014402914527238

*Goodwin AP and Ahn S (2013). 'A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children'. *Scientific Studies of Reading, 17*(4):257-285. https://doi.org/10.1080/10888438.2012.689791

Goss P and Sonnemann J (2016). *Widening gaps: what NAPLAN tells us about student progress*. https://grattan.edu.au/report/widening-gaps/

Graham S (2020). 'The sciences of reading and writing must become more fully integrated'. *Reading Research Quarterly, 55*, S35-S44. https://doi.org/10.1002/rrq.332

*Graham S, Liu X, Bartlett B, Ng C, Harris KR, Aitken A, Barkel A, Kavanaugh C and Talukdar J (2018). 'Reading for writing: A meta-analysis of the impact of reading interventions on writing'. *Review of Educational Research, 88*(2):243-284. https://doi.org/10.3102/0034654317746927

Graham S and Perin D (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York*. Alliance for Excellent Education. www.carnegie.org

*Hall-Mills SS and Marante LM (2022). 'Explicit text structure instruction supports expository text comprehension for adolescents with learning disabilities: A systematic review'. *Learning Disability Quarterly, 45*(1):55-68. https://doi.org/10.1177/0731948720906490

*Hall MS and Burns MK (2018). 'Meta-analysis of targeted small-group reading interventions'. *Journal of School Psychology, 66*, 54-66. https://doi.org/10.1016/j.jsp.2017.11.002

Hanushek EA (2021). 'Addressing cross-national generalizability in educational impact evaluation'. *International Journal of Educational Development*, 80. https://doi.org/10.1016/j.ijedudev.2020.102318

Harlacher JE, Walker NJN and Sanford AK (2010). 'The "I" in RTI: Research-based factors for intensifying instruction'. *Teaching Exceptional Children*, *42*(6):30-38. https://doi.org/10.1177/004005991004200601

Harn BA, Chard DJ, Biancarosa G and Kame'enui EJ (2011). 'Coordinating instructional supports to accelerate at-risk first-grade readers' performance: An essential mechanism for effective RTI'. *The Elementary School Journal*, *112*(2):332-355. https://doi.org/10.1086/661997

Hempenstall K (2016). *Read about it: Scientific evidence for effective teaching of reading (Research report 11)*. The Centre for Independent Studies. https://www.cis.org.au/wp-content/uploads/2016/03/rr11.pdf

Hempenstall K (2020). 'Teaching reading through Direct Instruction: A role for educational psychologists?'. *The Educational and Developmental Psychologist*, *37*(2):133-139. https://doi.org/10.1017/edp.2020.13

*Huddle S, Hosp J and Watt S (2017). 'Reading interventions for adolescent English language learners with reading difficulties: A synthesis of the evidence'. *Multiple Voices for Ethnically Diverse Exceptional Learners, 17*(2):3-18. https://doi.org/10.5555/2158-396X.17.2.3

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004). https://www.govinfo.gov/content/pkg/PLAW-108publ446/html/PLAW-108publ446.htm

Intervention Central (n.d.). *New York* state *RTI technical assistance center*. https://www.interventioncentral.org/new-york-state-rti-technical-assistance-center

Institute of Education Sciences (2022). *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*. Institute of Education Services, US Department of Education. www.whatworks.ed.gov/reviewprocess/study_standards_final.pdf

JBI Global (n.d.). *System for the Unified Management of the Assessment and Review of Information (SUMARI)*. In https://app.jbisumari.org/

Jitendra A and Dupuis D (2016). 'The role of Tier 1 mathematics instruction in elementary and middle schools: Promoting mathematics success'. In Jimerson S, Burns M, and VanDerHeyden A (Eds.), *Handbook of Response to Intervention* (pp. 215-234). Springer. https://doi.org/10.1007/978-1-4899-7568-3_32

*Jitendra AK, Lein AE, Im SH, Alghamdi AA, Hefte SB and Mouanoutoua J (2018). 'Mathematical interventions for secondary students with learning disabilities and mathematics difficulties: A meta-analysis'. *Exceptional Children, 84*(2):177-196. https://doi.org/10.1177/0014402917737467

*Joseph LM and Schisler R (2009). 'Should adolescents go back to the basics?: A review of teaching word reading skills to middle and high school students'. *Remedial and Special Education, 30*(3):131-147. https://doi.org/10.1177/0741932508315646

Juel C (1988). 'Learning to read and write: A longitudinal study of 54 children from first through 4th grades'. *Journal of Educational Psychology*, *80*(4):437–447. https://doi.org/10.1037/0022-0663.80.4.437

*Jung PG, McMaster KL, Kunkel AK, Shin J and Stecker PM (2018). 'Effects of data-based individualization for students with intensive learning needs: A meta-analysis'. *Learning disabilities research and practice, 33*(3):144-155. https://doi.org/10.1111/ldrp.12172

*Kaldenberg ER, Ganzeveld P, Hosp JL and Rodgers DB (2016). 'Common characteristics of writing interventions for students with learning disabilities: A synthesis of the literature'. *Psychology in the Schools, 53*(9):938-953. https://doi.org/10.1002/pits.21958

*Kaldenberg ER, Watt SJ and Therrien WJ (2015). 'Reading instruction in science for students with learning disabilities'. *Learning Disability Quarterly, 38*(3):160-173. https://doi.org/10.1177/0731948714550204

Kamil ML, Borman GD, Dole J, Kral CC, Salinger T and Torgeson J (2008). 'Improving adolescent literacy: Effective classroom and intervention practices; a practice guide' (NCEE #2008–4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

*Kang EY, McKenna JW, Arden S and Ciullo S (2015). 'Integrated reading and writing interventions for students with learning disabilities: A review of the literature'. *Learning disabilities research and practice, 31*(1):22-33. https://doi.org/10.1111/ldrp.12091

Kilpatrick, D. (2003). *Phonological Awareness Screening Test*. https://thepasttest.com/

*Kilgus SP, Methe SA, Maggin DM, and Tomasula JL (2014). 'Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening'. *Journal of School Psychology, 52*(4):377-405. https://doi.org/10.1016/j.jsp.2014.06.002

*Kim AH, Vaughn S, Wanzek J and Wei S (2004). 'Graphic organizers and their effects on the reading comprehension of students with LD: A synthesis of research'. *Journal of Learning Disabilities, 37*(2):105-118. https://doi.org/10.1177/00222194040370020201

*Kim W, Linan-Thompson S and Misquitta R (2012). 'Critical factors in reading comprehension instruction for students with learning disabilities: A research synthesis'. *Learning Disabilities Research and Practice, 27*(2):66-78. https://doi.org/10.1111/j.1540-5826.2012.00352.x

Kraft MA (2020). 'Interpreting effect sizes of education interventions'. *Educational Researcher,* 49(4):241-253. https://doi.org/10.3102/0013189x20912798

*Kuder SJ (2017). 'Vocabulary instruction for secondary students with reading disabilities: An updated research review'. *Learning Disability Quarterly, 40*(3):155-164. https://doi.org/10.1177/0731948717690113

*Lee J, Bryant D, Ok M and Shin M (2020). 'A systematic review of interventions for algebraic concepts and skills of secondary students with learning disabilities'. *Learning Disabilities Research and Practice, 35*(2):89-99. https://doi.org/10.1111/ldrp.12217

*Lee J and Yoon SY (2017). 'The effects of repeated reading on reading fluency for students with reading disabilities: A meta-analysis'. *Journal of Learning Disabilities, 50*(2):213-224. https://doi.org/10.1177/0022219415605194

*Lein AE, Jitendra AK and Harwell MR (2020). 'Effectiveness of mathematical word problem solving interventions for students with learning disabilities and/or mathematics difficulties: A meta-analysis'. *Journal of Educational Psychology, 112*(7):1388-1408. https://doi.org/10.1037/edu0000453

Lemons CJ, Kearns DM and Davidson KA (2014). 'Data-based individualization in reading: Intensifying interventions for students with significant reading disabilities'. *Teaching Exceptional Children*, *46*(4):20-29. https://doi.org/10.1177/0040059914522978

*Maccini P, Mulcahy CA and Wilson MG (2007). 'A follow-up of mathematics interventions for secondary students with learning disabilities'. *Learning disabilities research and practice, 22*(1):58-74. https://doi.org/10.1111/j.1540-5826.2007.00231.x

Masters G (2017). 'Using PISA to monitor trends and evaluate reforms in Australia'. In L. Volante (Ed.), *The PISA Effect on Global Educational Governance* (pp. 175-188). Routledge. https://doi.org/https://doi.org/10.4324/9781315440521

*McClain MB, Haverkamp CR, Benallie KJ, Schwartz SE and Simonsmeier V (2021). 'How effective are reading comprehension interventions for children with ASD? A meta-analysis of single-case design studies'. *School Psychology*, 36(2):107-121. https://doi.org/10.1037/spq0000424

Moretti GAS and Frandell T (2013). *Literacy from a Right to Education Perspective. Contribution prepared for the Final Evaluation of the Implementation of the International Plan of Action for the United Nations Literacy Decade (UNLD): Education for All*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000221427

National Center for Intensive Intervention (n.d.). *Academic Screening Tools Chart*. https://intensiveintervention.org/resource/academic-screening-tools-chart

National Center on Multi-Tiered System of Supports (2022). https://mtss4success.org

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel: Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction.* https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

National Joint Committee on Learning Disabilities (NJCLD) (2005). 'Responsiveness to intervention and learning disabilities'. *Learning Disability Quarterly*, *28*(4):249–260. https://doi.org/10.2307/4126964

Neale MD (1999). *The Neale Analysis of Reading Ability* (*3rd ed*). Melbourne, Australia: ACER.

*Newell KW, Codding RS and Fortune TW (2020). 'Oral reading fluency as a screening tool with English learners: A systematic review'. *Psychology in the Schools, 57*(8):1208-1239. https://doi.org/10.1002/pits.22406

Office of Special Education Programs (OSEP) (n.d.). *Database of federally funded technical assistance and research centers*. https://osepideasthatwork.org/find-center-or-grant/find-a-center

*Oxley E and de Cat C (2021). 'A systematic review of language and literacy interventions in children and adolescents with English as an additional language (EAL)'. *Language Learning Journal, 49*(3):265-287. https://doi.org/10.1080/09571736.2019.1597146

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P and Moher D (2021). 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews'. *Systematic reviews, 10*(1):1-11. https://doi.org/10.1136/bmj.n71

*Peterson AK, Fox CB and Israelsen M (2020). 'A systematic review of academic discourse interventions for school-aged children with language-related learning disabilities [Review]'. *Language, Speech, and Hearing Services in Schools, 51*(3):866-881. https://doi.org/10.1044/2020_LSHSS-19-00039

Poed S, Cologon K and Jackson R (2020). 'Gatekeeping and restrictive practices by Australian mainstream schools: Results of a national survey'. *International Journal of Inclusive Education, 26*(8):766-779. https://doi.org/10.1080/13603116.2020.1726512

Powell SR and Stecker PM (2014). 'Using data-based individualization to intensify mathematics intervention for students with disabilities'. *Teaching Exceptional Children*, *46*(4):31-37. https://doi.org/10.1177/0040059914523735

Preston AI, Wood CL and Stecker PM (2016). 'Response to Intervention: Where it came from and where it's going'. *Preventing School Failure: Alternative Education for Children and Youth*, 60(3):173-182. https://doi.org/10.1080/1045988X.2015.1065399

Pullen PC, van Dijk W, Gonsalves VE, Lane HB and Ashworth KE (2018). 'RTI and MTSS: Response to Intervention and Multi-Tiered Systems of Support: How do they differ and how are they the same, if at all?'. In Pullen PC and Kennedy MJ (Eds.), *Handbook of Response to Intervention and Multi-Tiered Systems of Support* (pp. 5-10). Routledge.

*Reed DK, Cummings KD, Schaper A and Biancarosa G (2014). 'Assessment fidelity in reading intervention research: A synthesis of the literature'. *Review of Educational Research, 84*(2):275-321. https://doi.org/10.3102/0034654314522131

*Roberts CA, Leko MM and Wilkerson KL (2013). 'New directions in reading instruction for adolescents with significant cognitive disabilities'. *Remedial and Special Education, 34*(5):305-317. https://doi.org/10.1177/0741932513485447

Rose J (2006). *Independent review of the teaching of early reading*. https://dera.ioe.ac.uk/5551/2/report.pdf

Rosenshine B and Stevens R (1986). 'Teaching functions'. In Wittrock MC (Ed.), Handbook of Research on Teaching (3rd ed.) (pp. 376-391). New York: Macmillan .

Rowe K (2005). *Teaching reading: Report and recommendations.* https://research.acer.edu.au/tll_misc/5/

Sailor W (2015). 'Advances in schoolwide inclusive school reform'. *Remedial and Special Education*, *36*(2):94-99. https://doi.org/10.1177/0741932514555021

*Scammacca N, Roberts G, Vaughn S, Edmonds M, Wexler J, Reutebuch CK and Torgesen JK (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. RMC Research Corporation, Center on Instruction. https://files.eric.ed.gov/fulltext/ED521837.pdf

*Scammacca NK, Roberts G, Vaughn S and Stuebing KK (2015). 'A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011'. *Journal of Learning Disabilities, 48*(4):369-390. https://doi.org/10.1177/0022219413504995

Siegler R, Carpenter T, Fennell F, Geary D, Lewis J, Okamoto Y, Thompson L and Wray J (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide (NCEE #2010-4039)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/fractions_pg_093010.pdf

Shanahan T, Callison K, Carriere C, Duke NK, Pearson PD, Schatschneider C., and Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide (NCEE 2010-4038)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/readingcomp_pg_092810.pdf

Shaywitz SE, Fletcher JM, Holahan JM, Schneider AE, Marchione KE, Stuebing KK, Francis DJ, Pugh KR and Shaywitz BA (1999). 'Persistence of dyslexia: The Connecticut longitudinal study at adolescence'. *Pediatrics*, 104(6):1351-1359. https://doi.org/10.1542/peds.104.6.1351

Sedita J (2011). Adolescent literacy: Addressing the needs of students in grades 4–12. In Birsh J (Ed.), *Multisensory teaching of basic language skills* (3rd ed., pp. 1-23). Baltimore Brookes. http://keystoliteracy.com/wp-content/uploads/2012/08/Adolescent-Literacy-addressing-the-needs-of-students.pdf

Senate Education and Employment Committee (2016). *Access to Real Learning: the Impact of Policy, Funding and Culture on Students with Disability*. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Employment/students_with_disability/Report

*Shin J and McMaster K (2019). 'Relations between CBM (oral reading and maze) and reading comprehension on state achievement tests: A meta-analysis'. *Journal of School Psychology, 73*. https://doi.org/10.1016/j.jsp.2019.03.005

*Shin M and Bryant DP (2015). 'Fraction interventions for students struggling to learn mathematics: A research synthesis'. *Remedial and Special Education, 36*(6):374-387. https://doi.org/10.1177/0741932515572910

*Shin M, Bryant DP, Powell SR, Jung PG, Ok MW and Hou F (2021). 'A meta-analysis of single-case research on word-problem instruction for students with learning disabilities'. *Remedial and Special Education, 42*(6):398-411. https://doi.org/10.1177/0741932520964918

Shinn MR, Windram HS and Bollman KA (2016). 'Implementing response to intervention in secondary schools'. In Jimerson S, Burns M, and VanDerHeyden A (Eds.), *Handbook of Response to Intervention* (pp. 563-586). Springer. https://doi.org/10.1007/978-1-4899-7568-3_32

Simpson A (2019). 'Separating arguments from conclusions: The mistaken role of effect size in educational policy research'. *Educational Research and Evaluation*, 25(1-2):99-109. https://doi.org/10.1080/13803611.2019.1617170

*Solis M, Ciullo S, Vaughn S, Pyle N, Hassaram B and Leroux A (2012). Reading comprehension interventions for middle school students with learning disabilities: A synthesis of 30 years of research. *Journal of Learning Disabilities, 45*(4):327-340. https://doi.org/10.1177/0022219411402691

Stanley G (2008). *National numeracy review report.* Human Capital Working Group Council of Australian Governments. https://apo.org.au/node/4016

*Steinle PK, Stevens E and Vaughn S (2022). 'Fluency interventions for struggling readers in grades 6 to 12: A research synthesis'. *Journal of Learning Disabilities, 55*(1):3-21. https://doi.org/10.1177/0022219421991249

*Stevens EA, Rodgers MA and Powell SR (2018). 'Mathematics interventions for upper elementary and secondary students: A meta-analysis of research'. *Remedial and Special Education,* 39(6):327-340. https://doi.org/10.1177/0741932517731887

*Stevens EA, Park S and Vaughn S (2019). 'A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978-2016'. *Remedial and Special Education, 40*(3):131-149. https://doi.org/10.1177/0741932517749940

*Stewart AA and Austin CR (2020). 'Reading interventions for students with or at risk of attention-deficit/hyperactivity disorder: A systematic review'. *Remedial and Special Education*, 41(6):352-367. https://doi.org/10.1177/0741932519849660

*Swanson E, Hairrell A, Kent S, Ciullo S, Wanzek JA and Vaughn S (2014). 'A synthesis and meta-analysis of reading interventions using social studies content for students with learning disabilities'. *Journal of Learning Disabilities, 47*(2):178-195. https://doi.org/10.1177/0022219412451131

Swanson HL and Hoskyn M (1998). 'Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. Review of Educational Research', 68(3):277-321. https://doi.org/10.3102/00346543068003277

Thomson S, De Bortoli L, Underwood C and Schmid M (2019). *PISA 2018: Reporting Australia's results*. Volume I student performance. Australian Council for Educational Research (ACER). https://research.acer.edu.au/ozpisa/35

Thomson S, Wernert N, Rodrigues S and O'Grady E (2020). *TIMSS 2019 Australia.* Volume I: Student performance. Australian Council for Educational Research. https://doi.org/10.37517/978-1-74286-614-7

Tindal G, Shinn M, Walz L and Germann G (1987). 'Mainstream consultation in secondary settings: The Pine County model'. *The Journal of Special Education*, *21*(3):94-106. https://doi.org/10.1177/0022466987021003

Torgerson CJ, Brooks G and Hall J (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling (Research report RR711)*. Department for Education and Skills, the University of Sheffield. http://wsassets.s3.amazonaws.com/ws/nso/pdf/f211f7a331f7a210fa7b64d0855f2afc.pdf

Torgesen JK, Wagner RK and Rashotte CA (1999). *Test of Word Reading Efficiency*. Pro-Ed.

Triano S (2000). 'Categorical eligibility for special education: The enshrinement of the medical model in disability policy'. *Disability Studies Quarterly*, *20*(4):1-16. https://doi.org/10.18061/dsq.v20i4.263

Tricco AC, Khalil H, Holly C, Feyissa G, Godfrey C, Evans C, Sawchuck D, Sudhakar M, Asahngwa C, Stannard D, Abdulahi M, Bonnano L, Aromataris E, McInerney P, Wilson R, Pang D, Wang Z, Cardoso AF, Peters MDJ, Marnie C, Barker T, Munn Z (2022). 'Rapid reviews and the methodological rigor of evidence synthesis: A JBI position statement'. *JBI Evidence Synthesis, 20*(4):944-949. https://doi.org/10.11124/JBIES-21-00371

Tunmer WE and Hoover WA (2019, 2019/01/02). 'The cognitive foundations of learning to read: a framework for preventing and remediating reading difficulties'. *Australian Journal of Learning Difficulties, 24*(1):75-93. https://doi.org/10.1080/19404158.2019.1614081

United Nations (1948). *Universal Declaration of Human Rights*. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations Educational, Scientific and Cultural Organization (1990). *World declaration on education for all and framework for action to meet basic learning needs basic learning needs.* https://digitallibrary.un.org/record/101572?ln=en

University of Oregon (n.d.). *Dynamic Indicators of Basic Early Literacy Skills*. https://dibels.uoregon.edu/

*Valasa LL, Mason LH and Hughes C (2014). 'Essay-writing interventions for adolescents with high incidence disabilities: A review of research'. *International Journal for Research in Learning Disabilities, 2*(1):72-97. https://eric.ed.gov/?id=EJ1155699

Van Camp AM, Wehby JH, Martin BLN, Wright JR and Sutherland KS (2020). 'Increasing opportunities to respond to intensify academic and behavioral interventions: A meta-analysis'. *School Psychology Review*, *49*(1):31-46. https://doi.org/10.1080/2372966X.2020.1717369

Vaughn S and Fletcher JM (2012). 'Response to intervention with secondary school students with reading difficulties'. *Journal of Learning Disabilities, 45*(3):244-256. https://doi.org/10.1177/0022219412442157

Vaughn S, Gersten R and Chard DJ (2000). 'The underlying message in LD intervention research: Findings from research syntheses'. *Exceptional Children*, *67*:99–114. https://doi.org/10.1177/001440290006700107

Vellutino FR, Scanlon DM, Sipay ER, Small SG, Pratt A, Chen R and Denckla MB (1996). 'Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability'. *Journal of Educational Psychology*, *88*(4):601–638. https://doi.org/10.1037/0022-0663.88.4.601

Walker HM, Severson HH, Feil EG, Hawken L and Sabey C (2012). 'Implementing universal screening systems within an RTI-PBS context'. In *Handbook of youth prevention science* (pp. 96-120). Routledge. https://doi.org/10.4324/9780203866412

Wang R and Papageorghiou AT (2021). 'Umbrella review—a useful tool when it's raining systematic reviews'. *BJOG: An International Journal of Obstetrics and Gynaecology, 128*(13):2150-2150. https://doi.org/10.1111/1471-0528.16827

*Wanzek J, Vaughn S, Scammacca NK, Metz K, Murray CS, Roberts G and Danielson L (2013). 'Extensive reading interventions for students with reading difficulties after Grade 3'. *Review of Educational Research, 83*(2):163-195. https://doi.org/10.3102/0034654313477212

*Watt SJ, Watkins JR and Abbitt J (2016). 'Teaching algebra to students with learning disabilities: Where have we come and where should we go?' *Journal of Learning Disabilities, 49*(4):437-447. https://doi.org/10.1177/0022219414564220

*Wexler J, Vaughn S, Edmonds M and Reutebuch CK (2008). 'A synthesis of fluency interventions for secondary struggling readers'. *Reading and Writing: An Interdisciplinary Journal, 21*(4):317-347. https://doi.org/10.1007/s11145-007-9085-7

*Williams KJ, Austin CR and Vaughn S (2018). 'A synthesis of spelling interventions for secondary students with disabilities'. *Journal of Special Education, 52*(1):3-15. https://doi.org/10.1177/0022466917732777

Williams KJ, Stevens E and Vaughn S (2018). 'RTI in Secondary Schools: Current issues and recommendations'. In Pullen PC and Kennedy MJ (Eds.), *Handbook of Response to Intervention and Multi-Tiered Systems of Support* (pp. 338-350). Routledge.

Wu, M (2015). 'What national testing data can tell us'. In Lingard B, Thompson G, and Sellar S (Eds.), *In National Testing in Schools* (pp. 18-29). Routledge. https://doi.org/ https://doi.org/10.4324/9781315659312

*Xin YP and Jitendra AK (1999). 'The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis'. *The Journal of Special Education, 32*(4):207-225. https://doi.org/10.1177/002246699903200402

Yell ML and Drasgow E (2007). 'Assessment for eligibility under IDEIA and the 2006 Regulations'. *Assessment for Effective Intervention*, *32*(4):202–211. https://doi.org/10.1177/15345084070320040201

Ysseldyke J and Marston D (1999). 'Origins of categorical special education services in schools and a rationale for changing them'. In Reschly D, Tilly D, and Grimes J (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 1-18). Iowa Department of Education. https://files.eric.ed.gov/fulltext/ED434454.pdf

Zigmond N and Sansone J (1986). 'Designing a program for the learning disabled'. *Remedial and Special Education, 7*(5):13-17. https://doi.org/10.1177/074193258600700504

Appendices

A.1. JBI Critical Appraisal Checklist for systematic reviews and research syntheses

Reviewer ..... Date .....

Author ..... Year ..... Record Number .....

Checklist questions	Yes	No	Unclear	N/A
1. Is the review question clearly and explicitly stated?	☐	☐	☐	☐
2. Were the inclusion criteria appropriate for the review question?	☐	☐	☐	☐
3. Was the search strategy appropriate?	☐	☐	☐	☐
4. Were the sources and resources used to search for studies adequate?	☐	☐	☐	☐
5. Were the criteria for appraising studies appropriate?	☐	☐	☐	☐
6. Was critical appraisal conducted by 2 or more reviewers independently?	☐	☐	☐	☐
7. Were there methods to minimize errors in data extraction?	☐	☐	☐	☐
8. Were the methods used to combine studies appropriate?	☐	☐	☐	☐
9. Was the likelihood of publication bias assessed?	☐	☐	☐	☐
10. Were recommendations for policy and/or practice supported by the reported data?	☐	☐	☐	☐
11. Were the specific directives for new research appropriate?	☐	☐	☐	☐

Overall appraisal

☐ Include ☐ Exclude ☐ Seek further info

Comments (Including reason for exclusion)

.....

.....

.....

.....

.....

A.2. JBI Critical Appraisal Checklist for text and opinion papers

Reviewer ..... Date .....

Author ..... Year ..... Record Number .....

Checklist questions	Yes	No	Unclear	Not applicable
1. Is the source of the opinion clearly identified?	☐	☐	☐	☐
2. Does the source of opinion have standing in the field of expertise?	☐	☐	☐	☐
3. Are the interests of the relevant population the central focus of the opinion?	☐	☐	☐	☐
4. Is the stated position the result of an analytical process and is there logic in the opinion expressed?	☐	☐	☐	☐
5. Is there reference to the extant literature?	☐	☐	☐	☐
6. Is any incongruence with the literature/sources logically defended?	☐	☐	☐	☐

Overall appraisal

☐ Include ☐ Exclude ☐ Seek further info

Comments (Including reason for exclusion)

.....

.....

.....

.....

.....

A.3. Data extraction tool

Author and year (APA citation)	
Objective(s)	
Participants (characteristics/total number)	
Interventions/phenomena of interest	
Setting/Context	
Number of databases/sources searched	
Date range of included studies	
Number of Studies/Type of Studies/ Country of origin of included studies	
Appraisal instrument and rating	
Type of Review/Method of analysis	
Outcome(s)	
Results/findings	
Comments	

A.4. NVIVO codebook

Adult code	Child code
Service delivery model	
	Background and history
	Defining characteristics of multi-tier models
	Facilitators of uptake
	Facilitators of uptake and impact
	Impact on school
	Impact on Student
	Impact on system
	MTSS
	Rationale for change
	RTI
Screening	
	Mathematics
	Reading
	Writing
Assessment	
	Progress monitoring
	Mathematics
	Reading
	Writing
Reading interventions	
	Comprehension
	decoding
	Fluency
	Phonemic awareness
	Vocabulary
Writing interventions	
	Grammar
	Spelling
	Transcription
	Writing fluency
	Writing sentences
	Writing texts
	Student classification
	Type of measure

Adult code	Child code
Mathematics interventions	
	Algebra
	Arithmetical skills
	Basic number fact fluency
Factors moderating intervention impact	
	Age
	Duration
	Frequency
	Group Size
	Interventionist
	Recency of publication
	Student classification
	Type of measure
	Writing
Supporting fidelity	
	Materials
	Professional learning
	Resourcing
Research gaps	
	Assessment
	Fidelity of implementation
	Maintenance and Generalisability
	Mathematics
	Reading
	Service delivery
	Study quality
	Writing

A.5. Descriptive summary of studies

Author	Study date	Timeframe	Journal	Method	Classification	No of Studies	Participant Type	Participant Age
Ardoin et al. (2013)	2013	1981–2010	Journal of School Psychology,	SR-O	A	102	K-12	Not reported
Asaro-Saddler et al. (2021)	2021	2016–2020	Exceptionality	SRMA-SC	W	10	Schoolwide	M=10.10yrs
Barrio et al. (2015)	2015	2003–2013	Action in Teacher Education	SR-O	SDM	10	Teachers	Not reported
Berkeley et al. (2009)	2009	2007–2008	Journal of Learning Disabilities	SR-O	SDM	0	E HS	Not reported
Berkeley et al. (2010)	2010	1995–2006	Journal of Learning Disabilities	SRMA-O	R	40	K-12	M=12.5
Berkeley et al. (2020)	2020	1995–2006	Remedial and Special Education	SR-O	SDM	50	E HS	Not reported
Boon et al. (2018)	2018	1975–2017	Learning Disabilities: A Multidisciplinary Journal	SR-SC	W	10	Gr3-12	Not reported
Bouck and Park (2018)	2018	1975–2017	Education and Treatment of Children	SR-M	N	36	PK-12	Not reported
Browder et al. (2006)	2006	1975–2003	Exceptional Children	SRMA-M	R	128	K-12	5–12yrs
Burke et al. (2015)	2015	1990–2014	Behavior Modification	SRMA-SC	R	11	Gr6-9	M=13.4
Burns and Symington (2002)	2002	open–2002	Journal of School Psychology	SRMA-G	SDM	9	Gr1-7	Not reported
Burns et al. (2005)	2005	open–2004	Journal of Psychoeducational Assessment	SRMA-O	SDM	31	K-12	0
Ciullo and Reutebuch (2013)	2013	open–2012	Learning Disabilities Research and Practice	SR-M	Imp	12	Gr6-12	0
Collins et al. (2018)	2018	1978–2014	Journal of Learning Disabilities	SRMA-G	R	82	K-12	5–19yrs
Daniel et al. (2021)	2021	1996–2019	Journal of Learning Disabilities	SRMA-G	R	10	Gr6-12	0
Datchuk et al. (2020)	2020	open–2018	Exceptional Children	SRMA-SC	W	18	Middle	7–18yrs
Datchuk et al. (2022)	2022	1980–2021	Exceptional Children	SRMA-SC	W	42	Schoolwide	6–19yrs
Dietrichson et al. (2020)	2020	1980–2018	Campbell Systematic Reviews	SRMA-G	N	71	Gr7–12	0

Author	Study date	Timeframe	Journal	Method	Classification	No of Studies	Participant Type	Participant Age
Donker et al. (2014)	2014	2000–2012	Educational Research Review	SRMA-O	N	58	Gr2-11	0
Edmonds et al. (2009)	2009	1994–2004	Review of Educational Research	SRMA-M	R	29	Gr6-12	11–21yrs
El Zein et al. (2014)	2014	1980–2012	Journal of Autism and Developmental Disorders	SR-M	R	12	Gr2-8	7–17.5yrs
Fallon et al. (2015)	2015	1960–2011	Exceptional Children	SR-SC	Imp	47	Teachers	0
Filderman et al. (2021)	2021	1975–2020	Remedial and Special Education	SRMA-G	A	33	Teachers	0
Filderman et al. (2022)	2022	1975–2019	Exceptional Children	SRMA-M	A	63	ES/C	0
Flynn et al. (2012)	2012	1994–2002	Learning Disabilities Research and Practice	SRMA-G	R	10	Gr5-09	9.5–13yrs
Garwood et al. (2014)	2014	2004–2012	Remedial and Special Education	SR-SC	R	9	Gr5-9	11–16yrs
Gillespie and Graham (2014)	2014	open–2011	Exceptional Children	SRMA-M	W	43	Gr1-12	0
Goodwin and Ahn (2013)	2013	1980–2012	Scientific Studies of Reading	SRMA-G	R	30	K-9	15–19yrs
Graham et al. (2018)	2018	open–2016	Review of Educational Research	SRMA-M	R	54	K-12	0
Hall and Burns (2018)	2018	1997–2019	Journal of School Psychology	SRMA-G	R	26	K-12	0
Hall-Mills and Marante (2022)	2022	1997–2019	Learning Disability Quarterly	SR-M	R	9	Gr7-12	0
Huddle et al. (2017)	2017	1980–2013	Multiple Voices for Ethnically Diverse Exceptional Learners	SR-G	R	8	Gr4-12	0
Jitendra et al. (2018)	2018	1990–2017	Exceptional Children	SRMA-G	N	19	Secondary	0
Joseph and Schisler (2009)	2009	1986–2006	Remedial and Special Education	SRMA-M	R	23	Gr6-12	13–17yrs
Jung et al. (2018)	2018	1980–2017	Learning Disabilities Research and Practice	SRMA-G	A	14	K-12	0
Kaldenberg et al. (2015)	2015	1998–2012	Learning Disability Quarterly	SRMA-SC	R	20	Gr5-12	0
Kaldenberg et al. (2016)	2016	1980–2012	Psychology in the Schools	SRMA-G	W	23	Gr4-12	0
Kang et al. (2015)	2015	1975–2015	Learning Disabilities Research and Practice	SR-M	R	10	Gr2-8	0

Author	Study date	Timeframe	Journal	Method	Classification	No of Studies	Participant Type	Participant Age
Kilgus et al. (2014)	2014	open–2013	Journal of School Psychology	SRMA-O	A	34	GrK-8	0
Kim et al. (2004)	2004	1963–2001	Journal of Learning Disabilities	SR-M	R	21	Gr1-12	0
Kim et al. (2012)	2012	1990–2010	Learning Disabilities Research and Practice	SR-O	R	14	Gr6-8	0
Kuder (2017)	2017	2003–2016	Learning Disability Quarterly	SR-O	R	7	Gr6-10	0
Lee and Yoon (2017)	2017	1969–2019	Journal of Learning Disabilities	SRMA-M	R	34	Gr1-11	0
Lee et al. (2020)	2020	1994–2014	Learning Disabilities Research and Practice	SRMA-M	N	12	Gr6-12	0
Lein et al. (2020)	2020	1987–2019	Journal of Educational Psychology	SRMA-G	N	31	K-12	0
Maccini et al. (2007)	2007	1995–2006	Learning Disabilities Research and Practice	SR-M	N	23	Secondary	M=11-15.6
McClain et al. (2021)	2021	open–2020	School Psychology	SRMA-SC	R	20	Gr2-10	6–17yrs
Newell et al. (2020)	2020	1980–2019	Psychology in the Schools	SR-O	A	31	GrK-8	0
Oxley and de Cat (2021)	2021	2014–2017	Language Learning Journal	SR-G	R	26	K-12	4–18yrs
Peterson et al. (2020)	2020	open–2019	Language, Speech and Hearing Services in Schools	SR-G	W	7	Gr4-8	9–14yrs
Reed and Cummings (2014)	2014	2001–2012	Review of Educational Research	SR-G	A	46	GrK-8	0
Roberts et al. (2013)	2013	1975–2011	Remedial and Special Education	SR-M	R	19	Secondary	13–21yrs
Scammacca et al. (2007)	2007	1980–2006	Interventions for Adolescent Struggling Readers: A Meta- analysis with Implications for Practice	SRMA-G	R	31	Gr5-12	0
Scammacca et al. (2015)	2015	1980–2011	Journal of Learning Disabilities	SRMA-G	R	82	Gr4-12	0
Shin and Bryant (2015)	2015	1975–2014	Remedial and Special Education	SR-M	N	17	Gr3-12	0
Shin and McMaster (2019)	2019	1975–2014	Journal of School Psychology	SRMA-G	A	61	Gr1-10	0
Shin et al. (2021)	2021	1975–2020	Remedial and Special Education	SRMA-SC	N	20	Gr1-12	0

Author	Study date	Timeframe	Journal	Method	Classification	No of Studies	Participant Type	Participant Age
Solis et al. (2012)	2012	1979–2012	Journal of Learning Disabilities	SR-M	R	12	Gr6-8	12–14yrs
Steinle et al. (2022)	2022	2006–2019	Journal of Learning Disabilities	SR-M	R	17	Gr6-12	0
Stevens et al. (2018)	2018	1978–2016	Remedial and Special Education	SRMA-M	N	24	Gr4-12	0
Stevens et al. (2019)	2019	1990–2015	Remedial and Special Education	SRMA-M	R	23	Gr3-12	0
Stewart and Austin (2020)	2020	2015–2017	Remedial and Special Education	SR-M	R	16	Gr4-7	0
Swanson et al. (2014)	2014	open–2010	Journal of Learning Disabilities	SRMA-M	R	27	K-12	0
Valasa et al. (2014)	2014	open–2013	International Journal for Research in Learning Disabilities	SRMA-M	W	26	Gr6-12	0
Wanzek et al. (2013)	2013	1995–2011	Review of Educational Research	SRMA-M	R	19	Gr6-9	0
Watt et al. (2016)	2016	1980–2014	Journal of Learning Disabilities	SRMA-M	N	15	Gr7-9	0
Wexler et al. (2008)	2008	1980–2005	Reading and Writing: An Interdisciplinary Journal	SRMA-M	R	19	Gr6-12	11–21yrs
Williams et al. (2018)	2018	2004–2016	The Journal of Special Education	SR-SC	W	13	Gr6-12	0
Xin and Jitendra (1999)	1999	1980–1996	The Journal of Special Education	SRMA-M	N	25	Gr1-12	0

A.6. Critical appraisal results from qualitative and quantitative systematic reviews

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Score
AbdullahBaghasi. 2018.	Y	Y	U	U	N	N	Y	N	N	N	U	9
Ardoin SP, Christ TJ, Morena LS, Cormier DC, Klingbeil DA. 2013.	U	U	U	Y	N	N	Y	Y	N	Y	Y	13
Arrimada, Torrance and Fidalgo, 2020.	U	U	U	U	U	U	N	U	N	N	Y	9
Asaro-Saddler K, Moeyaert M, Xu X, Yerden X. 2021.	U	U	U	U	N	N	Y	Y	N	Y	Y	11
Barrio BL, Lindo EJ, Combes BH, Hovey KA. 2015.	U	Y	U	Y	N	N	Y	U	Y	Y	U	13
Berkeley S, Scruggs TE, Mastropieri MA. 2010.	U	Y	U	Y	N	N	Y	Y	N	Y	Y	14
Boon RT, Barbetta PM, Paal M. 2018.	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	21
Bouck EC, Park J. 2018.	Y	Y	U	U	Y	U	Y	U	N	Y	Y	16
Browder DMW.	N	U	U	Y	Y	Y	Y	Y	U	Y	Y	17
Burke MD, Boon RT, Hatton H, Bowman-Perrott L. 2015.	U	Y	U	U	N	N	Y	Y	N	Y	Y	13
Burns MK, Appleton JJ, Stehouwer JD. 2005.	Y	Y	U	U	N	N	Y	Y	N	U	Y	13
Burns MK, Symington T. 2002.	U	Y	Y	Y	U	U	U	Y	Y	Y	Y	18
Ciullo SR and Reutebuch, C. 2013.	Y	Y	Y	U	N	N	Y	Y	N	Y	Y	15
Collins AA, Lindström ER, Compton DL. 2018.	Y	U	U	Y	N	N	Y	Y	Y	Y	U	14
Cook KB, Bennett KE. 2014.[50]	Y	Y	U	U	Y	Y	Y	Y	N	Y	Y	18
Daniel J, Capin P, Steinle P. 2021.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	22
Datchuk SM, Rodgers DB, Wagner K, Hier BO, Moore CT. 2022.	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	20
Datchuk SM, Wagner K, Hier BO. 2020.	Y	Y	U	Y	Y	Y	Y	U	Y	Y	Y	20
Dietrichson J, Filges T, Klokker RH, Viinholt BCA, Bøg M, Jensen UH. 2020.	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	20

50 While this study passed the critical appraisal, fewer than half of the studies included in this systematic review met quality standards. See Appendix A.8. This study was therefore excluded from our umbrella review on that basis.

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Score
Donker AS, de Boer H, Kostons D, Dignath van Ewijk CC, van der Werf MPC. 2014.	U	Y	U	U	N	N	Y	Y	Y	Y	U	13
Edmonds MS, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett KK, et al. 2009.	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	21
El Zein F, Solis M, Vaughn S, McCulley L. 2014.	Y	Y	N	Y	Y	U	Y	Y	Y	Y	Y	19
Elleman AM, Oslund EL, Griffin NM, Myers KE. 2019.	N	U	U	U	N	N	N	N	N	U	Y	6
Fallon LM, Collier-Meek MA, Maggin DM, Sanetti LMH, Johnson AH. 2015.	N	U	U	Y	Y	Y	Y	U	Y	U	Y	16
Filderman MJ, Austin CR, Boucher AN, O'Donnell K, Swanson EA. 2022.	Y	U	Y	U	Y	Y	Y	Y	Y	Y	Y	20
Filderman MJ, Toste JR, Didion L, Peng P. 2021.	Y	Y	Y	U	U	U	Y	Y	Y	Y	Y	19
Flynn LJ, Zheng X, Swanson HL. 2012.	U	Y	Y	U	N	N	Y	Y	N	Y	Y	14
Garwood JD, Brunsting NC, Fox LC. 2014.	Y	Y	Y	U	N	N	U	U	Y	Y	Y	15
Gillespie A, Graham S. 2014.	Y	Y	Y	Y	Y	U	Y	Y	Y	U	Y	20
Goodwin AP, Ahn S. 2013.	N	Y	U	Y	N	N	Y	U	Y	Y	U	13
Graham S, Liu X, Bartlett B, Ng C, Harris KR, Aitken A, et al. 2018.	Y	Y	Y	Y	U	Y	U	Y	Y	Y	Y	20
Hall and Burns, (2018)	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	21
Hall-Mills SS, Marante LM. 2022.	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	20
Huddle S, Hosp J, Watt S. 2017.	Y	Y	U	U	Y	Y	Y	Y	N	U	Y	17
Jitendra A, Xin YP. 1997.	Y	U	N	N	N	N	N	N	N	Y	Y	7
Jitendra AK, Lein AE, Im S, Alghamdi AA, Hefte SB, Mouanoutoua J. 2018.	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	18
Joseph LM, Schisler R. 2009.	U	Y	Y	U	U	N	N	Y	N	Y	Y	12
Jung P, McMaster KL, Kunkel AK, Shin J, Stecker PM. 2018.	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	20
Kaldenberg ER, Ganzeveld P, Hosp JL, Rodgers DB. 2016.	Y	Y	Y	U	N	N	Y	Y	Y	Y	Y	17
Kaldenberg ER, Watt SJ, Therrien WJ. 2015.	Y	Y	Y	U	Y	Y	Y	Y	U	Y	Y	19

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Score
Kang EY, McKenna JW, Arden S, Ciullo S. 2015.	Y	Y	Y	U	Y	Y	Y	Y	N	Y	Y	19
Kilgus SP, Methe SA, Maggin DM, Tomasula JL. 2014.	U	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	19
Kim A-H VS, Wanzek J WS. 2004.	N	Y	U	U	N	N	Y	Y	N	U	Y	11
Kim W, Linan-Thompson S, Misquitta R. 2012.	Y	Y	Y	U	N	N	Y	Y	U	Y	U	15
Kuder SJ. 2017.	N	Y	U	U	N	N	Y	Y	N	Y	Y	12
Lee J, Bryant D, Ok M, Shin M. 2020.	Y	Y	Y	U	N	N	Y	Y	N	Y	Y	15
Lee J, Yoon SY. 2017.	Y	Y	U	Y	N	N	N	Y	Y	U	U	13
Lein AE, Jitendra AK, Harwell MR. 2020.	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	18
Maccini P, Mulcahy CA, Wilson MG. 2007.	U	Y	Y	Y	U	N	Y	Y	U	Y	Y	17
McClain MB, Haverkamp CR, Benallie KJ, Schwartz SE, Simonsmeier V. 2021.	Y	U	Y	U	Y	U	Y	Y	Y	Y	Y	19
Newell KW, Codding RS, Fortune TW. 2020.	U	Y	U	Y	Y	Y	Y	U	Y	Y	Y	19
Oxley and de Cat C. 2021.	N	Y	U	U	N	N	Y	Y	Y	U	U	12
Paul S-AS, Clarke PJ. 2016.	Y	Y	U	U	N	N	U	N	N	N	U	8
Peterson AK, Fox CB, Israelsen M. 2020.	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	21
Reed DK, Cummings KD, Schaper A, Biancarosa G. 2014.	Y	Y	U	U	U	U	Y	Y	N	Y	Y	16
Roberts CA, Leko MM, Wilkerson KL. 2013.	N	Y	U	Y	N	N	U	N	Y	Y	Y	12
Scammacca N, Roberts G, Vaughn S, Edmonds M, Wexler J, Reutebuch CK, et al. 2007.	Y	Y	U	U	N	N	U	Y	N	Y	Y	13
Scammacca NK, Roberts G, Vaughn S, Stuebing KK. 2015.	Y	Y	U	U	U	Y	Y	Y	Y	Y	Y	19
See BH, Gorard S. 2014.	U	U	Y	Y	N	N	N	N	U	U	U	9
Sencibaugh JM. 2007.	N	U	U	U	N	N	N/A	Y	N	N	N/A	5
Shin J. 2017.	U	Y	U	Y	N	N	Y	Y	Y	Y	Y	16
Shin M, Bryant DP, Powell SR, Jung P-G, Ok MW, Hou F. 2021.	U	Y	U	U	Y	Y	Y	Y	U	Y	Y	18
Shin M, Bryant DP. 2015.	Y	Y	U	Y	N	N	Y	Y	Y	Y	Y	17

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Score
Solis M, Ciullo S, Vaughn S, Pyle N, Hassaram B, Leroux A. 2012.	Y	Y	U	U	N	N	Y	Y	N	Y	Y	14
Steinle PKS, Stevens E, Vaughn S.	Y	Y	U	U	U	Y	Y	Y	N	N	Y	15
Stevens EA, Park S, Vaughn S. 2019.	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	21
Stevens EA, Rodgers MA, Powell SR. 2018.	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	21
Stewart AA, Austin CR. 2020.	Y	Y	U	U	Y	Y	Y	Y	Y	Y	Y	20
Swanson E, Hairrell A, Kent S, Ciullo S, Wanzek JA, Vaughn S. 2014.	Y	Y	U	U	Y	Y	Y	Y	N	U	Y	17
Valasa LL, Mason LH, Hughes C. 2014.	Y	Y	U	U	Y	N	U	Y	N	Y	Y	15
Wanzek J, Vaughn S, Scammacca NK, Metz K, Murray CS, Roberts G, et al. 2013.	U	Y	U	U	Y	Y	Y	Y	Y	Y	Y	19
Watt SJ, Watkins JR, Abbitt J. 2016.	U	U	U	Y	N	N	Y	Y	U	Y	Y	14
Wexler J, Vaughn S, Edmonds M, Reutebuch CK. 2008.	U	U	U	U	Y	N	U	N	N	Y	Y	11
Williams KJ, Austin CR, Vaughn S. 2018.	Y	Y	Y	U	Y	U	Y	Y	Y	U	Y	19
Xin YP, Jitendra AK. 1999.	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	18
%	61.64	82.19	43.83	43.83	41.09	32.87	80.82	82.19	52.05	80.82	86.3	

A.7. Critical appraisal results from systematic reviews of text and opinion

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Score
Berkeley S, Scanlon D, Bailey TR, Sutton JC, Sacco DM. 2020.	Y	Y	Y	Y	Y	Y	6
Berkeley S. 2009.	Y	Y	Y	Y	Y	Y	6
%	100.0	100.0	100.0	100.0	100.0	100.0	

A.8. Data extraction table

Study	Ardoin SP, Christ TJ, Morena LS, Cormier DC, Klingbeil DA. 2013
Review objectives	To identify and summarise the psychometric and empirical evidence for Curriculum-Based Measurement in Reading (CBM-R) as it is used to monitor and evaluate student progress.
Descriptions of interventions/ phenomena of interest	The accuracy of CBM-R decision rules to monitor and evaluate individual student progress in oral reading fluency.
Descriptions of outcomes included in the review	1. The number of recommended data points that should be collected prior to decision making (count). 2. The accuracy of any CBM-R decision rule as measured through correspondence with another measure of student growth. 3. The accuracy of estimated rates of growth at the level of the individual child through evaluation of the error evaluated in individual student growth estimates.
Descriptions of contexts included in the review	Journal articles, book chapters and instructional manuals.
Search details	Three databases: ERIC, PsycINFO and Academic Search Complete. An ancestral review was also conducted.
Number of studies and participants included	102 documents were included (journal articles=75, chapters/books=18 and other=9). Of the 102 documents (47 empirical articles and 55 manuals), 78 documents discussed using CBM-R data for progress monitoring data and 60 of those documents explicitly stated how to employ one or more CBM-R decision rules.
Appraisal instruments used	Not reported.
Description of main result	• Most decision-making practices are based on expert opinion and there is very limited psychometric or empirical support for such practices. The overwhelming majority of authors discussed or recommended the trend line decision rules and OLS regression was the most common among them. A number of documents also referenced the data point decision rules. There was, however, scant evidence for either procedure. • There is a lack of published evidence to support program evaluation and progress monitoring with CBM-R. There is robust evidence that CBM-R is useful for summative assessment purposes, such as screening and benchmarking however, such evidence does not confer robust support for the interpretation and use of CBM-R progress monitoring practices as they are used to evaluate instructional effects for individual students. On the contrary, evidence has recently emerged to suggest that CBM-R progress monitoring outcomes lack reliability and validity unless extensive data are collected for an extended time period (for example, Christ, Zopluoglu, Long and Monaghen, 2012).

Study	Asaro-Saddler K, Moeyaert M, Xu X, Yerden X. 2021
Review objectives	To determine whether the: 1. self-regulated strategy development (SRSD) approach to teaching writing to students with autism spectrum disorder (ASD) improves significantly the number of words written and overall quality of writing 2. effects of SRSD were consistent or variable across studies, and whether the moderator of age could explain the potential variability in the number of words written and overall quality of writing.
Descriptions of interventions/ phenomena of interest	The impact of SRSD on student writing.
Descriptions of outcomes included in the review	Volume/length of writing holistic quality of writing.
Descriptions of contexts included in the review	Setting and context were not specified.
Search details	Seven databases included: Academic Search Complete, Education Full Text, Education Research Complete, Education Source, ERIC, Psychology and Behavioural Sciences Collection, and Teacher Reference Center.
Number of studies and participants included	Ten single-case design studies including at least one participant on the autism spectrum. The average age equalled 10 years and 10 months but the total number of participants was unknown.
Appraisal instruments used	Study quality was not assessed. Statistical methods were used to account for or capture: 1. small sample sizes 2. the effectiveness of SRSD across studies in addition to case-specific and study-specific intervention effect estimates 3. the variability in the effectiveness of SRSD between studies and between cases can be captured.
Description of main results	• SRSD was an effective approach to teaching writing to students with ASD. The increase in number of words and holistic quality was large, statistically significant and meaningful. • Older students benefitted more than younger students for number of words. • For holistic quality, all participants benefitted from the intervention.

Study	Barrio BL, Lindo EJ, Combes BH, Hovey KA. 2015
Review objectives	To examine research focusing on the teaching, learning, implementation, and evaluation of RTI in GE teacher preparation during the last decade (2003 – 2013).
Descriptions of interventions/ phenomena of interest	The extent of literature in the field of general educator teacher preparation focused on Response to Intervention and its components. What the highly rated (according to impact factor or general consensus by a panel of experts) general education peer-reviewed journals are publishing in regards to Response to Intervention and its components.
Descriptions of outcomes included in the review	• Demographic information (such as, author, journal and year of publication). • Participants in the study (such as, preservice teachers, in-service teachers, administrators, teacher educators). • Type of publication (such as, quantitative study, qualitative study, interview of scholars, commentary). • Purpose of the study (such as, exploratory, recommendation, historical, literature review, program review). • Grade level focus (such as, elementary school, middle and/or high school, none). • Certification program (such as, GE, special education, other). • Components of RTI addressed (such as, progress monitoring, evidence-based practices, cultural competence, screening, data-based decision making, RTI overall).
Descriptions of contexts included in the review	US peer-reviewed publications.
Search details	10 Databases: Academic Search Complete, ERIC via Ebscohost, Educational Research Complete, JSTOR, Professional Development Collection, ProQuest, Psychinfo, Sage Journals Online, Taylor and Francis Online and Google Scholar. Electronic search of 3 professional associations' databases: the American Association of Colleges for Teacher Education, the Association of Teacher Educators and the American Educational Research Association. 7 peer-reviewed journals in the field of GE and teacher preparation were also hand searched, such assuch as Journal of Teacher Education, Action in Teacher Education, Educational Researcher, Educational Forum, Teaching Education, and Teaching and Teacher Education. Published between January 2003 and May 2013.
Number of studies and participants included	10 publications preservice teachers, in-service teachers, administrators, teacher educators).
Appraisal instruments used	Not reported.

Study	Barrio BL, Lindo EJ, Combes BH, Hovey KA. 2015
Description of main results	• None of the RTI-focused publications in this review targeted preservice teachers as the main audience but instead were written for teacher–educators. By contrast, 33 publications included in this review of the literature were obtained from highly rated journals in special education (for example, Exceptional Children – impact factor: 2.766, Remedial and Special Education – impact factor: 0.890, and Teacher Education and Special Education – impact factor: N/A). • Results of this study show General Education (GE) preservice teachers are not a commonly targeted audience for published literature on RTI and its components. Findings from this study show that the majority of the peer-reviewed journals that met the search criteria focused on teacher educators as their main participants or audience. • The field of education must further explore how best to prepare all educators, especially GE teachers who serve as the initial service providers in the RTI model. The review authors propose that teacher educators to enhance preservice teachers' practical experiences through case-based instruction, action-research projects, and the integration of progress monitoring and implementation of evidence-based practices to their curriculum, and write about these experiences in the form of manuscripts for publication.

Study	Berkeley S, Scruggs TE, Mastropieri MA. 2010
Review objectives	To determine whether obtained effect sizes from more recent research are generally comparable to previous research and whether other differences would be observed, for example, differences in types of treatments.
Descriptions of interventions/ phenomena of interest	Interventions designed to improve student reading comprehension for students in K - 12 with learning disabilities (LD), such as participants specifically identified as having LD (students described as reading disabled or dyslexic were considered synonymous with LD). Reading comprehension interventions classified as: questioning/strategy instruction, text structure, fundamental reading skills and 'other.'
Descriptions of outcomes included in the review	Reading comprehension measures: Norm-referenced measures such as standard scores from the Woodcock Reading Mastery Test and grade equivalents on the Gates–MacGinitie Reading Test (n = 25); criterion-referenced measures such as an informal reading inventory (n = 3) and, informal reports such as the number of years below grade level (n = 5).
Descriptions of contexts included in the review	Elementary (n = 15), middle (n = 18), high (n = 6) schools, and a residential facility for adjudicated youth (n = 1). Treatments were delivered in large groups (42.5%), small groups (35.0%), and one-to-one settings (22.5%); one study did not clearly report the treatment setting.
Search details	Six databases: PsycINFO, ERIC, Expanded Academic, Social Sciences Citation Index, EBSCO, and Digital Dissertation online databases. Ancestral search. Hand search. Date range: 1995 – 2006.

Study	Berkeley S, Scruggs TE, Mastropieri MA. 2010
Number of studies and participants included	Total studies = 40. Most prevalent research design was experimental designs with random assignment of participants to condition (52.5%), followed by quasi-experimental designs with matching or non-random assignment (40.0%), and pre-post designs with the pretest as comparison (7.5%). Total participants = 1734.
Appraisal instruments used	Not reported.
Description of main results	• Interventions on reading comprehension were generally very effective. Questioning/strategy instruction and text enhancements were associated with mean weighted effect sizes that were moderate to large in magnitude. The highest effect sizes were associated with fundamental reading skills instruction, followed by questioning/strategy instruction and text enhancements. • Reading comprehension interventions were more effective for middle/high-school students. Mean treatment effects for criterion-referenced measures for middle and high-school students (0.80) were larger than mean effects for elementary school students (0.52). These means were statistically different. • There were no clear differences for intervention impact by setting (small group, 1:1, classroom). • For criterion-referenced measures, mean weighted treatment effect sizes were highest for treatments of medium duration (more than 1 week but less than 1 month). Differences among treatments of varying length were statistically different according to a homogeneity test, Q(2, N = 30) = 6.68, p = .04. However, differences on norm-referenced tests by study duration were not statistically significant (p = .83). • There were no clear findings regarding the duration of interventions. For criterion-referenced measures, mean weighted treatment effect sizes were highest for treatments of medium duration (more than 1 week but less than 1 month). Differences among treatments of varying length were statistically different according to a homogeneity test, Q(2, N = 30) = 6.68, p = .04. However, differences on norm-referenced tests by study duration were not statistically significant (p = .83). • There were no differences between interventions delivered by researcher, adult tutor or teacher. The mean effect size for researchers (0.83) was higher than the mean effect size for teachers (0.56), although these differences fell slightly short of statistical significance.

Study	Boon RT, Barbetta PM, Paal M. 2018
Review objectives	To present a research synthesis of single-case studies on the efficacy of graphic organiser interventions to improve the writing outcomes of students with learning disabilities in the K-12 classroom.
Descriptions of interventions/ phenomena of interest	Interventions to improve the writing skills of students enrolled in grades K-12 identified with a learning disability as the primary or secondary disability, using as graphic organiser as the sole or primary intervention medium.
Descriptions of outcomes included in the review	Adapted analytic trait-scoring rubrics, writing assessment holistic rubrics, and either researcher-made or adapted state assessment measures. None of the studies applied norm-referenced writing quality measures. Writing outcomes measures of: 1. syntax – number of T-units and overall syntax measures in a composition 2. vocabulary richness – the ratio between the number of different words to total of words 3. text length – number of written words or lines 4. genre elements – using genre elements in a written composition 5. quality measures – holistic assessment of compositions from several dimensions such as clarity, coherence, cohesion, elaboration, flow, organisation and voice etc.
Descriptions of contexts included in the review	Interventions were conducted in a pull-out setting (n = 4; an in-class resource room setting (n = 4), and in both a pull-out and in-class setting (n = 2). Intervention groups were: one-to-one format (n = 5) and, small-group format (n = 5) with group sizes ranging from 2 to 8 students. Administered by a special education teacher (n = 3), a teacher (n = 2, a special education teacher and a special education administrator (n = 1), researchers (n = 3) and, by researchers and a special education teacher (n = 1).
Search details	Initial computerised search using Library Quick Search (ProQuest). Four databases: Academic Search Complete, Education Full Text, ERIC (EBSCOhost) and PsycINFO (ProQuest) and a search of Google Scholar. Hand search. Ancestral search. Date ranges: 1975 to July 2017.
Number of studies and participants included	Total studies = 10 (single-case designs). Total students = 62 enrolled in grades 3 to 12. Males (n = 44) and females (n = 18). Forty-7 students were diagnosed with LD, one had a dual diagnosis of learning disability and ADHD, one had a dual diagnosis of LD and communication disorder (CD), 3 had LD and comorbid emotional and behavioural disorder (EBD), 3 students had LD as a secondary disability, of these, 2 had EBD and the other had an autism spectrum disorder (ASD) as their primary disability, respectively. The remaining 7 participant students were diagnosed with other disabilities, one student had ADHD, another was diagnosed with EBD, 2 students had mild intellectual disabilities (MID), 2 had a dual diagnosis of ASD and ADHD and the other was identified with EBD and ASD.
Appraisal instruments used	What Works Clearinghouse (WWC) design standards for single-case design (SCD) research (Kratochwill et al. 2010, 2013).

Study	Boon RT, Barbetta PM, Paal M. 2018
Description of main results	• All the graphic organiser interventions, with one exception, proved to be beneficial to improve the writing skills of students with learning disabilities on both narrative and expository writings. Among narrative writing studies, a wide range of effect sizes were found (Tau-U range = 0.37 – 0.58; Mean = 0.42). For expository writing studies, the 3 studies, which employed computer-based graphic organisers, yielded large and very large overall effect sizes (Tau-U range = 0.71 – 1.00; Mean = 0.72). Effects on different writing outcomes were not consistent within some studies where the intervention yielded large, moderate and small effects depending upon the writing outcomes (Gonzalez- Ledo et al. 2015; Li, 2007; Zipprich, 1995). • Findings indicated that using graphic organisers improved the writing skills of students with learning disabilities on several dimensions such as writing quality, genre elements, text productivity, syntax and vocabulary. In general, usinga graphic organiser increased the number of elements included in the students' compositions regardless of writing genre, with large effect sizes. On text length and syntax measures, overall effect size results on both measures ranged from moderate to large, for the most part, except for one study, which yielded a negligible impact on syntax (Zipprich, 1995). For the only study assessed vocabulary in students' compositions, findings were inconclusive. • Some graphic organisers were more effective than others. A large omnibus effect size (Tau-U = 0.65) was noted for computer-based graphic organiser interventions compared to a moderate omnibus effect size (Tau-U = 0.37) for paper-and-pencil graphic organiser interventions. The effectiveness of teacher versus student generated graphic organisers indicated that overall both types of graphic organiser interventions were similarly effective, with omnibus effect sizes in the moderate range. • Graphic organisers should not be used in isolation. Using graphic organisers in the writing process should be considered as a supplementary tool of effective writing instruction to help students with LD to visualise and internalise writing principles, knowledge of genre and text structure, and as a scaffold device to assist students to generate content during the planning stage and to organise their planning ideas during the drafting stage. Gaps: There is limited research on graphic organisers and their effects on expository writing.

Study	Bouck EC, Park J. 2018
Review objectives	To know the current state of research on manipulatives for students with disabilities and identify the implications for teachers.
Descriptions of interventions/ phenomena of interest	Use of manipulatives as a tool to support students with disabilities in mathematics.
Descriptions of outcomes included in the review	To aid in understanding or solving mathematics problems.

Study	Bouck EC, Park J. 2018
Descriptions of contexts included in the review	The settings for the studies were schools: a public elementary, middle or high school. One study in a clinic.
Search details	Four databases: ProQuest, EBSCO, PsycInfo, and ERIC. Also, hand search of 5 relevant journals plus ancestral search.
Number of studies and participants included	36 studies in total. 16 Single case design and 13 group designs (7 experimental/ quasi experimental, 6 pre/post-test design), 1029 school students with disability (LD, ID, ASD).
Appraisal instruments used	Different quality indicators for the respective design. Those proposed by Gersten et al. (2005) for group designs and those proposed by Horner et al. (2005), for SCD studies.
Description of main results	Studies on manipulatives: • the effect sizes provided represented strong effects for the use of manipulatives • focused on basic operations, such as subtraction; 2 addressed area and perimeter; and one linear algebra • most commonly reported manipulatives were actually generic concrete objects, rather than specifically purchased manipulatives, and base 10 blocks. Studies on manipulatives within CRA: • 10 of the studies provided a measure of effect size. The majority of effect sizes provided represented strong effects for use of the CRA sequence • basic operations, with subtraction with regrouping and multiplication with regrouping the most common. Other mathematical areas included algebra, place value, fractions, area and perimeter, and money • manipulatives for the basic operations studies included base 10 blocks, Cuisenaire rods, and Rekenrek (such as, calculating frames). Gap: lack of high-quality studies that exist to support manipulatives in and of themselves as well as the CRA instructional sequence. Need for well-designed group comparisons.

Study	Browder DM, Wakeman, SY, Spooner, F, Ahlgrim-Delzell, L, Algozzine, B. 2006
Review objectives	To provide a comprehensive synthesis of the research on reading for individuals with significant cognitive disabilities.
Descriptions of interventions/ phenomena of interest	Evidence-based practices currently exist for teaching each of the National Reading Panel's (NRP) components of reading to individuals with significant cognitive disabilities. Interventions that targeted teaching reading or picture identification skills as the primary focus and included experimental data.
Descriptions of outcomes included in the review	A count of the number of studies that targeted the components of reading (sight words, phonics/decoding, phonemic awareness, comprehension, fluency), picture identification or other (for example, object and letter identification). The effect size of reading component interventions measured by percentage of nonoverlapping data (PND). The number of reading component intervention studies that met quality criteria.

Study	Browder DM, Wakeman, SY, Spooner, F, Ahlgrim-Delzell, L, Algozzine, B. 2006
Descriptions of contexts included in the review	Most of the studies (n = 86, 67%) took place in research settings or self-contained special education classrooms. Only a few were implemented in general education classrooms (n of studies = 14, 11%), home (n = 13,10%), or community (n = 5, 4%). Ten studies (8%) were conducted in other settings (for example, institution).
Search details	Databases "including" InfoTrac, Masterfile Premier, ERIC, PsychInfo, Academic Search Elite, and Dissertation Abstracts. Hand search. Ancestral search. Date range: 1975 and 2003.
Number of studies and participants included	Total studies = 128. Single subject design = 88 (69%); group design = 40 (31%). Total participants (N = 1,123) included: 61(55%) individuals with moderate intellectual disability, 124 (11%) individuals with severe intellectual disability, 62 (6%) individuals with autism, and 114 (10%) individuals with other cognitive disabilities (for example, unspecified developmental disability); 204 (18%) participants were not specifically identified as moderate or severe intellectual disability. Participants were: elementary age students (5–11; n = 301, 27%); younger adolescents (12–14; n = 168, 15%) or high school/transition age (15–21; n = 100, 9%). Some participants were adults (n = 50, 4%) and a few were in preschool programs (n = 8, 1%). A number of studies included participants (n = 388) for whom an age range was provided. Finally, a group of studies did not indicate the age of the participants (n of participants = 108).
Appraisal instruments used	Indicators proposed for group (Gersten et al. 2005) and single subject (Horner et al. 2005) designs.

Study	Browder DM, Wakeman, SY, Spooner, F, Ahlgrim-Delzell, L, Algozzine, B. 2006
Description of main results	• This review reveals strong evidence for teaching students with significant cognitive disabilities to read sight words using systematic prompting techniques in a repeated (massed) trial format. Sight word vocabulary was the most frequent component found (41 studies) and had the second highest overall percentage of nonoverlapping data (PND) average (85%). Uniquely, this review reveals that the studies on teaching sight words to students with severe intellectual disability are also of sufficient number, quality, and effect size to conclude that this intervention is effective for this population as well as for students with moderate intellectual disability. • This review of studies on teaching picture or symbol identification also revealed that students with severe intellectual disability can learn symbols related to literacy. What is lacking in the evidence to date is how to teach students with either moderate or severe disabilities the other components of reading needed for literacy. Some researchers have measured fluency as error rate, but rarely have researchers taught fluency. • Some evidence exists for teaching comprehension using concrete referents like pictures or an activity to demonstrate understanding of the word. Most of the high-quality studies addressed comprehension using a sight word in the context of a functional activity; others used word-to-picture matching and systematic instruction to make the comprehension responses through massed trial training (for example, to match picture and word) with systematic prompting and fading. While only about half (n = 11) of the 23 studies that measured or taught comprehension to individuals with moderate and severe MR met the criteria for quality, this collection had an overall strong effect size and enough studies to glean some information on evidence-based practice. Gaps: What is lacking in the evidence to date is how to teach students with either moderate or severe disabilities the other components of reading needed for literacy. The least is known about phonics and phonemic awareness. Although studies related to phonics had the highest overall average PND (93%), only 3 studies addressed that component. The need exists for research on how to teach phonemic awareness and phonics to individuals with significant cognitive disabilities. Some researchers have measured fluency as error rate, but rarely have researchers taught fluency. While these studies provide overall evidence for addressing comprehension through concrete means (using a picture or activity) with stimulus control procedures, they rely on question answering. The need exists for research that will expand interventions in comprehension to address the other practices identified by NRP. More research is needed to determine how to promote reading skills across years of instruction. No studies were found demonstrating a longitudinal approach to literacy; all studies spanned a few months of a school year and targeted only one or 2 components of reading.

Study	Burke MD, Boon RT, Hatton H, Bowman-Perrott L. 2015
Review objectives	To provide a quantitative synthesis of the published, peer-reviewed, single-case research literature on reading interventions for students with or at-risk for EBD.
Descriptions of interventions/ phenomena of interest	A reading intervention or program as the independent variable that targeted a certain aspect of reading (for example, phonological awareness, phonics, fluency, comprehension, and/or vocabulary). Fluency interventions: sight-word instruction, choice instructional method, and a contingency strategy. Comprehension interventions: graphic organisers, listening and text maps multicomponent, reading programs such as Corrective Reading program, Teach Your Child to Read in 100 Easy Lessons and the Great Leaps Reading program, plus repeated reading and/or the use of peer-mediated repeated reading instruction.
Descriptions of outcomes included in the review	Improvement according to a variety of outcome measures including: words read correctly errors per minute section/chapter quiz text map completion Gray Oral Reading Test Woodcock reading mastery test.
Descriptions of contexts included in the review	Students were enrolled in Grades 6 through 9 in a middle and/or high school classroom. Interventions were conducted in small-group pull-out settings and self-contained classrooms.
Search details	Two databases: ERIC and Academic Search Complete also conducted archival search and hand search studies included were from 1996 – 2008.
Number of studies and participants included	11 studies single-case design were included in the review. These included 219 phase contrasts from 44 participants. Thirty-five students were identified with an EBD. Of those students, 23 were identified as EBD, and 12 were identified as EBD and had other related disorders, including LD, mild intellectual disability (MID), ADHD, speech language impairment (SLI), conduct disorder (CD), oppositional defiant disorder (ODD), pervasive personality disorder (PPD), depression (D), and other health impairment (OHI). The remaining 9 students included in the studies were identified with other disability categories, 4 students were classified as LD, 4 OHI, and one OHI and ADHD. The overall mean age of the students was 13 years 4 months and ranged from 11 years 4 months to 18 years 2 months.
Appraisal instruments used	NO critical appraisal undertaken TAU-U statistic used to calculate nonoverlap effect size as this is appropriate for use with data that violates assumptions of normality.
Description of main results	• Overall moderate and positive effects were found for students with EBD receiving a range of reading interventions across intervention types included in the review. • An omnibus effect size of .59 with a 95% CI = [.54, .64] across studies was obtained. Gap: The number of studies was insufficient for conducting a moderator analysis as done in other quantitative reviews of the literature (such as, meta-analysis). At this point, the research base for middle/secondary reading interventions for students with EBD is underdeveloped.

Study	Burns MK, Appleton JJ, Stehouwer JD. 2005
Review objectives	To review existing research on RTI and the 4 specific models of RTI currently in practice (Heartland Agency [Iowa] Model, Ohio's Intervention Based Assessment [IBA], Pennsylvania's Instructional Support Teams [IST], and Minneapolis Public School's Problem-Solving Model [MPSM]).
Descriptions of interventions/ phenomena of interest	1. The effectiveness of the large-scale RTI models currently in practice as compared to those developed for research. 2. Whether RTI leads to improved systemic and student outcomes? 3. The average percentage of the student population was determined to have a disability under RTI.
Descriptions of outcomes included in the review	Measures of either individual student learning, such as postintervention and/or growth assessments, or systemic outcomes (for example, number of children identified as LD).
Descriptions of contexts included in the review	Not reported.
Search details	3 databases: PsycINFO, ERIC, and Education Abstracts. Ancestral search. No grey literature. No a priori specified date range.
Number of studies and participants included	Total = 21 articles.
Appraisal instruments used	Not reported.

Study	Burns MK, Appleton JJ, Stehouwer JD. 2005
Description of main results	• Both field-based and university-based RTI models currently in practice led to strong effects, with the former stronger than the latter, but larger effects were found for systemic rather than student outcomes. • Consistently strong effects of RTI implementations were found for RTI models currently in practice (field-based). Whether calculated using Cohen's d or Unbiased Effect Estimates (UEE), field-based efforts consistently demonstrated stronger effects than university research-based efforts. • This study clarified that sites implementing RTI had both improved systemic and student outcomes (UEE = 1.54 and 1.02, respectively). An examination by type of study and outcome revealed that field-based efforts had slightly smaller improvements in student outcomes (but still substantial at UEE = 0.94), but systemic outcomes that were over 3 times larger than research-developed efforts (UEE = 1.80 vs. 0.47, respectively). • On average, less than 2% of the student population was identified as LD among studies examining field-based RTI models. It seems that large-scale implementation of various RTI models led to fewer students identified as LD. Thus, concerns about large numbers of children identified as nonresponsive leading to dramatic increases in LD (Hale et al. 2004) were not validated by the current data. Gaps: The effect of RTI on LD prevalence needs additional empirical scrutiny using controlled rather than quasi-experimental studies. Moreover, controlled studies of systemic outcomes in general appear warranted. Specifically, RCT studies are needed to examine the effect of RTI implementation on referrals to and placements in special education, student time in special education services, and number of students retained in a grade. Implementation fidelity remains a crucial aspect of any intervention or system of interventions. Therefore, the effect of fidelity in implementation on RTI appears to be crucial data.

Study	Burns MK, Symington T. 2002
Review objectives	The primary goal of the current paper is not to empirically investigate the effectiveness of the PIT approach, but to identify areas in need of future research.
Descriptions of interventions/ phenomena of interest	The impact of pre-referral teams on students and school systems.
Descriptions of outcomes included in the review	Student outcomes – included observations of time on task, student task completion, scores on behaviour rating scales, and observations of target behaviour. Systemic variables included referrals to special education, new placements in special education, percentage of referrals that are diagnosed with a disability, number of students retained in a grade, and an increase in consultative or counselling activity by school psychologists.
Descriptions of contexts included in the review	Public schools in the US.
Search details	Three databases: PycINFO, ERIC, and Education Abstracts databases were searched for articles on January 1, 2002.

Study	Burns MK, Symington T. 2002
Number of studies and participants included	The current article conducted an empirical meta-analysis of research on PITs by reviewing 72 articles. 9 of the articles matched the inclusion criteria for the study and 57 effect size (ES) coefficients were computed.
Appraisal instruments used	Not reported
Description of main results	• PIT approach has a strong effect on desired outcomes. • One potentially important finding specifically relevant to future research is the difference between the mean ES coefficients for randomised and nonrandomised studies. • Randomised studies resulted in a mean ES coefficient that was well within Cohen's (1988) large effect range, but the .64 coefficient for nonrandomised studies suggests only a medium effect. • The pre-referral intervention team approach had a strong effect on school systems and on students.

Study	Ciullo SR. 2013
Review objectives	A systematic review of the literature for studies in which computer based graphic organisers were used for students in grades K-12 with learning disabilities to identify the effectiveness of the treatments on academic outcomes, and selected integral instructional and methodological features.
Descriptions of interventions/ phenomena of interest	Interventions that utilised technology-based graphic organisers or concept maps for students with LD Foci: 1. Content area learning (social studies) – 5 interventions, written expression – 5 interventions, and comprehension – 2 interventions. 2. Instructional features: degree of instructional support, person implementing treatment, and the instructional delivery process.
Descriptions of outcomes included in the review	Outcomes measures included reading comprehension, content learning (science, social studies, or math), vocabulary, or writing.
Descriptions of contexts included in the review	Grades 4 to12 students. Implementers of treatments: teachers – 9 studies (75%), researchers – 3 studies (25%). Treatment times varied from 2 academic years, to less than a week of instruction. Instructional dosage was recorded in sessions that ranged from 30 to 90 minutes. Research design: Experimental (2), quasi-experimental (2), single-case (3), and single-group design studies (5). Fidelity of implementation (4): Instructional features: support provided for completing the information on the graphic organiser, the person implementing each treatment, instructional delivery (principles of effective instruction for students with LD were included or not).

Study	Ciullo SR. 2013
Search details	1. Electronic database search of ERIC and PyschINFO. 2. Hand-search of the following journals (2009–2012): Journal of Learning Disabilities, Journal of Special Education Technology, Learning Disabilities – A Contemporary Journal, Learning Disabilities Research and Practice, and Learning Disability Quarterly. 3. Examination of previously published systematic reviews of the literature for interventions that included graphic organisers and/or technology across all academic subjects. Date range: "earliest available" through February 2012.
Number of studies and participants included	12 studies: 8 studies conducted with students classified as LD in secondary school (Grades 6–12); 3 for students in upper elementary grades; 1 included students in upper elementary grades and middle school. 162 students with LD.
Appraisal instruments used	None reported
Description of main results	Overall: Findings revealed high effect sizes on social studies measures and encouraging results for written expression, while comprehension results were less promising. No evidence was found suggesting that these treatments were efficacious without using explicit instruction and guided practice. Main findings for interventions that utilised technology-based graphic organisers or concept maps for students with LD. • Content acquisition interventions in social studies were associated with encouraging outcomes. Some promising advances for writing were present, and comprehension results were less promising. • Studies with the highest effect sizes incorporated principles of effective instruction for students with LD and literacy recommendations from the IES. (for example, The investigations in social studies that integrated systematic routines including explicit instruction, guided practice, extended practice opportunities, allocated time for studying content, and feedback were associated with learning advances.). • There was no evidence to suggest that computer-based mapping was effective for improving learning, without principles of explicit instruction. Teacher implemented interventions appeared to have more positive outcomes than researcher-led interventions. (For example, high-effect sizes were present in the teacher implemented studies (ES range = .84–2.41), positive gains were evident in 2 single-group interventions taught by teachers, special education and general education teachers collaboratively taught in 3 studies with promising outcomes. However, there were positive effects in one study in which the researcher implemented interventions with high PND across all students, and 2 researcher-led interventions were less effective for comprehension and writing.).

Study	Collins AA, Lindström ER, Compton DL. 2018
Review objectives	To further explore how different assessment methods and reader characteristics, especially students with reading difficulties (RD) and typically developing (TD) students, contribute to variability in reading comprehension test scores.
Descriptions of interventions/ phenomena of interest	1. Differences in reading comprehension test scores between students with RD and TD students across 6 response formats: multiple choice, cloze, sentence verification, open-ended questions, retell, or picture selection. 2. Moderating effects of other assessment methods (for example, time limits, text genre) and reader characteristics (for example, grade/age, methods for identifying students with RD) on response format when comparing score differences between students with RD and TD students.
Descriptions of outcomes included in the review	Student scores on reading comprehension tests
Descriptions of contexts included in the review	Individual administration, small group, in class/large group, not specified
Search details	Eight databases: PsycINFO, ERIC, ProQuest Dissertations and Theses, ProQuest Dissertations and Theses UK and Ireland, International Bibliography of the Social Sciences, PsycARTICLES, Social Services Abstracts, and Sociological Abstracts; also, Google Scholar. Ancestral search. Hand search. Grey literature. Date range: 1975 – 2014.
Number of studies and participants included	82 studies. Participants in Grades K - 12. A total of 5,055 students with RD and 6,454 TD students.
Appraisal instruments used	Not reported

Study	Collins AA, Lindström ER, Compton DL. 2018
Description of main results	• Students with RD performed, on average, less well than TD students across all 6 reading comprehension response formats. Specifically, the magnitude of average weighted effect sizes for students with RD and TD students varied by response format. All were statistically significant except the average weighted effect size estimate for sentence verification. • The achievement gap between students with RD and TD students was larger for picture selection and open-ended questions than for retell tasks. open-ended questions, multiple choice, and picture selection. The achievement gap narrows for response formats pertaining to low-level comprehension tasks (such as, cloze and sentence verification). • Certain assessment methods may correspond to outcomes on some response formats, such as time limits on multiple choice or basal/ceiling rules on open-ended question tests. However, although some variables significantly predicted variance in effect sizes on one response format, none predicted heterogeneity in effect sizes on more than one response format. Moreover, none of the reader characteristic variables explained the observed heterogeneity in study effect sizes. Gaps: Future studies should further investigate the effects of response format on the reading comprehension test scores of students with RD, with attention to how reader characteristics and assessment methods moderate student outcomes. It is critical for future investigations on reading comprehension assessments to report fidelity of test implementation, as well as reliability and validity of measures. Inconsistent findings across meta-regression models may warrant future investigations of moderating factors beyond reader characteristics and assessment methods.

Study	Cook KB, Bennett KE. 2014
Review objectives	To review writing interventions for high-school students with disabilities to identify those practices that have been shown to be effective in supporting writing development.
Descriptions of interventions/ phenomena of interest	Interventions focused on improving written expression (for example, planning, composing, editing, revising) specifically conducted with high-school students with disabilities – specific writing interventions used and disabilities targeted. Evaluated single-case research on writing interventions for high-school students with disabilities.
Descriptions of outcomes included in the review	1. Writing quality (not consistently defined or measured across studies): Likert-type scoring procedures, rubrics, or scoring guides (n = 7). 2. Number of essay elements (n = 7). 3. Number of words (n = 5). 4. Other: Correct word sequences (CWS; n = 3), paragraph content (n = 1), percent of descriptive words (n = 1) and number of lexical ties (n = 1).

Study	Cook KB, Bennett KE. 2014
Descriptions of contexts included in the review	Public high schools (n = 12); private office (n = 1); private school for students with disabilities (n = 1). US (n = 13) and New Zealand (n = 1). Implementers: Special education teachers (n = 3); a researcher who was also a teacher (n = 3); researchers (n = 8).
Search details	Five databases: Academic Search Complete, Education Research Complete, Eric, Medline, and PsychINFO. Hand search. Ancestral search. Date range: 1965 to 2011.
Number of studies and participants included	Total studies = 14. Total participants = 51 (31 males and 20 females) high-school students (Grades 9 – 12) and 14 to 19 years old. Participants were identified with LD (n = 27), ID or cognitive disabilities (n = 12), attention deficit hyperactivity disorders (ADHD; n = 5), emotional or behavioural disorders (EBD; n = 4), Asperger syndrome (AS; n = 2), or speech-language impairments (n = 1).
Appraisal instruments used	What Works Clearinghouse (WWC) single-case design standards (Kratochwill et al. 2010)
Description of main results	• No practice met the minimum WWC requirements for an overall summary rating. Only 6 of the 14 single-case design studies met WWC design standards with reservations, and of those, only 2 of 7 SRSD studies and the 1 pentop computer study demonstrated sufficient evidence of experimental control with an outcome variable (such as, Strong or Moderate Evidence of Causal Relations). Three SRSD studies that were reported as effective in previous reviews failed to meet evidence standards for single-case designs under WWC standards. The WWC standards are not and should not be the definitive tool for determining effective interventions in single-case design research. • SRSD holds the most promise as an effective writing intervention that will also promote maintenance of writing skills for compare–contrast and persuasive essays. Two SRSD programs could be tentatively recommended as positively affecting writing outcomes for high-school students with disabilities. The SDSR studies showed strong effects for compare–contrast (increased use of text structures, such as introduction, comparisons, details, summary, and conclusion) and moderate effects for persuasive essay writing (increased number of essential elements, including providing the context for the problem, a problem statement, support for the position taken, and a concluding statement). Skills were for up to 8 weeks after the intervention supports were removed. • The pentop computer technology could be tentatively recommended as positively affecting writing outcomes for high-school students with disabilities. This intervention shown to have moderate effects by a study that met WWC standards with reservations and resulted in an improvement in the overall quality of both compare–contrast and persuasive essays (Bouck et al. 2010). Quality ratings on students' planning, organization, content, style, and grammar showed immediate improvements when the pentop computer was used to plan and organise essays. Participants did not maintain quality in their essay writing after discontinuing the use of the pentop computer. Gaps: Even though a small number of writing practices have demonstrated tentative effectiveness, there is a continued need for more systematic and well-executed research to guide educators in best practices of writing instruction and intervention.

Study	Daniel J, Capin P, Steinle P. 2021
Review objectives	To examine follow-up intervention effects of reading interventions involving adolescent struggling readers in Grades 6 to 12.
Descriptions of interventions/ phenomena of interest	The effects of reading interventions provided in small-group settings on reading outcomes for struggling readers in Grades 6 to 12 at immediate post-test and follow-up time points. Intervention studies targeting English language reading-related skills, such as decoding, fluency, vocabulary, reading comprehension, or multicomponent reading interventions.
Descriptions of outcomes included in the review	25 different reading measures across 10 studies: 10 were standardised norm-referenced reading measures and included standardised measures of reading comprehension, reading fluency, word reading, and vocabulary. The 15 researcher-developed reading comprehension measures required students to read text and either generate a main idea statement or answer open-ended or multiple-choice questions.
Descriptions of contexts included in the review	Reading-related interventions conducted in school settings outside the general education classroom.
Search details	Four databases: Education source, Education Resources Information Center (ERIC), PsycINFO, and ProQuest Dissertation and Theses Global to locate unpublished and published studies. Date range: 1996 and August 2019. A hand search of 14 prominent educational journals spanning from January 2017 through August 2019. Ancestry searches of articles that fit the inclusion criteria, and of existing reviews.
Number of studies and participants included	Total studies = 10; 6 peer-reviewed articles and 4 dissertations. Total participants = 856 (immediate post-test data) and 693 (follow-up data) adolescent struggling readers. Of these, at post-test, 263 students were identified as having an LD while 593 were identified as struggling readers. LD participants were selected due to their school/district identification of LD (k = 6); struggling readers without an LD identification were below grade level on a standardised or state-based reading measure. Participants were in grades 6 -9 (k = 9) and grade 10 (k = 1).
Appraisal instruments used	Guide for Education-Related Intervention Study Syntheses (Vaughn et al. 2014)

Study	Daniel J, Capin P, Steinle P. 2021
Description of main results	• Remedial reading interventions for struggling adolescent readers showed a large significant effect at post-test, which, on average, reduced to a moderate effect at follow-up. Across all studies, the effect of treatment was g[w] = 0.78 at post-test and g[w] = 0.27 at follow-up. Of the 25 reading measures students were assessed on, 15 were researcher-developed reading measures. For researcher developed reading measures, weighted ES between the treatment and control groups at immediate post-test was w = 0.86, 95% CI [0.30, 1.42], and at the follow- up time point, it was 0.35, 95% CI [-0.20, 0.91], For standardised measures, weighted ES at immediate post-test was g = 0.05, 95% CI [-0.15, 0.25]. No mean effect on standardised measures at follow-up was calculated because only 2 of the 4 studies that administered standardised measures collected follow-up data on control group students. • In general, providing intensive reading comprehension strategy instruction, either with or without direct instruction, was beneficial to students' progress in reading. The estimated average weighted ES between treatment and control groups on all reading comprehension measures at immediate post-test was g[w] = 0.67, 95% CI[0.10, 1.25], and at follow-up was g[w] = 0.33, 95% CI [-0.19, 0.85]. Most studies (n = 5) focused on improving students' comprehension of expository or narrative texts. All 5 studies taught students to use various comprehension strategies, however, only 3 studies reported employing a combination of direct (such as, modelling, guided, and independent practice) and strategy instruction. Of the 26 ESs measured to compare treatment and control groups' performance at immediate post-test and follow-up, 14 were positive and significant in favour of the treatment group. The CIs of these 14 ESs overlapped with 12 follow-up ES CIs, denoting sustainability, and stability of intervention effects. The median follow-up time was 6 weeks (range = 2 weeks to 2 years). • Reading comprehension–related interventions for 6 or more hours generally reported positive effects of intervention on reading comprehension measures at immediate post-test and follow-up time points. 3 studies provided 6 to 10 hr of comprehension-related interventions, and follow-up results varied for different measures. One study (n = 59) reported stable maintenance effects at the follow-up time point on a non-standardised measure of main idea summarisation (g = 1.12); however, no significant differences were observed for students in treatment and control groups at immediate post-test and follow-up time points on another non-standardised measures. Another study (n = 33) reported stable positive maintenance effects at follow-up on a researcher-developed near transfer measure of comprehension (g = 1.42), but the magnitude of difference on a researcher-developed far transfer measure was only significant at immediate post-test (g = 2.27) and not at the follow-up time point measure. The third study (n = 236) moderate positive effects of intervention on researcher developed (g = 0.34, 0.49, n = 226) and standardised measures (g = 0.33; n = 226) at the follow-up time point. Gaps: Due to the small number of studies included in this synthesis, it was not possible to conduct moderator analyses to analyze intervention elements that influenced the strength of association between treatment and follow-up effects. More studies are needed to substantiate the claim that effects of interventions are sustained at follow-up time points.

Study	Datchuk SM, Rodgers DB, Wagner K, Hier BO, Moore CT. 2022
Review objectives	To examine the effects of writing interventions on the writing fluency of students with and without disabilities.
Descriptions of interventions/ phenomena of interest	Impact of skill/strategy acquisition intervention on writing fluency Impact of fluency intervention on writing fluency.
Descriptions of outcomes included in the review	Total number of words written per minute.
Descriptions of contexts included in the review	Primary and secondary school students.
Search details	Four databases: ERIC; PsycINFO, Web of Science and Google Scholar Grey literature (dissertations) included. Hand search of journals Ancestral search of reference lists of included articles 1980 – 2021.
Number of studies and participants included	42 Single Case Experimental Designs: • 15 fluency studies • 27 acquisition studies.
Appraisal instruments used	What Works Clearinghouse standards for single case designs.
Description of main results	Writing interventions had a positive effect on the level and trend of TWWpM across studies. Acquisition interventions: • acquisition interventions improved writing fluency, and had an average gain of 0.16 TWWpM per session • A variety of acquisition interventions were used, with SRSD and graphic organiser instruction being the most frequent. Both instructional approaches (such as, SRSD and graphic organiser instruction) were primarily used to improve aspects of writing other than TWW, such as genre elements (for example, essay parts) and holistic writing quality, meaning gains in TWW largely resulted as an auxiliary benefit from focusing on other aspects of writing. Fluency interventions: – positive effects for fluency interventions with an average gain of 0.10 TWWpM per session • fluency interventions featured timed practice paired with one or more components of initial instruction (for example, explicit instruction or SRSD), performance feedback, goal setting, and graphing of performance) Found writing fluency varied by age but did not substantially vary by other student demographics of disability status, gender, and race. Specifically, older students (such as, above the mean age of 12 years old) tended to show higher levels of TWWpM than younger students.

Study	Datchuk SM, Wagner K, Hier BO. 2020
Review objectives	To examine level and trend effects of writing interventions on the text-writing sequences of students with disabilities and writing difficulties.
Descriptions of interventions/ phenomena of interest	The effects of intervention (direct instruction or SRSD) on the level and trend of text-writing sequences of students with disabilities and writing difficulties. Also interested in whether effects differ by demographic characteristics related to disability, age, gender, and race or ethnic background.
Descriptions of outcomes included in the review	CWSpM (Correctly Writing Sentences per Minute)
Descriptions of contexts included in the review	Schools-non-specified
Search details	Two databases: ERIC and PsycINFO no restrictions on the type of publication or date range of included studies: 2005-2018
Number of studies and participants included	Reviewed 18 single-case experimental design studies (SCED) studies with repeated measurement and multiple opportunities to detect an experimental effect. We excluded group-design studies because these lacked student-level data needed to estimate level and trend effects. The 18 studies had a total of 96 students. The majority of students were White, were male, were of middle-school age, and had a disability. The average age of students was 12.6 years old (SD = 2.6) with a range of 7 years old to 18 years old. Subsequently meta-analysed 15 of these studies with a total of 79 students using mixed-effects linear regression and an information-theoretic ranking of competing models.
Appraisal instruments used	What Works Clearinghouse pilot single-case design standards. All multiple-baseline and multiple-probe designs except 2 (Datchuk and Dembek, 2018; Lewandowski, 2011), as well as the one study that used an alternating-treatment design (Schneider et al. 2013), met or exceeded this standard. All studies reported fidelity and interobserver agreement.

Study	Datchuk SM, Wagner K, Hier BO. 2020
Description of main results	• Positive intervention effects were found across studies. • The number of writing sequences gradually trended upward by 0.39, CWSpM per intervention session, a gain of approximately 4 CWSpM per 10 sessions of intervention. This growth is appreciable given that students tend to gain a maximum of 0.44 CWS per 3 min per week (such as, 0.15 CWSpM per week). • Different interventions produced similar effects on writing fluency. Although interventions produced similar effects on writing fluency, interventions using direct instruction tended to last a shorter duration than SRSD, indicating a difference in efficiency - have to last approximately 368 min (such as, 36.8 min per session). This difference in efficiency highlights an important distinction in the focus of studies that used direct instruction or an SRSD approach all students - regardless of disability type, gender, or race – may benefit equally from writing intervention. • The type of disability also did not significantly contribute to the variance in writing fluency; however, given the small number of students within each disability type, only the 2 most-frequently-occurring disability types were examined (such as, LD and EBD). • The presence of a disability was found not to be a substantial moderator above and beyond the presence of a writing difficulty. • One demographic variable, age, was found to explain variance in text writing fluency intervention impact. • Results suggest that older students tended to produce a higher CWSpM rate (such as, average of .78 CWSpM) during sessions compared to younger students. Younger students, however, showed steeper trends during intervention, indicating they benefited more from intervention. • Results suggest differential effects for type of writing task; students tended to show higher rates of CWSpM on sentence tasks compared to discourse tasks of narratives and essays. Specifically, scores on sentence tasks tended to be approximately 1.3 CWSpM higher than discourse tasks during intervention. • Although the research literature on measuring writing fluency highlights multiple ways to measure writing sequences (for example, McMaster et al. 2009), only a few studies used measures other than CWS, such as CIWS and %CWS. • In the majority of studies, text-writing sequences were measured at the discourse level (such as, essay or narrative), and most studies used a standard definition of CWS as 2 adjacent words within the context of a phrase or sentence (Videen et al. 1982).

Study	Dietrichson J, Filges T, Klokker RH, Viinholt BCA, Bøg M, Jensen UH. 2020
Review objectives	To assess the effectiveness of a broad range of school-based interventions targeting students with, or at risk of, academic difficulties on standardised tests in reading and maths.
Descriptions of interventions/ phenomena of interest	Targeted interventions that aim to improve achievement for students either with academic difficulties, or are deemed at risk of such difficulties on the basis of their background, and not all students.
Descriptions of outcomes included in the review	Scores on standardised tests of mathematics and reading.
Descriptions of contexts included in the review	Students in grades 7–12 in regular schools in OECD countries (US, Canada, UK, Germany, Netherlands, and Australia).
Search details	14 databases and 7 national repositories. Grey literature search Hand search Ancestral search Date range: 1980 to July 2018.
Number of studies and participants included	71 studies: 59 are from the US, 4 from Canada, 3 from the UK, 2 from Germany, 2 from the Netherlands and one from Australia. Fifty-2 studies are randomised controlled trials (RCTs) and 19 are quasi-experimental studies (QESs). Total number of student observations was around 105,700. The target group consisted of, on average, 47% girls, 73% minority students, and 62% low income students.
Appraisal instruments used	Not reported
Description of main results	• The interventions studied have on average positive and statistically significant short-run effects on standardised tests in reading and maths. This effect size (weighted average ES = 0.22) is of an educationally meaningful magnitude, for example, in relation to the gap between groups of at-risk and not at-risk students. This means that the most effective interventions have the potential of making a considerable dent in this gap. • There is little evidence of longer-run effects of the interventions on students' performances on standardised tests in reading and maths. Only 7 included studies tested effects more than 3 months after the end of intervention. The average effect size was small and not statistically significant (ES = 0.05), but there was substantial variation. • Small-group instruction has significantly larger effect sizes than CAI and incentive components. Interventions that included small-group instruction (ES = 0.38), peer-assisted instruction (ES = 0.19), progress monitoring (ES = 0.19), CAI (ES = 0.17) and coaching of personnel (ES = 0.10) had positive and significant average effect sizes. Interventions that provided incentives for students did not have a significant average effect size (ES = 0.05). • Effects are very similar across reading domains. Interventions targeting fluency, vocabulary, multiple reading areas, meta-cognitive, social-emotional, or general academic skills, comprehension, spelling and writing, and decoding had average effect sizes ranging from 0.14 to 0.22, all of them statistically significant.

Study	Dietrichson J, Filges T, Klokker RH, Viinholt BCA, Bøg M, Jensen UH. 2020
Description of main results	• Interventions have larger effects on standardised tests in maths than on reading tests. Effect sizes based on mathematics tests had a relatively large effect size (ES = 0.34). Gaps: We do not know whether the short-run effects are lasting. More research about long-run effects would therefore be a welcome addition to the literature. More research is also needed from non-English speaking countries. There are also more interventions that have been tested by reading tests than maths tests, and interventions targeting maths seem like a promising research agenda. Improvement to research designs are required, for example, increase the number of units and students in intervention and control groups. Detailed descriptions of the instruction given to control groups is required.

Study	Donker AS, de Boer H, Kostons D, Dignath van Ewijk CC, van der Werf MPC. 2014
Review objectives	To examine the results of studies on learning strategy instruction focused on improving self-regulated learning were brought together to determine which specific strategies were the most effective in increasing academic performance.
Descriptions of interventions/ phenomena of interest	The most effective strategies, such as cognitive, metacognitive and management strategies (and their respective sub-strategies), motivational aspects and metacognitive knowledge, for improving academic performance. Moderation on strategy effectiveness of different types of students, type of measurement instrument and the influence of publication bias on results.
Descriptions of outcomes included in the review	Researcher-developed tests and intervention-independent tests, for example, standardised achievement tests, of 'academic achievement' (operationalised as performance in one or more school subject domains)
Descriptions of contexts included in the review	Schools - core academic subjects only. Strategy interventions (total n = 95): Mathematics (n = 44), followed by (comprehensive) reading, writing and science (n = 23, n = 16, and n = 9, respectively).
Search details	Two databases: ERIC and PsychInfo. No ancestral or hand searches. No grey literature. Date range: January 1st 2000 to January 1st 2012.
Number of studies and participants included	Total studies = 58. Primary school or secondary students up to and including the twelfth grade.
Appraisal instruments used	Not reported

Study	Donker AS, de Boer H, Kostons D, Dignath van Ewijk CC, van der Werf MPC. 2014
Description of main results	• The instruction of learning strategies can improve academic performances for all students in primary and secondary schools. The studies' average mean effect size of Hedges' g = .66 (S.E. = 0.05). The meta-analysis of variance revealed no significant between groups differences; the effect sizes show that the learning strategy interventions were highly effective for all types of students. • Certain subject domains appeared to be more suitable for strategy instruction than other course fields: in writing the highest effects were found, regardless of the exact content of the intervention (such as, the strategies taught). Effect sizes were: writing (Hedges' g = 1.25), science (.73), mathematics (.66) and comprehensive reading (.36). • The analyses show significant positive effects of the inclusion of general metacognitive knowledge, the learning strategies planning and rehearsal, and the motivational aspect task value. Interventions which include 'general metacognitive knowledge', 'planning' or 'task value' enhanced student performance the most effectively. In writing the combination of (general) metacognitive knowledge and evaluation was B = .75 (.24) and .57 (.26), respectively ($p < .05$), indicating that interventions which included both strategies were the most effective here. In mathematics, elaboration proved to be very effective. This substrategy entails attaching meaning to new information by connecting the old to the new material. • Comprehensive reading was the only subject in which intervention-independent tests resulted in a significant lower effect than self-developed tests. The average effect size of Hedges g .78 for the self-developed tests and of .45 for the intervention independent tests. This difference was significant ($p < .01$). For writing, math and science, the difference in effect size between the 2 types of tests was not significant. 5. Results indicated that there was some publication bias in the domain of science, whereas overall it was low. The effects we presented seem to be representative of the outcomes which would have been obtained if more studies had been included. Gaps: What we do not know is whether the same learning strategies are instructed in different ways to regular and disabled students. What would be interesting is to first differentiate among student groups and then conduct the analyses again, to see which strategies in which domain are the most effective for which type of student.

Study	Edmonds MS, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett KK, et al. 2009
Review objectives	To determine the outcome of comprehension, word study, vocabulary, and fluency interventions on reading comprehension of all struggling readers in Grades 6 through 12
Descriptions of interventions/ phenomena of interest	The influence of interventions targeting decoding, fluency, vocabulary, and comprehension on comprehension outcomes for older students (Grades 6 through 12) with reading difficulties or disabilities?

Study	Edmonds MS, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett KK, et al. 2009
Descriptions of outcomes included in the review	Measures of reading comprehension: Researcher developed (n = 9); standardised (n = 7); and "all measures" (n = 13)
Descriptions of contexts included in the review	Middle schools and high schools. Implementers: Researchers (n = 12), teachers (n = 13), both teachers and researchers (n = 2) or not reported (n = 2)
Search details	Two databases: ERIC and PsycINFO. Ancestral search. Hand search. Date range: 1994 to 2004
Number of studies and participants included	Total studies = 29. 17 treatment–comparison, 9 single subject, and 3 single-group design studies. NOTE: When reporting weighted mean effects, only outcomes from studies with treatment–comparison conditions were included. Total participants = 976 students - middle school students (n = 19 studies), high-school students (n = 5), 2 on both middle and high-school students (n = 2), and only students' ages reported (n = 3). Most studies included students with learning or reading disabilities (n = 17) or a combination of both students with and without disabilities (n = 4); only 8 studies included samples of struggling readers without disabilities. Participants were struggling readers. Struggling readers were defined as low achievers or students with unidentified reading difficulties, with dyslexia, and/or with reading, learning, or speech or language disabilities.
Appraisal instruments used	What Works Clearinghouse Design and Implementation Assessment Device (Institute of Education Sciences, 2003)
Description of main results	• A primary finding from this synthesis is that struggling readers in older grades (6 through 12) can improve in their reading comprehension when taught reading comprehension practices. The random effects model showed the weighted average of the difference in comprehension outcomes between students in the treatment conditions and students in the comparison conditions was large (effect size = 0.89; 95% confidence interval (CI) = 0.42, 1.36). That is, students in the treatment conditions scored, on average, more than 2 thirds of a standard deviation higher than students in the comparison conditions on measures of comprehension, and the effect was significantly different from zero. Even when using standardised measures, which offer a more generalised measure of comprehension, the effect was moderate (ES = 0.47), providing students with an average of a half standard deviation advantage compared to their peers without the treatment.

Study	Edmonds MS, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett KK, et al. 2009
Description of main results	• Interventions that developed students' strategy knowledge and use showed large effects whereas other types of interventions showed relatively lower effects of interventions on reading comprehension outcomes for struggling readers. Results of the meta-analysis of a subset of studies with contrasts between a treatment condition and a no-treatment comparison condition (k = 13) and with theoretically similar measures of reading comprehension showed average weighted effect sizes of 1.23 for targeted reading interventions in comprehension (n = 7), 0.72 for multiple reading components (n = 3) and, 0.34 for word reading strategies (word study = 0.34, n = 2; fluency ES = - 0.03, n = 1). Results from this synthesis suggest that older struggling readers benefit from explicit comprehension strategy instruction—that is, modelling and thinking aloud how to self-question and reflect during and after reading and engaging students to become actively involved in monitoring their understanding and processing text meaning. There may be a diminishing relationship between accuracy (for example, word recognition and fluent reading) and comprehension with secondary students. • Overall narrative text was associated with higher effect sizes from comprehension interventions than expository text. In an analysis of studies that reported the type of text used, the weighted average effect for interventions using expository text was moderate (n = 3, effect size = 0.53), whereas the average effect for those focusing on narrative text was high (n = 6, effect size = 1.30). Eleven of the 13 studies included in the meta-analysis used reading of connected text as part of the intervention. • Reading comprehension outcomes were higher when interventions were implemented by researchers in contrast to when implemented by teachers. Whether an intervention was implemented by the researcher (n = 4, average effect size = 1.15) or the students' teacher (n = 8, effect size = 0.77), the effects were large. The 95% CIs for these 2 conditions did not overlap, suggesting that they were significantly different. 5. Interventions that specifically targeted students with learning disabilities were associated with the highest gains in reading comprehension. Effects on comprehension were different depending on the student population. Moderate average effects were found for samples of struggling readers (n = 5, effect size = 0.45) or both struggling readers and students with disabilities (n = 4, effect size = 0.68), but a large effect (n = 4, effect size = 1.50) was found for studies with samples of only students with disabilities.

Study	El Zein F, Solis M, Vaughn S, McCulley L. 2014
Review objectives	To examine the effects of reading comprehension interventions for students with Autism Spectrum Disorders (ASD)
Descriptions of interventions/ phenomena of interest	The effectiveness of reading comprehension interventions in improving reading comprehension outcomes for students identified with ASD
Descriptions of outcomes included in the review	Researcher-developed "probes" (n = 9 studies), and standardised measures (n = 2), or both (n = 1) of reading comprehension

Study	El Zein F, Solis M, Vaughn S, McCulley L. 2014
Descriptions of contexts included in the review	Not reported
Search details	Five databases: PsychINFO, ERIC, Psychology and Behavioral Sciences Collection, Medline, and ProQuest Dissertation and Theses. Ancestral search. Date range: 1980 - 2012.
Number of studies and participants included	Total studies = 12. Total participants = 110, K - 12 students, including 64 diagnosed with ASD and 4 diagnosed with both attention deficit hyperactivity disorder (ADHD) and an intellectual disability (ID). The remaining 39 participants were typically developing peers of students with ASD. IQ scores of participants ranged from 42 to 125 (which represents a wide range of cognitive abilities).
Appraisal instruments used	Certainty of evidence evaluation system utilised in previous syntheses (Schlosser and Sigafoos 2007).
Description of main results	• Findings from the studies indicate that modifying instructional interventions associated with improved comprehension for students with reading difficulties may improve reading comprehension in students with ASD in Grades K - 12. • The findings from this synthesis indicated a preliminary body of evidence to support using strategy instruction, student grouping practices, and explicit instruction, to improve reading comprehension for students with ASD. However, after appraising the studies for certainty of evidence, 3 were found to be of conclusive evidence, 5 of suggestive evidence, and 4 of inconclusive evidence.

Study	Fallon LM, Collier-Meek MA, Maggin DM, Sanetti LMH, Johnson AH. 2015
Review objectives	To evaluate whether performance feedback (PF) is an evidence-based practice for increasing treatment fidelity according to WWC guidelines (Kratochwill et al. 2010)
Descriptions of interventions/ phenomena of interest	Study elements and characteristics related to implementing PF, the position of the consultant and consultee, and the disability status of the client involved in PF delivery
Descriptions of outcomes included in the review	Number of demonstrations of intervention effects and non-effects, with a ratio of greater than 3:1 indicating evidence without reservations, a ratio of 3:1 indicating evidence with reservations, and a ratio of less than 3:1 indicating no evidence. Specifically, the analyses included the computation of the (a) improvement rate difference (IRD; Parker, Vannest, and Brown, 2009), (b) percentage of nonoverlapping data (PND; Scruggs, Mastropieri, and Casto, 1987), and (c) Busk and Serlin (1992) standardised mean difference (SMD).
Descriptions of contexts included in the review	PF was delivered (a) verbally in person, as an individual or group meeting (for example, RTI meeting), (b) written via paper, (c) written via e-mail, (d) on videotape, (e) via videoconferencing or teleconferencing (for example, Skype), or (f) by another method.
Search details	Five databases: ERIC, PsycARTICLES, PsycINFO, Proquest Dissertation Abstracts, and Scopus. Ancestral search. Date range: 1960 to 2011.
Number of studies and participants included	Total studies = 47 (169 individual cases); single-case research designs

Study	Fallon LM, Collier-Meek MA, Maggin DM, Sanetti LMH, Johnson AH. 2015
Appraisal instruments used	What Works Clearinghouse (WWC; Kratochwill et al. 2010) technical guidelines for single-case design
Description of main results	• The number of studies with strong and moderate evidence reviewed is sufficient to determine that PF can, in fact, be considered an evidence-based practice to increase implementers' treatment fidelity when providing academic and behavioural interventions to students in education settings. Nearly half of eligible cases were determined to meet design standards (n = 81; 47.9%), about a quarter were determined to meet design standards with reservation (n = 45; 26.6%) and the remainder did not meet design standards (n = 43; 25.4%), primarily due to the presence of fewer than 3 opportunities to demonstrate a treatment effect or fewer than 3 data points for each phase. • In studies demonstrating strong to moderate evidence, PF typically included at least an individual, in-person discussion of what is going well and poorly with implementation but often includes much more (such as, graphic review of implementer data, discussion of student outcomes, etc.). Verbal PF was provided in nearly all studies reviewed (n = 28; 96.6%). Often, at least one other element accompanied verbal PF: graphic PF (n = 20; 69.0%), review of student data (n = 20, 69.0%), or problem solving to support implementation (n = 18; 62.1%). PF was most often delivered to consultees in-person during individual meetings (n = 19; 65.5%) or occasionally in the context of group meetings (for example, RTI team meetings, n = 2; 6.9%). • Studies demonstrated that PF could be a daily occurrence or take place once a month and still demonstrate strong to moderate evidence. Several studies provided PF daily (such as, 5 times per week; n = 8; 27.6%), multiple times a week (such as, 2-4 times per week; n = 5, 17.2%), weekly (once per week; n = 6; 20.7%), or monthly (1 time per month; n = 1; 3.4%). In some cases, PF was provided only when needed (such as, when fidelity fell to unacceptable levels; n = 2, 6.9%). Studies that demonstrated strong or moderate evidence also varied greatly in the immediacy with which PF was provided. • Most consultants involved in studies demonstrating evidence were university researchers; consultee participants held a variety of positions. University researchers were graduate student researchers or, professors (n = 20; 69.0%). In other cases, an outside consultant served this role (n = 2; 6.9%). No study utilised a consultant from within the school building. In most studies, the consultees were general education teachers (n = 20; 69.0%) or special educators (n = 9; 31.0%). In other studies, consultees included (a) paraeducators (n = 4; 13.7%), (b) students in undergraduate or graduate programs (n = 3; 10.3%), (c) assistant teachers (n = 1; 3.4%, (d) Head Start teachers (n = 1; 3.5%), (d) social workers (n = 1; 3.5%), or (e) parents (n = 1; 3.5%).

Study	Fallon LM, Collier-Meek MA, Maggin DM, Sanetti LMH, Johnson AH. 2015
Description of main results	• Studies represented a range of client participants, including students with various abilities and learning differences. Gaps: The need for further research around determining or developing an appropriate effect size for examining the relationship between treatment integrity and PF. Specifically, there is little empirical guidance related to (a) how to define intervention steps specifically or more generally, (b) varied methods for assessing treatment fidelity (for example, permanent products, self-report, direct observation), (c) determining "acceptable" versus "unacceptable" levels of treatment fidelity, and (d) how often treatment fidelity should be assessed and supported, among other issues (Gresham,1989; Sanetti et al. 2011; Sanetti and Kratochwill,2009). Without significant research that tackles these conceptual and practical issues, it is unlikely that treatment fidelity will be regularly assessed and that PF will be efficiently and effectively utilised (Cochrane and Laux, 2008; Sanetti and DiGennaro Reed, 2012).

Study	Filderman MJ, Austin CR, Boucher AN, O'Donnell K, Swanson EA. 2022
Review objectives	To understand the relative effects of targeting various component skills instructionally within the context of a reading comprehension intervention on the reading comprehension outcomes of third through twelfth grade struggling readers.
Descriptions of interventions/ phenomena of interest	Reading interventions that provide reading comprehension intervention in isolation. The extent to which participant (grade level), intervention (instructional approach and measure type), and study (quality) characteristics moderate the effects?
Descriptions of outcomes included in the review	Researcher-designed measures or standardised measures of reading interventions.
Descriptions of contexts included in the review	Elementary and secondary schools. Researcher and teacher implemented. Group sizes not reported.
Search details	Four databases: ERIC, PsycINFO, Education Source, and Academic Search Complete. Hand search: Articles published in the past 5 years across major journals. Ancestral search. Date range: January 1975 to March 2020.
Number of studies and participants included	Total studies = 64 (61 articles) with 341 effect sizes. Studies were experimental and quasi-experimental designs with treatment and comparison. Total participants = 6,349; 3,555 elementary students, 2,007 secondary students, and 787 from a range of elementary and secondary grades. Students were primarily at risk (n = 4,020), followed by identified with a learning disability (n = 2,329).
Appraisal instruments used	Study quality was rated across 3 quality indicators—research design, comparison-group instruction, and implementation fidelity—using procedures applied in previous meta-analyses (Austin et al. 2019; Goldstein et al. 2014; Stevens et al. 2021). We rated each study as exemplary, acceptable, or unacceptable across each quality indicator.

Study	Filderman MJ, Austin CR, Boucher AN, O'Donnell K, Swanson EA. 2022
Description of main results	• A meta-analysis of 64 studies indicated that comprehension intervention for struggling readers in Grades K through 12 had a significant and strong positive effect on reading comprehension outcomes. The immediate post-intervention weighted mean effect size for comprehension interventions was g=.59 ($p < .001$). A meta-analysis of 28 maintenance effects was conducted separately and indicated a nonsignificant but moderately positive effect ($g = .36$, $p = .06$). • Background knowledge and strategy instruction were associated with significantly larger effects. Background knowledge significantly moderated effects, with effects associated with studies that included this instructional approach being significantly higher ($\beta = .54$, $p = .03$, 95% CI [0.06, 1.02]). Background knowledge alone was not a significant moderator of effects ($\beta = .06$, $p = .75$, 95% CI [– 0.34, 0.47]). Strategy instruction without background knowledge instruction also significantly moderated effects—specifically, studies that had strategy instruction alone had significantly higher effects ($\beta = .77$, $p = .01$, 95% CI [0.26, 1.27]). Of all strategies, main idea strategy instruction yielded the strongest effect of $g = .72$, with the effects of inferencing, retell, and prediction ranging from $g = .56$ to $g = .60$. Text structure interventions yielded the smallest effect noted for strategy instruction ($g = .47$). Instructional enhancements and standardised measurement were associated with significantly lower effects ($g = .45$). • Grade level did not moderate effects, suggesting that elementary and secondary students both benefit from comprehension instruction. Grade level, which was divided into elementary and secondary and did not moderate effects ($\beta = .05$, $p = .74$, 95% CI [– 0.25, 0.34]). However, the average effect for elementary students ($g = .47$) was smaller than that observed for secondary students ($g = .67$). Gaps: There were few studies ($k = 11$) that included only background knowledge instruction. Future research is needed with studies that include only background knowledge in the absence of strategy instruction to better understand the unique value of background knowledge instruction on reading comprehension. Research is needed to examine how background knowledge supports reading comprehension in students with varying levels of word-reading ability and language comprehension. Research that strategically pairs background knowledge with selected strategies would further support our understanding of the active ingredients of comprehension intervention. Research that includes only one type of strategy instruction in isolation (for example, main idea) rather than multiple related strategies (for example, main idea and retell) would more clearly demonstrate which strategies are most effective. Future research should make a concerted effort to collect delayed post-test data for all reading instruction and in particular for comprehension instruction as so little is currently known about the long-term effects of such instruction.

Study	Filderman MJ, Toste JR, Didion L, Peng P. 2021
Review objectives	To describe the features of data literacy training (defined as any quantifiable information that helps teachers know more about their students for instructional decision-making) and to determine the effects of data literacy training on teacher outcomes (such as, knowledge, skills, and beliefs)
Descriptions of interventions/ phenomena of interest	The features of data literacy training for kindergarten through 12th grade teachers. The effects of data literacy training on kindergarten through 12th-grade teacher outcomes (such as, data literacy knowledge and skills, beliefs). Training characteristics (such as, coaching, delivery format, inclusion of multiple data literacy skills, and inclusion of data interpretation skills) that may moderate the effects of data literacy training on teacher outcomes.
Descriptions of outcomes included in the review	Knowledge and skills were measured by direct assessments of teacher conceptual understanding (for example, What is curriculum-based measurement?) and/or application of that knowledge (for example, accurate scoring of a curriculum-based measurement). Beliefs were tapped by measures of teacher perceptions of both their ability to use data (such as, self-efficacy) and the value of data use. Researcher developed tools and previously validated scales.
Descriptions of contexts included in the review	A majority of studies included collaborative training components (k = 24). Intervention was delivered in collaborative groups and part in lecture or online learning formats (k = 10) or collaborative groups only (k = 14).
Search details	Databases within the EBSCO electronic library, including PsycINFO, Education Resources Information Center (ERIC), Education Source, and Academic Search Complete and ProQuest. Hand search Date range: 1975 to 2019.
Number of studies and participants included	Total = 33 studies, 163 effect sizes - independent group designs (such as, experimental, quasi-experimental) and repeated measures design (such as, single group pre–post design) study designs. Total participants = 4,844, including 3,361 elementary teachers, 728 secondary teachers, and 265 teachers from a range of elementary and secondary grades, and not reported (k = 12). For studies that reported outcomes of teachers' knowledge and skills (k = 27), participants taught elementary (k= 6, n = 825), secondary (k = 5, n = 728), a range from elementary to secondary levels (k = 5, n = 227), and not reported (k = 11). Participants taught general education (k = 15, n = 1,924), then special education (k = 8, n = 246), a combination (k = 2, n = 103), and not reported (k = 2). The majority of participants were in-service teachers (k = 21, n = 2,001), with fewer preservice teachers (k = 5, n = 434), and one study including preservice and in-service teachers (n = 76). For studies that reported outcomes examining teacher beliefs (k = 11), participants taught elementary (k = 5, n = 2,286), secondary (k = 1, n = 90), a range of elementary and secondary (k = 2, n = 114), and not reported (k = 3). Participants taught special education (k = 1, n = 20), general education (k = 5, n = 249), and a combination of general and special education (k = 3, n = 2,237), or not reported (k = 2). A majority of participants were at the in-service level (k = 7, n = 2,414), with fewer participants at the preservice level (k = 3, n = 135) and one study reporting teachers at the preservice and in-service level (n = 76).
Appraisal instruments used	Not reported

Study	Filderman MJ, Toste JR, Didion L, Peng P. 2021
Description of main results	• Overall, findings suggest that teacher training in data literacy has significant positive effects on both teacher knowledge and skills, and teacher beliefs. The average effect was large for knowledge and skills (g = .67), and moderate for teacher beliefs (g = .48; Cohen, 1988), with both more than large enough to be of interest to policy makers. • There were more negative effects associated with beliefs about the value of data use than with self-efficacy measures. Belief outcomes were separated into self-efficacy and the value of data use. There was a wide range of effects for measures of self-efficacy (g = –0.74 to g = 3.68) and value (g = –0.43 to g = 1.25). • Active learning and collective participation moderated effects, in that the effects of trainings with a collaborative format were significantly higher. Twenty-2 studies included a collaborative format, and this had a significant positive impact on teacher knowledge and skills outcomes (β = 1.31, 95% CI = [0.41,2.21]). Gaps: The real-world applications of data literacy training remain largely unexplored in experimental research. Across all of the studies on training, only one included a maintenance measure did not examine whether teachers transferred this into practice (such as, sustained data use in the classroom). Additional research that compares collaborative training alone with collaborative training along with additional coaching is needed to determine whether and in what conditions coaching may support data literacy training. Some of the planned moderator analyses (for example, duration, participant characteristics) could not be conducted because there were few studies that included the variable of interest or reported the information needed. We require additional knowledge about the mechanisms to maintain these outcomes and transfer learning into practice.

Study	Flynn LJ, Zheng X, Swanson HL. 2012
Review objectives	To synthesize the experimental literature on reading interventions for upper elementary and middle school students (specifically students in grades 5–9) identified with reading disabilities.
Descriptions of interventions/ phenomena of interest	1. Experimental reading interventions for adolescents with reading scores falling below the 25th percentile (standard score of 90) on a norm-referenced reading test. 2. Variation in effect sizes between experimental and control conditions on norm-referenced measures of reading as a function of moderator variables related to instructional conditions (for example, focus of instruction), intervention intensity (for example, length of intervention), and type of outcome measure (for example., word identification, reading comprehension). 2. Variation in effect sizes between experimental and control conditions on norm-referenced measures of reading as a function of sample characteristics, such as grade level, reading level, and intelligence? 3. Effect sizes variations related to treatment outcomes as a function of reported instructional components?

Study	Flynn LJ, Zheng X, Swanson HL. 2012
Descriptions of outcomes included in the review	Pre-test and post-test norm-referenced measures of word recognition, decoding, fluency, comprehension, and math.
Descriptions of contexts included in the review	The majority of studies were conducted in the US, with 4 in Canada, and one in France.
Search details	Four databases: ERIC, PsychINFO, and Web of Science. Hand search. Date range: 1960 to 2009.
Number of studies and participants included	Ten studies (12 independent samples). Total students = 509 (average across studies = 66). Approximately 51% were males.
Appraisal instruments used	Not reported
Description of main results	• The magnitude of treatment outcomes across reading interventions for upper elementary and middle school students identified with reading disabilities on norm-referenced reading measures was small by Cohen's (1988) standards (-0.29 to 0.73). • The magnitude of reading intervention outcome did not vary significantly as a function of the type of reading skills addressed, focus of reading instruction, and/or variations in sample characteristics. A comparison in the magnitude of ESs as a function of the type of dependent measure, such as word identification (M = 0.41), decoding (M = 0.43), comprehension (M = 0.73), reading fluency (M = - 0.29) and math (M = - 0.33) was not significant, $\chi 2$ (4, N = 59) = 9.47, p > .05.

Study	Garwood JD, Brunsting NC, Fox LC. 2014
Review objectives	To review the specific reading comprehension and fluency strategies attempted with secondary students with EBD to present the quantity and characteristics of the research focused on adolescent students' reading. skills.
Descriptions of interventions/ phenomena of interest	1. The number and nature of published studies focused on reading fluency or comprehension outcomes for secondary students with EBD in special education. 2. The research designs and elements of the studies completed since 2004, for example, possible increases in rate of publication and treatment integrity (TI). 3. Effect sizes of the identified interventions.
Descriptions of outcomes included in the review	Words correct per minute (WCPM; 56%); reading comprehension levels (56%); error rate (such as, errors per minute or errors per 30 s; 44%), reading comprehension rates (22%), and behavior (22%). Standardised measures were used across the eight studies reporting measure.
Descriptions of contexts included in the review	US middle or high schools (56% were in middle schools; 22% in high schools and 22% in combined middle and high schools). Non-general education classroom settings, such as self-contained classrooms (67%), resource room (11%) or other unused room (22%) settings.
Search details	Academic Search Complete, Electronic Resources Information Center (ERIC), PsycINFO, and Education Full Text (H. W. Wilson). Hand search and ancestral review. Date range: 2004 to 2012.

Study	Garwood JD, Brunsting NC, Fox LC. 2014
Number of studies and participants included	Total studies = 9, all of which employed a single-subject design. Total participants = 38. Males represented 79%. Age range between 11.4 and 16 years. Reading levels of the participants were reported in 78% of the studies, with a mean range between Grades 3 and 7.3.
Appraisal instruments used	Not reported
Description of main results	• The number of studies focused on reading comprehension or fluency for secondary students with EBD increased since 2004. The current review found 9 studies since 2004 represents a marked improvement by the research community at a rate of approximately one study per year. All 9 studies employed a single-subject research design. • All 9 included studies reported treatment integrity (TI), with eight of the 9 reporting mean scores. The studies included in this review assessed and reported TI at a rate of almost double the "almost half" of studies reporting TI as documented by Griffith et al. (2009). • The range of effects for fluency measures was moderate to very large. The improvement rate difference (IRD) of individual cases across studies on oral reading fluency ranged from 0.51 to 1.00 during the priming phase, from 0.05 to 1.00 during the intervention phase, and from –0.14 to 0.50 for the maintenance phase. Omnibus IRD for oral reading fluency variables during the intervention phase across studies ranged from 0.60 to 0.94 (M = 0.84). • The range for comprehension was questionable to very large. The IRD for individual cases across studies on reading comprehension ranged from –0.25 to 1.00 during the priming phase, from –0.03 to 1.00 during the phase, and from –1.00 to 0.67 for the maintenance phase. Omnibus IRD for reading comprehension variables during the intervention phase across studies ranged from 0.29 to 0.93 (M = 0.57). Story mapping was the only intervention to yield an effect size beyond moderate for reading comprehension. • 60% of the studies reporting very large effects were implemented by teachers. The 2 studies reporting small to questionable effect sizes were both implemented by researchers (Hale et al. 2005; Schmitt et al. 2009).

Study	Gillespie A, Graham S. 2014
Review objectives	To assess the impact of writing interventions on the quality of writing produced by students with learning disabilities (LD)
Descriptions of interventions/ phenomena of interest	Writing interventions for students with LD (for example, strategy instruction, process writing, prewriting, procedural facilitation, goal setting, peer tutoring, creativity training, the addition of self-regulation procedures to strategy or skill instruction, dictation to a scribe or into a tape recorder, adding instruction to process writing, activities to increase writing motivation, instruction on how to write sentences, self-evaluation using a rubric, collaborative writing where students created compositions with peers; and, comprehensive writing programs where multiple writing treatments were combined, for example, strategy instruction, process writing, and prewriting were combined).

Study	Gillespie A, Graham S. 2014
Descriptions of outcomes included in the review	Quality measures of students' writing (included holistic scales, analytic scales, and norm-referenced assessments)
Descriptions of contexts included in the review	Writing interventions were delivered in a variety of settings, including (a) resource room/pull-out (n = 15), special education (n = 9), general education (n = 6), and after school programs (n = 3). Interventions delivered by researchers, teachers, or peers.
Search details	Five databases: Education Abstracts, ERIC, PsycInfo, ProQuest, and Dissertation Abstracts International. Ancestral search. Hand search of National Technical Information Service (NTIS) to identify technical reports. Date range: to Dec 2011.
Number of studies and participants included	Total studies = 43. Students in Grades 1–12 with documented LD (for example, test scores, Individualised Education Plans, special education placement). Most studies (n = 35) involved students in upper elementary and middle grades (such as, Grades 4-8). The remaining studies included students in primary grades (n = 3) and high school (n = 5).
Appraisal instruments used	Studies were further coded and scored for 7 quality indicators: (a) design (true-experiment = 1 point; quasi-experiment = .5 point; within-subject design = zero points) (b) treatment fidelity (1 point if fidelity was .80 or greater) (c) control for instructor effects (1 point if methods were used to control instructor effects; for example, instructors randomly assigned to conditions, instructors taught both conditions) (d) number of instructors (1 point if 2 or more instructors were assigned to each condition) (e) study attrition (1 point if at least 80% of students completed the study) (f) equal attrition across conditions (1 point if there was no more than 10% difference in attrition between groups), and (g) reliability of writing quality measure (1 point if reliability was .80 or higher).

Study	Gillespie A, Graham S. 2014
Description of main results	• The writing quality of students with LD was improved through intervention. Overall, writing interventions had a statistically significant positive impact on the writing quality of students with LD resulting in an average weighted ES of 0.74, ($p < .001$) across the 43 studies. Thirty-eight of the studies (86%) of studies yielded a positive ES. These effects were found across all grades, with most involving students in Grades 4 to 8. None of the moderators (such as, study quality score, year of publication, type of publication, and who delivered the intervention) had statistically significant relationships with the average weighted ES. • Four of the writing treatments (such as, strategy instruction, dictation, goal setting, and process writing) had statistically significant effects on the writing quality of students with LD. The average weighted ESs for the 6 writing treatments containing 4 or more ESs were all positive. • Treatments designed to enhance a specific writing process were only effective when time was devoted to teaching the writing skill or process. To learn to use these writing processes independently, students with LD need systematic instruction, teacher support, and scaffolding. Gaps: It is especially important that future writing intervention studies are true experiments that control for instructor effects, report reliability of outcome measures, and provide treatment fidelity data, as these were weaknesses in the studies reviewed here. Future research should evaluate comprehensive writing programs and multicomponent interventions that involve teaching a wide range of writing skills to students with LD.

Study	Goodwin AP, Ahn S. 2013
Review objectives	To aggregate findings from 30 English studies and examines statistical differences related to study features to better understand whether and under what circumstances morphological instruction improves literacy achievement for school-aged students learning in English.
Descriptions of interventions/ phenomena of interest	Interventions with a morphological component with an emphasis on morphemes as units of meaning
Descriptions of outcomes included in the review	Standardised or researcher designed measures of literacy achievement outcomes (reading comprehension, decoding, fluency, morphological knowledge, phonological awareness, spelling, and vocabulary)
Descriptions of contexts included in the review	Preschool–12th grade; instruction in English. Unit of intervention was categorised into 3 groups: individualised, small groups with less than 12, and large group
Search details	Manual and computerised searches of databases including ERIC, Education Full Text, PsycINFO, and Dissertation Abstracts Online Database. Random, quasi-experimental, or nonexperimental designs
Number of studies and participants included	30 independent studies with 92 standardised mean differences. Random, quasi-experimental, or nonexperimental design. Average sample size ranged from 12 to 1,569 ($M = 116.14$, $SD = 254.09$) and participants from preschool to ninth grade were included.

Study	Goodwin AP, Ahn S. 2013
Appraisal instruments used	Not reported
Description of main results	• Overall, children in the morphological intervention groups yielded statistically higher mean scores on literacy outcomes when compared to comparison groups. Under the random-effects model, the weighted-mean effect size was 0.32. However, no statistically significant intervention effect was found for high school (d = 0.45). The weighted means under the mixed-effects model were statistically significant for upper elementary (d = 0.29), and middle school (d = 0.34). Also, a statistically significant intervention effect was found for children from a broad range of grade levels (d = 0.19). • Moderator analyses suggest morphological instruction supported proximal outcomes, but effects did not transfer to supralexical processing (such as, reading comprehension and fluency). Decoding showed the largest statistically significant mean effects (d = 0.59), followed by phonological awareness (d = 0.48) and morphological knowledge (d = 0.44). Significant effects were also found on vocabulary (d = 0.34) and spelling (d = 0.30). However, no statistically significant intervention effects were found for reading comprehension (d = 0.09) and fluency (d = –0.05). • Most interventions with students with literacy challenges occurred in small groups or individualised settings compared to work with general education students occurring in larger, classroom-sized groups. No statistically significant mean difference was found by unit of intervention, Q between (3) = 4.14, p = .13, although largest mean values were for individualised intervention (d = 0.42), followed by small group (d = 0.35), and then large group interventions (d = 0.29), which were all statistically significant. Findings suggest instruction is most effective when unit size is designed around the needs of students who are receiving the instruction. • Interventions showed larger effects on researcher-designed measures than on standardised measures. The significant Q-statistics of 36.28 indicates the mean effect of morphological intervention based on standardised tests (d = 0.23) was statistically lower than one from researcher-made tests (d = 0.36), which both were statistically significant. • ELLs, poor readers and spellers, and children with learning disabilities responded similarly to the morphological interventions. Mean values were statistically significant for all types of learners but were not statistically different, Q between (3) = 4.90, p = .18. ELLs showed the largest intervention effect (d = 0.54), followed by children with learning disabilities (d= 0.37), then poor readers/spellers (d = 0.35), and typical achievers (d = 0.29). Gaps: Because confidence in making causal inferences is highest when considering experimental designs, future research should continue to include experimental studies like the most recent addition of a large randomised control trial (Lesaux et al. 2012).

Study	Graham S, Liu X, Bartlett B, Ng C, Harris KR, Aitken A, et al. 2018
Review objectives	To examine if students' writing performance is improved by reading interventions
Descriptions of interventions/ phenomena of interest	Writing interventions where students were taught how to read (phonological awareness instruction [39%], phonics instruction [28%], vocabulary instruction [2%], reading fluency instruction [2%], comprehension instruction [22%], and multicomponent reading instruction[7%] (this treatment involved reading instruction at both the word and comprehension level) and studies where students' interaction with words or text was increased through reading or observing others read (self-teaching where students read words or text and effects on spelling was assessed (Share, 1999); more reading where the amount of text students read was increased (this did not include studies that only assessed spelling, as it would overlap with self-teaching); (c) observing readers as they engage in reading or use text to carry out an activity; (d) reading and analysing another person's text; and, reading model text to emulate it when writing.
Descriptions of outcomes included in the review	Measures of student performances on spelling, writing quality and words
Descriptions of contexts included in the review	Experiments were published in a variety of countries including: the US, Canada, Israel, Netherlands, the United Kingdom, Italy, Spain, Netherlands, Hong Kong, Slovenia, South Korea, Australia, France, Hong Kong, Norway, Portugal, and Taiwan.
Search details	Six search strategies: Electronic databases - ERIC, PsychINFO, ProQuest Dissertations and Theses Global, Linguistics and Language Behavior Abstracts, and EBSCOhost. Hand searches of 16 journals; ancestral search of previous reviews, Google search, ancestral search of all documents. Date range: Ended January 1, 2016, but extended backward as far as possible.
Number of studies and participants included	Total experiments Reading taught interventions :54, true- or quasi-experimental studies Total students = 5 018 in preschool to 12th grade (60% of studies with Grade 1 to 6 students, and 27% of studies with preschool to kindergarten children). Almost one-half of the studies involved students who experienced difficulties mastering literacy. Increased reading of text interventions: 36 studies, true- or quasi-experimental studies or involving a participant as own control designs Total students = 3,060 students in preschool to Grade 12 (58% of the studies were conducted with students in Grades 1 to 6, and 11% of studies were conducted with preschool to kindergarten children). Eighty-three percent of studies involved typically developing students.
Appraisal instruments used	Nine quality indicators: scored 1.0 if met, otherwise 0. The indicators were: 1. high-quality design (true experiment); 2. not an N of 1 design (more than 2 groups or classes in each condition); 3. teacher effects controlled (teachers randomly assigned to condition or taught in each); 4. attrition (not greater than 10%); 5. differential attrition (differences in attrition between each condition was 5% or less);

Study	Graham S, Liu X, Bartlett B, Ng C, Harris KR, Aitken A, et al. 2018
Appraisal instruments used	6. pre-test equivalence (treatment and control conditions demonstrated equivalence on pre-test writing measure as mean scores for each condition did not differ by more than the smallest standard deviation for the 2 conditions at pre-test; true-experiments that did not include a pre-test were credited with meeting the criterion); 7. no floor or ceiling effects at pre-test or post-test (mean score for a writing measure was not more than one standard deviation from the lowest and highest scale for the measure); 8. measures were reliable (a reliability coefficient of .70 or greater was reported for the writing measure; norm-referenced measures were assumed to be reliable); and 9. treatment fidelity (evidence of treatment fidelity provided).
Description of main results	• Teaching reading improved students' writing performance. Teaching reading resulted in statistically significant effects for an overall measure of writing (effect size [ES] = 0.57) and specific measures of writing quality (ES = 0.63), words written (ES = 0.37), or spelling (ES = 0.56). The impact of teaching reading on writing was maintained over time (ES = 0.37). More specifically, phonological awareness (ES = 0.69), phonics (ES = 0.39), and reading comprehension instruction (ES = 0.66) significantly strengthened students' writing performance. • The average weighted ESs at all 3 grade levels and each type of experiment were statistically greater than zero. When individual study characteristics were tested, type of experiment (true experiment ES = 0.46 and quasi-experiment ES = 0.99; Q [between] = 6.83, degrees of freedom [df] = 1, $p < .01$) and grade level (preschool ES = 0.61, elementary ES = 0.46, and secondary = 0.57; Q [between] = 9.30, df = 2, $p < .02$) were statistically significant moderators. • Increasing students' interaction with text improved their writing performance. Having students read text or observe others interact with text also enhanced writing performance, producing a statistically significant impact on an overall measure of writing (ES = 0.35) and specific measures of writing quality (ES = 0.44) or spelling (ES = 0.28). The long-term impact of these effects, however, is not certain as writing gains were not maintained over time. The average weighted ES for these studies for writing overall was 0.15, and this effect was not statistically significant. Only 2 of the 5 studies produced positive results. Gaps: Research is needed to provide additional verifications that the 4 treatments identified (phonological awareness, phonics, and comprehension instruction as well as reading words or text) enhance writing along with further investigation of the impact of other reading treatments on writing, such as reading fluency instruction, vocabulary instruction, more reading, teaching students about the functions and structure of text, reading and analysing text, observing readers interact with text, multicomponent reading interventions or reading model text. Future research needs to expand its focus to include other product measures (for example, textual organization and cohesion, writing vocabulary, structural elements) as well as measures of writing knowledge, process, and motivation. An especially critical goal for future studies examining the impact of reading interventions on writing is to determine if obtained effects are maintained over time.

Study	Hall MS, Burns, MK. 2018
Review objectives	To apply meta-analytic techniques to examine the effects of targeted small group reading interventions and to identify components of small-group instruction that were most effective
Descriptions of interventions/ phenomena of interest	The effectiveness of small-group (3 to 15 students) reading interventions targeting specific skills, such as phonemic awareness or phonics (n = 3 studies), reading fluency (repeated reading and reading widely; n = 1), vocabulary (explicit instruction; n = 5), reading comprehension (n = 4), compared to a more comprehensive approach that addresses multiple skills (n = 14). The impact of intervention characteristics such as interventionist, dose, and group size on the effect of small-group reading interventions. The impact of participant-level variables (for example, student grade) and study characteristics such as research design, outcome measure used, and control group.
Descriptions of outcomes included in the review	Standardised norm-referenced measures, researcher-developed measures or curriculum-based measurement of reading or literacy performance
Descriptions of contexts included in the review	Small-group (3 to 15 students) reading interventions for grades K through 12. Intervention was delivered by an adult rather than peer to peer.
Search details	Three databases: PsycINFO, ERIC, and Education Full Text. Ancestral search
Number of studies and participants included	Total studies = 26. Quantitative group comparison research design, either experimental or quasi-experimental. Students in Grades K - 12; all native English speakers. Total number of students not reported.
Appraisal instruments used	What Works Clearinghouse (2008) Procedures and Standards

Study	Hall MS, Burns, MK. 2018
Description of main results	• An overall weighted effect size for small-group reading interventions of g = 0.54 was found, suggesting that small-group reading interventions in general were moderately effective. The confidence interval for the overall median effect did not include 0, and 46 studies with a 0 effect would have to be found to change the score to fall below the criterion for a small effect of g = 0.20. Thus, across the examined studies, small-group reading interventions reliably resulted in a positive effect for students. • Interventions that were targeted to a single reading skill area were found to be more effective than more general interventions that combined multiple reading skill areas; however, the confidence intervals did overlap. Providing targeted reading interventions to students in small-groups through a multi-tiered system of support or other models appears to be an effective method for increasing students' reading skills. • Small-group reading interventions implemented by teachers had a small to moderate median effect size and were less effective than intervention was administered by researchers or by trained interventionists from outside the school. Effect sizes by interventionist were: teachers (g = 0.59); researchers (g = 0.51) or by trained interventionists from outside the school (g = 0.52). There was considerable overlap between the 3 confidence intervals and the magnitudes of the weighted effect sizes were quite similar. • Targeted interventions and group size were the 2 most important intervention variables among the ones examined. Using targeted interventions and group size accounted for 19% of the variance, which is a large effect (Cohen, 1988). The 2 variables were roughly equal in magnitude of effect (r2 =0.09 and r2= 0.11) respectively. • The effect size of small-group reading interventions for secondary students was small. Interventions delivered to elementary school students resulted in an effect size that was over 3 times as large as the one for secondary students, but the confidence intervals for the 2 estimates of effect did overlap. For students in secondary grades (8th through 12th grades) the effect size was small (g = 0.20, 95% CI= -0.01 –0.41), whereas the effect size for elementary school students (such as, students in grades K through 5th grades) was moderate (g = 0.64, 95% CI= 0.38–0.90).

Study	Hall-Mills SS, Marante LM. 2022
Review objectives	To summarise the literature of the past 2 decades regarding the effects of explicit text structure instruction for adolescents with or at risk of LD and illuminate trends in the current evidence supporting the use of text structure literacy interventions with adolescents in high incidence populations who have literacy deficits.
Descriptions of interventions/ phenomena of interest	The outcomes of explicit text structure instruction for adolescents with or at risk for LD

Study	Hall-Mills SS, Marante LM. 2022
Descriptions of outcomes included in the review	Outcome measures ranged from norm-referenced standardised tests (distal measures) to researcher made and criterion referenced measures (proximal measures). Proximal outcome measures included: 1. Number of idea units, central and incidental information recalled from science and social studies texts; 2. Pre/post surveys on text structure strategy knowledge; 3. Comprehension on short and long comparison structure texts measured by scoring for top-level structure, number of issues, and competence; identification of signalling words in a short comparison text, scoring for top-level structure and competence based on a problem-solution structure text, identifying the main idea of a comparison text; 4. Researcher-designed measures for main idea, compare and contrast, cause and effect, and multiple-choice reading comprehension of history; 5. Total idea units identified for cause/effect and compare/contrast texts; 6. Maze procedure to measure reading comprehension and ability to discern meaning of text structures; 7. Curriculum-based measurement maze assessment; 8. Comprehension of content area (social studies, science) expository reading passage; and 9. Percentage correct science reading comprehension questions.
Descriptions of contexts included in the review	Intervention settings: whole class (n = 2), small group (n = 3), individual (n = 2). [Meta-analyses excluded]. Delivered by classroom teachers, researchers, and a self-paced, web-based software application.
Search details	Four databases (Academic Search Complete; Educational Resources Information Center [ERIC; ProQuest]; PsycINFO; ComDisDome). Date range: January 1997 to January 2019.
Number of studies and participants included	Total studies = 9 (included 2 previous meta-analyses). Total number of students across 7 studies (excluding meta-analyses) = 417. Adolescents (aged 10 to 19 years; Grades 7 through 12) with or at risk of LDs (for example, diagnosed with SLD, LD, or reading disability [RD] or identified by their school as being at risk for reading failure).
Appraisal instruments used	American Speech-Language-Hearing Association as reflected in the Quality of Evidence Summary Table (QuEST; Wolter et al. 2011).

Study	Hall-Mills SS, Marante LM. 2022
Description of main results	• The results revealed that explicit text structure instruction leads to improvements in expository reading comprehension for adolescent students with or at risk for LD. In each study reviewed, there was a main effect of text structure instruction on expository reading comprehension outcomes for adolescents with LD. Most effect sizes were large, and effects ranged from small to large (d = 0.31–2.17). Six studies reported statistically significant gains in text comprehension after intervention. • There is some evidence that explicit text structure instruction transfers across content areas and gains are maintained over longer periods of time. Transfer effects (for example, transfer to a different text structure, different content in same text structure, etc.) were reported in 3 studies in the present review (Bakken et al. 1997; Ihle, 2011; O'Connor et al. 2017). Gaps: There was variability in the inclusion criteria for students with LD. Future researchers should consider clear definitions of the sample to avoid confusion among subgroups. More research is needed to confirm the effectiveness of text structure approaches for adolescents with or at risk for LD (only a small number of studies actually included participants with LD). At the present time, there is not enough evidence to conclude that transfer and maintenance occur consistently with text structure intervention for adolescents with LD.

Study	Huddle S, Hosp J, Watt S. 2017
Review objectives	To examine reading interventions for adolescent English Language Learners (ELLs) with reading difficulties and disabilities
Descriptions of interventions/ phenomena of interest	The characteristics of interventions for adolescent ELLs with reading difficulties (for example, multicomponent and extensive). The characteristics of ELLs, who benefit from the interventions (e.g., English language proficiency and home language). The effectiveness of reading interventions for improving reading outcomes for adolescent ELLs with reading difficulties. The effect of between treatment and comparison conditions on ESs by outcome measure?
Descriptions of outcomes included in the review	Measures of: Comprehension – any measure that required students to read the text and answer questions or complete sentences within the text. Fluency – any measures that included rate and accuracy during text reading. Word – measures of word identification. Decoding – a measure of decoding or word attack skills. English language skills – any direct measures of vocabulary knowledge or listening comprehension skills. All of the studies used at least one standardised outcome measure.
Descriptions of contexts included in the review	The language of intervention instruction was English. Reading interventions were conducted in a school setting (grades 4–12). Six of the studies were conducted in middle schools (grades 6–8), one in an elementary school (grade 4), and one in a high school (grades 9–12). All of the studies were conducted in large urban schools with high percentages of students who were economically disadvantaged (for example, above 80% free and reduced lunch). Delivered by a teacher (k = 4) or researcher (k = 4). Group sizes ranged from 2 - 10 students with 5 studies providing intervention in groups of 4 or smaller and 3 studies providing intervention to groups of 5 - 10.

Study	Huddle S, Hosp J, Watt S. 2017
Search details	Databases EBSCOhost and PsycINFO, including Academic Search Elite, ERIC, and Teacher Reference. Date range: 1980 to Dec 2012.
Number of studies and participants included	Total of 8 studies (11 comparisons) – experimental (k = 6) or quasi-experimental (k = 2) designs. Total students = 844.
Appraisal instruments used	What Works Clearinghouse Procedures and Handbook Version 3.0 (IES, 2008), and the Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education (Gersten et al. 2005).
Description of main results	• The limited research examining intervention for adolescent English Language Learners (ELL)s and variability in effect sizes meant that meaningful conclusions are problematic. Overall effect sizes (n = 46) ranged from -0.20 to 0.71. Reading interventions for ELLs are generally multicomponent and are not generally extensive (more than 75 sessions). • Reading intervention ESs for adolescent ELL demonstrated a variability both across outcome measures and within outcome measures. Small to medium effects were detected for fluency outcomes (ES = 0.40–0.64). No effects to small effects were evident for word reading and language outcome measures (ES = -0.06 to 0.47). Comprehension ESs were variable ranging from -0.20 to 0.71. • Contextual variables are not universally described in a way that will facilitate knowing what individual ELLs would benefit from reading interventions. The ESs demonstrated a considerable amount of variability. Different methods for classifying ELLs were used in the studies, students likely had different levels of home language proficiency, and students likely received different types of ELL instruction. However, 7 of the eight studies included information about how students were classified as ELL and 6 of the eight studies provided information about students' English proficiency levels. Four studies linked limited English proficiency (LEP) status or low receptive vocabulary to lower comprehension outcomes. Gaps: It is clear from this review that more research is needed to not only determine what interventions are effective for adolescent ELLs but also what interventions are effective for individual ELLs based on their unique characteristics. In addition, research is needed that examines different components of interventions to determine if certain components of intervention (for example, vocabulary and fluency) are more effective for adolescent ELLs. More research is also needed on interventions that are extensive.

Study	Jitendra AK, Lein AE, Im S, Alghamdi AA, Hefte SB, Mouanoutoua J. 2018
Review objectives	to provide a quantitative synthesis of empirical evaluations of mathematical intervention programs implemented in secondary schools for students with learning disabilities and mathematics difficulties
Descriptions of interventions/ phenomena of interest	Effect of mathematical interventions for secondary students with LD and MD. Moderators of student outcomes (instructional characteristics (setting, instructional time, implementer, fidelity of implementation).

Study	Jitendra AK, Lein AE, Im S, Alghamdi AA, Hefte SB, Mouanoutoua J. 2018
Descriptions of outcomes included in the review	Quantitative measures of mathematics performance in: foundational skills such as operations and algebraic thinking (for example, arithmetic word problem solving) and number system (for example, fractions, decimals), higher-level mathematics topics (such as, ratio and proportional relationships, expressions, and equations). Quantitative measures of maths performance included researcher designed and standardised.
Descriptions of contexts included in the review	Middle and high schools
Search details	Three databases: online databases ERIC, PsycINFO, and ProQuest Dissertation and Theses Plus ancestral search Plus hand search 1989 – 2017.
Number of studies and participants included	19 treatment-control studies containing 20 independent samples were included. 68% were RCT, the remainder were quasi-experimental total sample size of 1,959. Nine studies included middle school students; 6 involved high-school students; and 4 included both. About 8% of the studies comprised middle school students, and 3% included both elementary and middle school students. Fourteen studies (74%) involved students with LD.
Appraisal instruments used	Critical appraisal of study quality was not undertaken. Efforts to minimise bias were included. For example, Hedges's g correction was used to reduce this small sample size bias, also conducted an analysis of publication bias, evaluated heterogeneity of variance using the Q statistic and conducted testing for moderators.

Study	Jitendra AK, Lein AE, Im S, Alghamdi AA, Hefte SB, Mouanoutoua J. 2018
Description of main results	• Overall, providing mathematics interventions to secondary school students is moderately effective (g = 0.37). This impact varied according to instructional and methodological features. • Mathematics interventions for older students were equally as effective in any setting (general education vs special education setting/classroom. Effect sizes were similar for studies conducted in general education (g = 0.46) or special education settings (g = 0.41), and the difference was not statistically significant (p =.79). • Mathematics interventions for older students were equally as effective when implemented in large (more than ten) and small (less than ten) groups. Large- and small-group implementation (g = 0.37 and 0.38, respectively) • Mathematics interventions for older students were equally as effective when implemented to enhance foundational content and higher-level mathematical content (g = 0.36 and 0.38, respectively). • Mathematics interventions for older students had a stronger impact for students in middle than high school (middle school, g = 0.44, vs. high school, g = 0.23) and the advantage for middle school was significant (such as, the confidence interval of the effect sizes did not contain 0). • Results from the moderator analysis of the present study support the need for engaging students in learning for an extended period, especially when learning higher-level mathematical content. • A key finding of this meta-analysis was that the effect of these interventions appeared to be greater when the intervention lasted >10 hr than when it was ≤10 hr. The mean effect size for instruction provided for >10 hr (g = 0.58) was greater than that for instructional time ≤10 hr (g = 0.11). This finding is consistent with research regarding the length of instructional time provided to students being an essential variable to learning. Gaps: Insufficient studies to inform the field about effective intervention approaches. The strongest intervention effects were found when combining visual strategies (for example, fraction circles or squares, manipulative objects, diagrams, mental representations) with other strategies (for example, priming the underlying problem structure, cognitive strategy, contextualised instruction such as problems used real-world scenarios). visuals only, g = 0.28, vs. visuals combined with other strategies, g = 0.52), but the contrast was not statistically significant. Although not statistically significant, it is practically significant and can be interpreted as showing that approximately 70% of students in the treatment group performed above the mean of students in the control group. Additional testing of these interventions needs to be addressed in future research. Few studies with students with LD and MD so power is lower. More studies required to increase confidence in findings

Study	Joseph LM, Schisler R. 2009
Review objectives	To provide a synthesis and critical analysis from a review of research in the past 20 years that sought to examine the effectiveness of word reading instruction methods on reading achievement of students in the middle and high school grades.
Descriptions of interventions/ phenomena of interest	Reading instructional methods for teaching word reading skills to middle and high school pupils
Descriptions of outcomes included in the review	Quantitative measures of 4 basic reading skill outcome variables: word identification, pseudoword reading, fluency, and comprehension
Descriptions of contexts included in the review	Middle schools (Grades 6 through 8), and high schools (Grades 9 through 12). Remedial reading classes/programs, resource rooms, special education classes or not stated.
Search details	4 databases: ERIC, Educational Abstracts, Educational Research Complete, and PsycINFO. Hand search. Ancestral search. Date range: 1986 to 2006.
Number of studies and participants included	Total articles = 23. Fifteen studies used large-group quasi-experimental or experimental research designs, and 9 studies used single-subject research designs to determine the effectiveness of word reading instruction. A total of 1,181 participants. Eleven studies included only middle school students, (n = 346 participants). Nine studies consisted of high-school students (n = 203). Four studies included a combination of students at middle and high school levels (n = 632). Most participants were characterised as having a disability in the area of reading or as demonstrating significant delays in reading achievement. Three studies included participants with mild intellectual disability, and 3 studies included participants with serious emotional and/or behavioural disabilities.
Appraisal instruments used	Not reported

Study	Joseph LM, Schisler R. 2009
Description of main results	• In general, teaching word reading skills to adolescents has a medium-to-large effect on reading achievement performance for middle and high-school students with reading disabilities or significant delays in reading achievement. Across all reading achievement outcomes in these investigations, the mean effect size was .83, the median effect size was .68, and the mode was .55 (SD = .82). • Overall, the reading instructional methods and programs that were implemented appeared to have the greatest impact on reading fluency and rather moderate influence on comprehension. The mean effect size of teaching basic reading skills reflected overall medium effect sizes (M ES = 0.55) on word recognition, pseudoword reading and reading comprehension, and an overall large effect size (M ES = 1.31) for reading fluency. It appears that programs that taught multiple basic component reading skills were effective for helping adolescents achieve a multitude of basic reading skills. The investigations in this review revealed that word reading skills were mainly taught through the use of established programs; however, few investigations reported the use of specific strategies. Before any conclusions can be made about the use of one type of explicit, systematic, basic reading skill instruction program, method, or strategy over others, more research is needed to validate the current findings and to extend upon them and measure longitudinal results. Findings from this review suggest that explicit systematic instruction on basic reading skills is critical for helping students progress from a learning-to-read to a reading-to-learn stage. What seems to be most clear from this review of studies is that programs and specific strategies including components that matched the needs of students produced desirable results. For instance, if students needed help with increasing oral reading fluency, instructional methods that included repeated readings was clearly beneficial. Gaps: All of the programs and methods described in this review need further investigations to substantiate their use with adolescent populations. Future studies should consider including larger numbers of ethnically diverse populations as well as adolescents with severe or less common types of disabilities (such as, autism, serious emotional disturbance, behavioural disorders, and moderate-to- severe intellectual disability). Likewise, it would be important to determine which word reading instruction methods are most effective for teaching ELLs in middle or high school grades. Researchers may wish to examine instructional efficiency as well as effectiveness, given the instructional time constraints within a school day. Of all studies that were reviewed, none reported the effects of basic reading instruction on understanding expository text. Thus, research is required to explore the effects on academic achievement in other content areas such as science, social studies, or history given the increasing expectation to read expository text as students enter middle school.

Study	Jung P, McMaster KL, Kunkel AK, Shin J, Stecker PM. 2018
Review objectives	To examine the effects of teachers', use of data-based individualisation (DBI) to improve academic performance for K-12 students with intensive learning needs, including those with disabilities
Descriptions of interventions/ phenomena of interest	The mean effect of teachers' uses of DBI on student achievement across academic areas (reading, mathematics, and spelling/writing). Factors that influence the strength of the effects of DBI on student outcomes.
Descriptions of outcomes included in the review	Measures of academic performance using curriculum-based measurement (CBM) or CBM-like tasks (mimicked the tasks used for progress monitoring but were designated specifically for pre- and post-testing and were not the same assessments used for progress monitoring), commercial norm-referenced tests, or researcher-developed tests (used standardised administration procedures and usually addressed more generalised performance, or included broader skills that spanned grade levels). CBM administrators were DBI implementers, computer software, or both).
Descriptions of contexts included in the review	Elementary, middle, and high schools (not explicitly reported); grades 1 – 9
Search details	Hand search of relevant literature from previous reviews on teachers' use of CBM for instructional decision-making. Database search using Academic Search Premier, ERIC, PsychINFO, Education Full Text, and ProQuest Dissertations and Theses. Most studies were conducted in the 1980s and 1990s. Ancestral search by examining the reference lists of all included studies.
Number of studies and participants included	Total studies = 14 experimental or quasi-experimental designs. Total participants = 872 (excluding one study that did not report N). Students participants with disabilities receiving special education services.
Appraisal instruments used	Not reported

Study	Jung P, McMaster KL, Kunkel AK, Shin J, Stecker PM. 2018
Description of main results	• Findings of this study provide promising evidence of DBI for improving student outcomes across reading, mathematics, and spelling/writing. The mean effect of DBI Only on student performance across academic areas was g=0.37 (k=14, CI95=0.17, 0.58), indicating statistically significant effects of DBI Only for enhancing student performance. The overall mean effect of DBI Plus (incorporated additional information to facilitate teachers' instructional decision-making) on student performance was similar: g = 0.38 (k = 6; CI95= 0.15, 0.61), indicating statistically significant effects on student performance. • Analyses revealed that average effects of DBI Only did not differ reliably by academic area. Moderator analyses were conducted on DBI Only but not DBI Plus studies, given significant heterogeneity in DBI Only effects but not among DBI Plus effects. The effects of DBI Only on student performance were not statistically different across reading, mathematics, and spelling/writing, though there was little remaining heterogeneity between group effects (Q[2] = 1.56, p = .46). Most effects (k = 20) came from studies of DBI in reading. Effects from studies of spelling/writing were the largest (g = 0.47), followed by mathematics (g = 0.37) and reading (g = 0.28). Confidence intervals in this category did not include zero, indicating statistical significance of all 3 effects. • Results of moderator analyses revealed differential effects of DBI Only depending on CBM characteristics. DBI Only effects were larger in studies that used teacher-generated CBM (g = 0.80), compared to those that used researcher developed CBM (g = 0.39). DBI Only effects also varied depending on the frequency of CBM administration, with twice-weekly administration yielding the strongest effects (g = 0.47); NB. most effect sizes were drawn from the twice-weekly variable. More frequent (daily) CBM administration yielded negative effects. • Effects of DBI Only varied depending on the type and frequency of support provided for teachers. Effects were largest in the one study that incorporated small-group, collaborative problem solving with individual consultation (g = 0.86; Jung, McMaster, and del Mas, 2017); effects also were larger for more frequent (weekly or twice-weekly) support than for less frequent support (every 1–3 weeks g = 0.23; every 2–3 weeks g = -0.02). Gaps: Further research is needed to determine the influence of frequency of CBM administration (for example, weekly) and to determine the optimal support needed to promote strong student outcomes.

Study	Kaldenberg ER, Ganzeveld P, Hosp JL, Rodgers DB. 2016
Review objectives	To expand the current writing intervention literature base by conducting a meta-analysis of single-subject writing intervention studies for students with LD.
Descriptions of interventions/ phenomena of interest	Instruments and metrics used to measure the efficacy of writing interventions for students with LD. The effect sizes of the different writing interventions for this population. Key similarities and differences between SRSD and non-SRSD interventions used to teach writing.

Study	Kaldenberg ER, Ganzeveld P, Hosp JL, Rodgers DB. 2016
Descriptions of outcomes included in the review	39 different dependent measures were reported across the 23 studies (see Table 2), only 9 of which were used in at least 3 studies. The 9 dependent measures were: total number of words written (TWW; n = 16); score from a holistic quality rubric (n = 15); number of functional essay elements (n = 8); percentage of words spelled correctly (n = 3); coherence score (n = 3); number of story grammar elements (n = 3); score based from a nonholistic quality rubric (n = 3); number of correct word sequences (CWS; n = 3), and number of T-units (n = 3). T-units have been used by linguists to measure sentence development in developing writers. Each dominant clause and corresponding dependent clauses counts for one T-Unit (Hunt, 1965).
Descriptions of contexts included in the review	Public or private school in the US (grades 4 to 12). Eighty-five percent of the interventions took place outside of the general education classroom and were implemented by a member of the research team (78%).
Search details	Hand search of 2 previous reviews and meta-analyses for single subjects writing interventions. 4 electronic databases: Education Full Text, Educational Resources Information Center (ERIC), and PsycInfo for post 2005 articles. Hand search of relevant education and psychology journals from 2005 to 2012 related to special education, LD, and literacy. No grey literature. Date range to 2012.
Number of studies and participants included	23 studies in 21 articles published between 1988 and 2012. Total students = 102.
Appraisal instruments used	Not reported
Description of main results	• In general, researchers used a single piece of writing to score 1–12 different metrics to determine the effectiveness of writing interventions for students with learning disabilities. There was significant variation in how the outcome measures were administered, including the amount of time the student was given to write. Even though some dependent variables were used in multiple studies, variation still existed. For instance, 16 studies recorded the total number of words written (TWW) by each student. However, the amount of time given to students to write varied from 3 minutes to unlimited. • A variety of instructional strategies improved the writing of students with LD. Studies that used an SRSD (k = 13) intervention produced the highest mean phi (.83; SD = .15), whereas studies that fell into our non-SRSD category (k = 10) also produced a high effect size (.71; SD = .23). There was no significant difference between the 2 intervention categories; a Mann–Whitney test resulted in p = .257. The non-SRSD category consisted of 2 direct instruction (DI) studies (M = .87), 2 assistive technology (AT) studies (M= .44), and 6 studies that fell into a category we called strategy instruction (SI; M= .75). Gaps: Writing measures were not used consistently, making comparisons across studies difficult. In addition, reliability and validity data were not reported for the individual studies. These data are important and future researchers should focus on investigating the psychometric properties of these writing measures.

Study	Kaldenberg ER, Watt SJ, Therrien WJ. 2015
Review objectives	To expand on the findings from previous literature on effective strategies for teaching reading comprehension of science text for students with learning disabilities (LD).
Descriptions of interventions/ phenomena of interest	Reading comprehension interventions for science text geared toward students with LD. Interventions that emphasize teaching vocabulary for the purpose of increasing comprehension. Moderating variables (if any) affected the effect sizes of interventions used to teach reading comprehension of science text among students with LD.
Descriptions of outcomes included in the review	Assessment of reading comprehension using science text (for example, filling in missing words within science passages, reading text and answering question, or recalling the main ideas or details of a text).
Descriptions of contexts included in the review	Students in Grades 3 to 12; no other details were reported.
Search details	Hand search of 2 previous reviews. Two databases: ERIC and PyschInfo. Hand search of the special education journals identified by Berkeley et al. (2010). Date range: 1980 to 2012.
Number of studies and participants included	Total studies = 20 - quantitative group study (included quasi-experimental, excluded single subject); Average sample size = 76 students (Mdn = 54), ranging in size from 10 to 211. Students in Grades 5 to 11.
Appraisal instruments used	Council for Exceptional Children Standards for Evidence-Based Practices in Special Education (Cook et al. 2014).

Study	Kaldenberg ER, Watt SJ, Therrien WJ. 2015
Description of main results	• All of the interventions aimed at increasing the reading comprehension of science texts for students with LD were effective (ESs > 0.50). Under a random effects model, the studies produced a combined ES of 0.98 (standard error [SE] = 0.15). Difference between those that reported a combined ES for students with LD and general education students (n = 18) compared with those that reported an ES for students with LD (n = 2): Results indicated that there was not a significant difference between the studies (Q = .120, p = .729) in this regard. 7 studies reported delayed measure data with a combined ES of 0.69 (SE = 0.21). • Teaching students the meaning of vocabulary words found in science text had a significantly greater impact on the comprehension of that text for students with LD than other non-vocabulary interventions. Comparing ESs between the 2 types of studies showed a greater combined ES for the vocabulary studies (ES = 1.25; k = 11) than the non-vocabulary studies (ES = 0.64; k = 9). Findings suggested that effective vocabulary instruction should focus on semantic mapping (Bos and Anders, 1990, 1992), use direct instruction (for example, precision teaching, Lovitt et al. 1985, 1986), or supply the students with mnemonics (Mastropieri et al. 1987; Scruggs et al. 1987). Explicitly teaching (using a direct instructional approach or a cognitive strategy approach) students the definitions of vocabulary words can increase reading comprehension. In addition, supplementing a meta-cognitive component (for example, students are actively reflecting on their thinking) in combination with providing direct instruction of vocabulary to students was found to be more effective on overall reading comprehension (Bos and Anders, 1990). • Findings from the non-vocabulary studies suggest that students with LD also benefit from instruction that helps students to better understand expository text structure (Bakken et al. 1997). Students with little background relating to the text content (common among students in content area reading) do rely heavily on these features for understanding (Goldman and Rakestraw, 2000). It is important to teach students to identify main ideas and supporting details while reading. To accommodate students through this process, graphic organisers have been found to be effective (Billingsley and Wildman, 1988), along with explicitly teaching students the main ideas and important details prior to reading (Griffin et al. 1991). It is also important to teach and require students to be active learners constantly reflecting on what is being read (for example, having the student ask or write down notes about why something would be true; Mastropieri et al. 1996). Gaps: Further investigations examining the use of non-vocabulary and vocabulary interventions in inclusive settings to increase the reading comprehension of science text for students with LD are needed.

Study	Kang EY, McKenna JW, Arden S, Ciullo S. 2015
Review objectives	To understand the extant research, identify encouraging practices, and guide future research by evaluating integrated reading and writing interventions for students with learning disabilities (LD) or students with academic difficulties.
Descriptions of interventions/ phenomena of interest	The teaching characteristics of integrated reading and writing interventions (interventions that taught reading and writing to improve reading skills and/ or writing skills) for students with or at risk for LD. The quality of the extant research on reading and writing interventions for students with or at risk for LD according to the What Works Clearinghouse (WWC) design standards. For studies that met WWC design standards with or without reservations, the effectiveness of integrated reading and writing interventions at improving literacy outcomes for students with or at risk for LD.
Descriptions of outcomes included in the review	Researcher developed or standardised reading or writing measures
Descriptions of contexts included in the review	The language of instruction was English. Participants in grades K through 12 or were 5–18 years of age. The students were identified as having LD or struggling readers or writers. Implementers were researchers and teachers (general education, special education, computer lab).
Search details	Three databases: ERIC, Psych- INFO, and Education Resource. Date range: earliest available through 2014. Hand searches of relevant journals from January 2011 to March 2014. Ancestral search pf references from previously published syntheses and meta-analyses (such as, Graham and Hebert, 2011; Hebert et al. 2013; Mason, 2013) of reading, writing, and reading–writing relationships.
Number of studies and participants included	Total studies = 10. The designs were 6 treatment-control group studies (60 percent), 2 single group studies (20 percent), and 2 single-case studies (20 percent). 1975–2015 Total students = 628 in grades 2 through 8 were included. Of those, 248 were students had LD and 380 were students labeled at risk for literacy failure.
Appraisal instruments used	WWC Procedures and Standards Handbook Version 3.0 (Institute of Education Sciences, 2014)

Study	Kang EY, McKenna JW, Arden S, Ciullo S. 2015
Description of main results	• Only 10 studies met selection criteria, 6 of which did not meet the WWC design standards. Three of the 4 studies (Faggella-Luby et al. 2007; Mason et al. 2006, 2012) that met WWC design standards with or without reservations were more recent. • Students with LD responded to interventions that were multicomponent and included explicit training in cognitive strategies which followed the SRSD instructional sequence. This finding is based on Mason et al. 2006 (n = 2) and Mason et al. 2012; (n = 26) used Think before reading, think While reading, think After reading (TWA) + Pick goals, List ways to meet goals, and make Notes, Sequence notes (PLANS) with 4th grade students. • Cognitive strategy instruction interventions (self-questioning, comprehension strategy training, self-regulated comprehension, and explicit writing frameworks) yielded large effects on reading and writing measures for secondary students with LD. Fagella-Luby et al. 2007 (n = 14; 9th grade [Treatment, n = 7 and Control, n = 7] implemented the Embedded Story-Structure (ESS). Students with LD in the treatment condition outperformed students in the control condition on 3 researcher-developed measures: strategy-use (ES = 4.16), knowledge (ES = 3.77), and reading comprehension (ES = 0.87). This study met the WWC design standards without reservations. Gaps: In addition to employing stronger experimental designs and additional replications of encouraging studies, future research should explore the utility of integrated reading and writing interventions with secondary students with who have academic difficulties.

Study	Kilgus SP, Methe SA, Maggin DM, Tomasula JL. 2014
Review objectives	To conduct the first meta-analysis of curriculum-based measurement of oral reading (R-CBM) universal screening and diagnostic accuracy research. to employ a diagnostic test accuracy meta-analysis (DTAM) of the R-CBM literature to evaluate R-CMB evidence to date (a) to generate generalizable estimates of diagnostic accuracy (b) to examine the consistency and robustness of R-CBM performance across multiple investigations, and (c) to identify variables that account for any inconsistency and consider what implications these findings may have to the application of R-CBM.
Descriptions of interventions/ phenomena of interest	Curriculum-based measurement of oral reading (R-CBM) class of tools employing standardised procedures and content in the evaluation of student performance for universal screening for developing services supporting prevention and early intervention services read aloud measure or indicator of oral reading fluency, defined as a measure of a student's ability to read connected text with accuracy and speed and indicative of a range of reading skills, being influenced by the retention of previously learned skills (for example, phonemic awareness) and the acquisition of new skills (for example, vocabulary, fluency, and comprehension).
Descriptions of outcomes included in the review	R-CBM diagnostic accuracy to identify the cut score yielding an optimal SE value (for example, ≥.80). to identify whether the chosen cut score also yields adequate levels of remaining conditional probabilities.

Study	Kilgus SP, Methe SA, Maggin DM, Tomasula JL. 2014
Descriptions of contexts included in the review	Students kindergarten to 8th grade in typical school settings; independent, stand-alone investigations student instructions and scoring of digits or words correct per minute the sample to which results corresponded included individuals at all levels of reading ability, such as those at low, some, and high risk for reading difficulty (per a criterion). Article employed one or more criterion measures that were either a state-wide achievement tests (SWAT) or published norm-referenced tests (PNRT).
Search details	2 databases: PsycINFO and Education Research Complete (ERC) Hand-search of journals R-CBM related technical report repositories Ancestral search Repositories: University of Oregon Center on Teaching and Learning (https://dibels.uoregon.edu/ research/), the Florida Center for Reading Research (http://www.fcrr.org/science/scienceTechnicalReports.shtm), and the University of Oregon Behavioral Research and Teaching (http://www.brtprojects.org/publications/technical-reports). A review of these databases.
Number of studies and participants included	34 studies, including 20 peer-reviewed articles, 7 dissertations, and 7 technical reports
Appraisal instruments used	Authors were unable to control for study quality
Description of main results	• Bivariate hierarchical linear models yielded generalised estimates of diagnostic accuracy statistics, which predominantly exceeded standards for acceptable universal screener performance. • When predicting criterion outcomes within a school year (≤9 months), R-CBM sensitivity ranged between .80 and .83 and specificity ranged between .71 and .73. • Multiple moderators of R-CBM diagnostic accuracy were identified, including: (a) R-CBM cut score used to define risk, (b) lag in time between R-CBM and criterion test administration, and (c) percentile rank, corresponding to the criterion test cut score through which students were identified as either truly at risk or not at risk. Follow-up analyses revealed substantial variability of extracted cut scores within grade and time of year (such as, fall, winter, and spring). The substantial variability in cut scores across studies, as well as the difficulty identifying reliable cut scores linked specifically to grade levels and time of year, point to the potential need for binary diagnostics to shift towards using local normative data for decision making. Gaps: the inflexible application of a single cut score across contexts and suggested the potential necessity of local cut scores. The moderating influence of several important student demographic characteristics, such as gender, ethnicity, socioeconomic status, special education status, and English language learner status.

Study	Kim W, Linan-Thompson S, Misquitta R. 2012
Review objectives	To serve as a guide to educators when applying reading comprehension strategies in their middle school classrooms and enable them to differentiate instruction and tailor instruction to students' needs.
Descriptions of interventions/ phenomena of interest	The effects of critical factors in intervention (type of instructional methods, self-monitoring, components of reading incorporated, fidelity of instruction (scripted vs. non-scripted, and researcher vs. teacher), and group size) on the reading comprehension of middle school students with LD.
Descriptions of outcomes included in the review	Standardised- or researcher-developed comprehension measures of reading comprehension
Descriptions of contexts included in the review	Whole group instruction (12 or more students; k = 3), instruction in groups (6–12 students; k = 5), instruction in pairs (k = 4), and individual instruction (k = 2). Interventions were delivered in one of 4 ways: researcher-delivered instruction with script (k = 5), researcher-delivered instruction with no script (k = 3), teacher-delivered instruction with script, and teacher delivered instruction with no script.
Search details	Educational databases including Academic Search Alumni Edition, Academic Search Complete, ERIC, and PsycINFO. Examination of articles reviewed in previously published syntheses (see Dexter and Hughes, 2011; Gajria et al. 2007; Kim et al. 2004; Swanson, 2008). Date range: 1990 to 2010 in peer-reviewed journals.
Number of studies and participants included	Total studies = 14 (11 articles), experimental, or quasi-experimental designs and included a control or comparison group. A total of 465 students with LD, middle school students (grades 6–8) classified as LD according to the federal and state criteria for eligibility of LD.
Appraisal instruments used	Not reported
Description of main results	• The findings of this review indicate that instructional modifications are helpful in improving reading comprehension skills for middle school students with LD, but their effects vary. For example, some studies (k = 5), using graphic organisers (GOs) resulted in high ESs, while other studies demonstrated little or no difference between intervention and control groups (ES range: 0.07 - 1.77). Additionally, the results on follow-up tests demonstrated that the effects were not maintained. • Middle-school students with LD who used strategies consistently, revealed high reading comprehension scores. In particular, main idea, summarisation, and targeting underlying structures were more beneficial than the other strategies for students. The ESs were high on post-tests as well as on the standardised test (ES range: 0.57 - 5.23). It appeared that the effects of instruction on comprehension varied regardless of the reading components they incorporated. Although this body of literature is not comprehensive, when providing balanced instruction (multiple reading components together), results do point toward the beneficial effects of specifically incorporating vocabulary versus other components in reading comprehension instruction. Among the balanced instruction studies that targeted instructional modifications, instruction that incorporated vocabulary (Bos and Anders, 1990, Studies 1, 2, 3) yielded consistently higher effects (mean ES =1.53) than instruction that incorporated vocabulary and word decoding (DiCecco and Gleason, 2002) on post-tests (mean ES = 0.28).

Study	Kim W, Linan-Thompson S, Misquitta R. 2012
Description of main results	• Overall, the findings showed that the effects of interventions were consistently high when delivered by a researcher, relative to instruction delivered by a teacher. Interventions delivered by researchers with script (k = 5) indicated high effects across post-tests (range ES = 0.86–3.40), but medium to high effects on follow-up tests (range ES = 0.57–3.12). Interventions delivered by researcher with no script (k = 3) reported high effects on post-tests (range ES=1.33–1.78), but smaller effects on follow-up tests (range ES = 0.40–0.47). The effect sizes for studies incorporating teacher-delivered instruction, ranged from small to high. • Overall, analysis revealed that the individual instruction was most effective on post-tests (range ES =1.20–2.57) and follow-up tests (range ES = 1.23–3.12). Effects were inconsistent for the other 3 groups (whole group, groups, and pairs) across studies and measures. The range of ESs was small to large on post-tests (range ES =0.07–5.23), follow-up tests (range ES = 0.40–2.00), and standardised tests (range ES = 0.34–0.84). It appeared that learning in groups (6–12 students) was more effective than the other formats when instructional modifications were employed. Individual instruction was effective in enhancing students' comprehension performance for the studies that employed strategy instruction. Gaps: Durations of the studies examined were too varied to draw any conclusive findings. Duration of instruction in the studies reviewed ranged from a minimum of 3 days to maximum 12 weeks. Also, the frequency per week and duration per session varied significantly across studies. None of the studies reviewed implemented small-group instruction (groups of 3 to 5), a common grouping format in special education (Vaughn et al. 2003).

Study	Kuder SJ. 2017
Review objectives	To update the literature on vocabulary instruction for secondary students with learning disabilities (LD).
Descriptions of interventions/ phenomena of interest	New methods for teaching vocabulary to secondary-age students with LD. Methods that are most effective for this population.
Descriptions of outcomes included in the review	Researcher developed measures including: Strategy use word knowledge, morphological analysis, satisfaction survey; mastery probes; 30-item multiple-choice instrument to assess knowledge of target vocabulary; measures for reading fluency, vocabulary knowledge, comprehension; 10-item quiz. Teacher developed: Assessments of social studies units from the school district, 20-item weekly vocabulary quiz.
Descriptions of contexts included in the review	Instruction was conducted in general education (inclusive) classrooms (k = 6) and a self-contained special education classroom (k = 1). Delivered by researchers (k = 2) and teachers (k = 5).
Search details	3 databases: PsychINFO, ERIC, and Education Full Text databases Date range: 2003 to 2016.

Study	Kuder SJ. 2017
Number of studies and participants included	Total studies = 7; experimental (k = 3), quasi-experimental (k = 2), and single subject (k = 2) designs. Total participants (with LD) = 162 (94 [58%] male; 68 [42%] female) 6th to 10th grade students with LD. Most of the students were receiving special educational services under the category of LDs (114 [88%]). The remainder included students with behaviour disorders (5), intellectual disabilities (4), other health impairments (5), and students with autism (one).
Appraisal instruments used	Not reported
Description of main results	• The current review found that vocabulary instruction for secondary-age students with LD could lead to improvements in word knowledge. Findings from research since 2003, showed that mnemonic instruction (k = 1, n = 8), learning strategies that utilised morphemic analysis (k = 1; n = 24), and direct instruction (k = 1, n = 20) had large effect sizes. • Using a mnemonic technique such as the keyword method can be highly effective for improving the vocabulary acquisition of secondary- age students with LDs. Terrill et al. (2004) found that vocabulary recall of Grade 10 students with LD (n = 8) was much better when they used the keyword method than when the non-mnemonic methods were used. In the mnemonic (keyword) condition, students recalled nearly 92% of the vocabulary words whereas in the non-mnemonic condition they recalled only less than 50% (48.8 %). The effect size was large (d = 2.743). • Students with disabilities are able to utilize a strategy that goes beyond simple memorization of word meanings. Harris et al. (2011) found that students (9th grade English classes) in the morphological strategy, Word Mapping Strategy (WMS) group outperformed their peers who utilised the alternative meaning memorisation (Vocabulary LINCS; Ellis, 1992) strategy as well as those in the control group on a test of morphological analysis skills. Although students with disabilities (n = 24) in both of the intervention groups made significant gains on both the tests of word knowledge (d = 4.226, p = .003) and morphological analysis (d = 6.942, p < .01), students without disabilities (n = 206) outperformed them on both measures. • For students with LD, instruction that focuses on vocabulary knowledge is needed. Seifert and Espin (2012) found that 10th-grade students with LD who were enrolled in regular education biology classes learned more vocabulary words in the vocabulary learning (direct instruction; p = .000, ES = 1.12,) and the combined (p = .00, ES = 1.11) conditions than in the text reading (repeated reading) condition. Neither the text reading (repeated reading) nor vocabulary learning conditions resulted in improved reading comprehension. Gaps: There is a need for more well-designed research that includes larger sample sizes, appropriate comparison groups, and the consistent use of measures to ensure implementation fidelity. Additionally, using both standardised and researcher developed outcomes measures would be useful for determining the extent to which intervention methods lead to generalization of learning to new vocabulary items.

Study	Lee J, Bryant D, Ok M, Shin M. 2020
Review objectives	To identify and understand intervention studies that focus on algebraic concepts and skills for secondary school students with learning disabilities (LD)
Descriptions of interventions/ phenomena of interest	Algebraic concepts and skills taught to secondary school students with LD during the interventions. Instructional components of the effective interventions.
Descriptions of outcomes included in the review	Mathematical performances: Researcher developed measures (n = 11); standardised assessments (n = 1); and, researcher-developed and standardised tests (n = 1).
Descriptions of contexts included in the review	Settings: vacant classrooms (n = 3), conference or supply room (n = 3), 2 math classrooms (n = 1), resource classroom (n = 1), an office (n = 1); no reported (n = 1). United States, and Canada (n = 1). Interventionist: Researchers (n = 9) or teacher implemented interventions (n = 3), including a math teacher (n = 1), and a special education teacher (n = 1). Instructional grouping: One-on-one instruction (n = 8); pair (n = 2); small-group (n = 1); and, whole class (n = 1).
Search details	Four databases: Education Resources Information Center, Academic Search Complete, Education Source, and PsycINFO Hand search. Ancestral search. Date range: January1969 to February 2019.
Number of studies and participants included	Total = 12 studies. Of the 12 studies, 10 implemented single-case design (SCD, and 2 used a quasi-experimental (QED) treatment comparison design. Total students = 89 in grades 6 through 12: 6 in grade 6, 5 in grade 7, 20 in grade 8, 19 in grade 9, 7 in grade 10, and 32 in grades 11 and 12. Identification criteria for LD using 2 different approaches: IQ-achievement discrepancy model (n = 3); school districts' RTI processes (n = 3); the schools' processes following each state's guidelines (no detailed information; n = 3); not reported (n = 4).
Appraisal instruments used	Not reported.

Study	Lee J, Bryant D, Ok M, Shin M. 2020
Description of main results	• Overall, students with LDs showed improved performances on algebraic concepts and skills following the interventions. The Tau-U values and Hedges' g were typically large or very large for most participants. However, the outcomes should be interpreted with caution because of variability for individual participants, and because the values to determine the magnitude of the ES can be considered arbitrary. Also, most studies only used researcher-developed measures to determine student improvement on algebraic concepts and skills; more distal standardised measures may have shown minimal changes in general student performances. • Using multiple representations in combination with a sequence and/or range of examples were very effective. Ten out of 11 studies using these instructional components demonstrated highly positive outcomes on algebraic notions and competencies. Half of the studies used CRA sequencing instruction to teach algebra principles. The other half involved multiple representations (for example, concrete and virtual manipulatives, pictures) with a logical sequence of examples (for example, addition, subtraction, multiplication, division) to teach algebraic concepts and abilities. • Explicit instruction in conjunction with multiple representations and a sequence and/or range of examples led to growth in students' algebra concepts and skills. All the 10 studies that applied this combination of instructional components demonstrated students' growth in relation to algebra concepts and skills. This study's findings corroborate the effectiveness of the features of explicit instruction (including modelling the steps for solving problems, sufficient practice opportunities, and ongoing feedback) to enhance instruction with combinations of various visuals and work examples. • Student verbalisation with combinations of multiple representations and a sequence and/or range of examples led to improved algebraic concepts and skills for secondary students with LD. Eight of the 12 studies that used this combination of instructional components led to students' improvement regarding algebraic concepts and skills. In the studies, students were encouraged to think aloud their strategies for solving problems, and to communicate their reasoning with their peers. Gaps: Future intervention studies should investigate applying the understanding of numbers to the rational number system, as well as defining and using functions to model relationships between quantities (in addition to creating and solving linear equations, expressions, and inequalities with one or 2 variables and forming quadratic expressions).

Study	Lee J, Yoon SY. 2017
Review objectives	To provide statistical support through a meta-analysis of pedagogies that promote Repeated Reading (RR) intervention.
Descriptions of interventions/ phenomena of interest	Responses to repeated oral reading intervention for students with reading disabilities (RD) on reading fluency by comparing add-on intervention variables (for example, word preview, listening passage preview, systematic error correction [SEC], performance feedback).
Descriptions of outcomes included in the review	Correct words per minute (CWPM).
Descriptions of contexts included in the review	Reading fluency interventions conducted in English. No other details reported.
Search details	Electronic databases including Education Resources Information Center, PsycINFO, ProQuest Dissertation and Thesis Database, and ArticleFirst and search engines such as Google Scholar. References of identified articles through online searching. Additional in-depth searches were carried out on the most relevant journals, researchers, and ancestry searches. Date range: 1990 to 2014.
Number of studies and participants included	39 independent effect sizes were identified out of 34 experimental studies Students with RD or at risk for RD (grouped as RD for this article as pre-test differences in reading fluency were ns). Total students (in 29 ESs) = 274; elementary (grades 1 - 5) = 142, middle and high school (grades 6 - 11) = 132. Less than 10% of students had a reading age of grade 6 or above (max RA = grade 8).
Appraisal instruments used	Not reported.
Description of main results	• Repeated Reading has a large positive effect on reading fluency for students with reading disabilities or at risk of reading disability. Under the random-effects model, considering that all the studies were not conducted under the same conditions, the estimated overall average Hedges' g for the RR effects was 1.41 ($z = 11.18$, $p < .001$) with a 95% confidence interval between 0.99 and 1.41, which is a relatively large effect. • Repeated Reading was effective for both elementary and secondary students. However, the RR intervention was more effective on students at the elementary reading level (Hedges' g = 1.25) than on students at the secondary reading level (Hedges' g = 0.80). The difference between the 2 subgroups was statistically significant, $Q(1) = 7.89$, $p = .005$. Less than 10% of all students had a reading age of grade 6 or above (max RA = grade 8). • The responsiveness to the RR intervention appears to be attributed to a listening passage preview prior to the RR. The effect of the RR intervention with the listening passage preview (Hedges' g = 1.95) was significantly higher than the RR intervention without the listening passage preview (Hedges' g = 0.94). The difference between the conditions was statistically significant, $Q(1) = 8.97$, $p = .003$.

Study	Lee J, Yoon SY. 2017
Description of main results	• Reading the passage at least 4 times increased reading fluency more than 2 to 3 times. When the maximum number of repeats during the intervention was sub-grouped by 2, 3, and 4 and more repeats, the difference among the 3 groups was statistically significant, Q(2) = 12.68, p = .002. When the maximum number of repeats was 4 and more, the effect of RR intervention (Hedges' g = 1.73) was significantly different from 2 and 3 repeats (Hedges' g = 1.45 and 0.82, respectively). Even though the effect of 3 repeats was smaller than 2 repeats, the difference between the effects was not statistically significant. • The effects of RR on reading rate was observed only by using non-transfer practiced passages for students with RD. The effect of generalised transfer passages (Hedges' g = 0.97) was smaller than the effect of non-transfer practiced passages (Hedges' g = 1.94). The differences between the textual factor condition was statistically significant, Q(1) = 10.32, p = .001. • The analyses showed a significant difference in the estimated mean effect size by type of documentation. This implies the existence of a publication bias since studies published as journal articles showed, on average, a larger RR intervention effect (Hedges' g = 1.55) than studies published as dissertations (Hedges' g = 0.82). Gaps: The studies on RR that we examined provided insufficient information (for example, levels of training, duration, and total sessions of the interventions) and reading textual factors (for example, text difficulties), as was similarly concluded by Chard et al. (2009, p. 277). The consistency of dependent variables also varies among the studies reported. We used a standard CWPM measure; however, studies may have used this measure on passages previously read, on novel passages, or on norm-referenced passages.

Study	Lein AE, Jitendra AK, Harwell MR. 2020
Review objectives	The overall effectiveness of word problem-solving interventions on the mathematical performance of school-aged (K–12) students with LD and/or MD.
Descriptions of interventions/ phenomena of interest	Mathematics intervention impacts variation by characteristics of participants, interventions, and study design.
Descriptions of outcomes included in the review	Researcher-developed measures and standardised measures.
Descriptions of contexts included in the review	General education and special education classrooms (and non-specified other).
Search details	Three databases: ERIC, PsycINFO, ProQuest Digital Dissertation, also ancestral and hand search as well as date range: 1989 – 2019.
Number of studies and participants included	31 studies 2,841 participants (1,258 were in year 7).
Appraisal instruments used	No critical appraisal homogeneity was assessed, accounting for bias in samples.

Study	Lein AE, Jitendra AK, Harwell MR. 2020
Description of main results	• Word problem-solving interventions yielded a moderate positive mean effect size (g = 0.56, 95% confidence interval [0.40, 0.72]). The finding was large enough and yielded statistically significant and meaningful effect to support the conclusion that mathematics word problem solving interventions can have positive effects for students with LD and/or MD. • The relation between word problem-solving interventions and student learning was moderated by grade level with a larger average effect size for elementary students than for secondary students (elementary g = 0.63, secondary g = 0.33). • Statistically significant differences between effect sizes obtained for intervention model type were as follows: schema-based instruction g = 0.40, schema broadening (and transfer) instruction (SBTI) g = 1.06, strategy (cognitive and explicit) instruction g = 0.28, "other" (for example, computer-assisted instruction, diagramming, inquiry-oriented practices) g = 0.11. • There was a statistically significant difference between results by implementer: researcher g = 0.71, school personnel g = 0.28, and researcher and school personnel g = 0.76. • There was no statistical difference by implementation of intervention (whole class vs. small group; p = .90), intervention dosage (> 16 hr vs. > or equal to 10 to < or equal to 16 hr vs. < 10 hr; p = .94), or mathematics task (arithmetic word problem solving vs. word problem solving with fraction, ratio, proportion, percent, and algebra; p = .81).

Study	Maccini P, Mulcahy CA, Wilson MG. 2007
Review objectives	To determine: (1) the nature and focus of math interventions that are effective for assisting secondary school students with LD and (2) implications for both research and practice.
Descriptions of interventions/ phenomena of interest	Type of intervention (behavioural, cognitive, or alternative systems delivery), focus of intervention, academic skill or concept targeted, and nature of results.
Descriptions of outcomes included in the review	Various measures of mathematics performance. ([Six mathematics domains: basic skills (addition, subtraction, multiplication, division) (4), fractions (6), decimals (3), geometry (1), algebraic skills and concepts (5), as well as problem-solving tasks (7).)
Descriptions of contexts included in the review	Secondary school students (grades 6-12) with learning disabilities LD. Intervention settings: resource room/class (n = 6), special education classroom/ self-contained class (n = 3), regular classroom (n = 3), learning support class (n = 1), separate class (n = 2), school conference room (n = 1), remedial math/ math classes (n = 5), not reported (n = 2).
Search details	1. Research Port - a computerised search engine - that provided access to multiple databases, including PsychInfo, ERIC, and the Exceptional Child Education Resources. 2. Hand search of prominent special education journals. 3. Ancestral search of references Date range: September 1995 to May 2006.

Study	Maccini P, Mulcahy CA, Wilson MG. 2007
Number of studies and participants included	23 studies (single-subject or group design). Total number of participants was 1,134; 384 identified as having LD. 14 studies included only participants with LD; the remaining studies included students with various other disabilities (such as, behavioural or emotional disorders, ADHD), students with remedial mathematics needs, and students with no disabilities.
Appraisal instruments used	None described.
Description of main results	• A number of practices produced significant gains for secondary school students with LD in mathematics in comparison to control conditions on post-test measures. • The practices resulting in significant ES included: (1) mnemonic strategy instruction (range of d: 1.68 to 2.22); (2) graduated instructional approach (range of d: 0.43 to 3.87); (3) cognitive strategy instruction involving planning (range of d: 0.2 to 1.4); (4) schema-based instruction (range of d = 1.69); and (5) contextualised videodisc instruction (range of d: 0.56 to 0.81). • Studies with significant treatment effects incorporated features of effective instruction (modelling, guided practice, independent practice, monitoring student performance, and/or corrective feedback). These interventions increased students' performance and retention in computational skills with decimals, fractions, geometry, integers, solving linear equations and problem-solving tasks.

Study	McClain MB, Haverkamp CR, Benallie KJ, Schwartz SE, Simonsmeier V. 2021
Review objectives	To add to the existing literature by determining the current state of reading comprehension intervention studies and the effectiveness of various reading comprehension interventions for children and adolescents with ASD by conducting a meta-analysis of single-case design (SCD) studies.
Descriptions of interventions/ phenomena of interest	Overall effectiveness of reading comprehension interventions for children with ASD (reading comprehension interventions are conceptualised as one-on-one interventions designed to improve reading comprehension skills). Moderator variables: race/ethnicity, age, disability, intervention length, cognitive ability, and SCD method quality.
Descriptions of outcomes included in the review	Percent increase in reading comprehension (number of questions answered correctly).
Descriptions of contexts included in the review	Schools and clinics Interventionists were researchers and teachers.
Search details	ERIC, PsycINFO, and PubMED databases ancestral search grey literature (from PsycINFO). No specified date range (not restricted).
Number of studies and participants included	20 studies, single case designs. 60 participants. the specific studies included children between the ages of 6 and 17 with the ages of 8 and 10 being the most common.
Appraisal instruments used	What Works Clearinghouse (WWC) SCD standards (Kratochwill et al. 2010).

Study	McClain MB, Haverkamp CR, Benallie KJ, Schwartz SE, Simonsmeier V. 2021
Description of main results	• Overall, the reviewed reading comprehension interventions implemented improved reading comprehension skills for children with ASD according to effect size analyses. BC-SMD (k = 14) results indicated that, although variable across studies, the number of comprehension questions answered correctly in the intervention phase increased almost 3 standard deviations (BC-SMD [overall] = 2.97, 95% CI = 1.85–3.97, p < .001) from the baseline phase at a 10 session follow up). • Reading comprehension interventions were differentially effective for children with ASD and were dependent on child race/ethnicity, disability, and age. Further analyses indicate that intervention effectiveness, although broadly effective, differs across child race/ethnicity, $\chi2(4) = 40.4$, p < .001; age, $\chi2(2) = 52.0$, p < .001; participant disability (ASD vs. ASD + speech language impairment), $\chi2(2) = 44.3$, p < .001; and study method quality (meeting WWC vs. meeting WWC with reservations), $\chi2(2) = 150.4$, p < .001. • specific to race/ethnicity moderator analyses should be interpreted with caution due to low power. Within race/ethnicity, Black (184% increase, p = .006), Latinx (230% increase, p = .018), and White (451% increase, p = .003) children saw significant increases in reading comprehension skills from baseline to intervention. Conversely, Asian children did not show significant increases in reading comprehension from baseline to intervention (204% increase, p = .090). • Both older and younger children with ASD groups showed significant improvements in reading comprehension post intervention. Younger children (240% increase, p < .001) saw greater gains than older children (135% increase, p = .003). • Children with ASD only exhibited significant gains in reading comprehension skills. Regarding disability, children with ASD (193% increase, p < .001). Children with ASD and SLI did not show significant increases in reading comprehension skills (189% increase, p = .092). • Research method quality also contributed to intervention outcomes. Participants in studies with designs that met WWC standards with reservations (342% increase, p < .001) showed greater improvements in reading comprehension than participants in studies with designs that met WWC standards (66% increase, p = .002). Gaps: Researchers should continue conducting SCD research that specifically evaluates the effectiveness of reading comprehension interventions for children with ASD. The body of research on this topic is emerging, but still limited. At this time, there are insufficient studies to determine evidence-based reading interventions for improving comprehension. for children with ASD.

Study	Newell KW, Codding RS, Fortune TW. 2020
Review objectives	To critically evaluate the available validity evidence in support of oral reading fluency (ORF) as a screening tool for use with emerging biliterates with different language backgrounds.
Descriptions of interventions/ phenomena of interest	The relationship between ORF scores and reading comprehension measures for emerging biliterates. The relationship between ORF scores and other reading outcome measures for emerging biliterates. Evidence supporting the use of ORF as a tool to accurately identify emerging biliterates at-risk for poor reading outcomes.
Descriptions of outcomes included in the review	ORF probes administered in the standard manner with WRC scores calculated (see Shinn, 1989), which included (a) only words read correct in connected text, not word lists, and (b) passages levelled or controlled for difficulty in some manner. ORF probes in any language were accepted. All studies but 3 (10%; Baker, 2007; Baker et al. 2012; Keller-Margulis et al. 2012) used criterion measures only in English. Sixteen studies (52%) measured reading comprehension. The most common measure used was the Stanford achievement test—10th Edition reading comprehension task (SAT10; Harcourt Brace, 2003; n = 4; 13%). Many studies used available state assessment results (n = 16; 52%) as criterion measures. Finally, 3 studies (10%; Baker and Good, 1995; Betts, Muyskens, and Marston, 2006; Jimerson, Hong, Stage, and Gerber, 2013) used a broad reading measure not part of a state accountability test.
Descriptions of contexts included in the review	English language learners (ELs) US (k = 30) and ELs in South Africa (k = 1).
Search details	3 EBSCO databases: ERIC, EducationSource, and PsycINFO. A search for publisher technical reports or manuals was conducted on publisher websites (AIMSweb, DMG/ DIBELS, FastBridge Learning, easyCBM) as well as the National Center on Response to Intervention website (www.rti4success.org). Ancestral search. Date range: 1980 to 2019.
Number of studies and participants included	Total articles = 31; 18 peer-reviewed, published articles, 12 dissertations, and one technical report. Emerging biliterates in Kindergarten through eighth grade. The majority of studies included Spanish-speaking students learning English (n = 19; 61%). Nearly half of the studies provided no information about the English language proficiency of participants (n = 13; 42%). In those studies that described language proficiency of participants with categories, most studies had the majority of students at the intermediate English (L2) proficiency level. English (L2) proficiency tended to increase as grade levels increased (Millett, 2011).

Study	Newell KW, Codding RS, Fortune TW. 2020
Appraisal instruments used	To capture the quality of the literature, articles were also coded based on 6 indicators (Gersten et al. 2005; Jitendra, Burgess, and Gajria, 2011; Talbott, Maggin, Van Acker, and Kumm, 2017). The quality indicators are presented in Table 1. Quality indicator one (Q1) considers whether authors sufficiently described the participants, including participant language proficiency in L1 and L2, and languages spoken. Quality indicator 2 (Q2) rates the extent to which authors described the type of L2 instructional program or model received by participants. The third quality indicator (Q3) rates authors on whether outcome or criterion measures were sufficiently described. Quality indicator 4 (Q4) addresses whether authors considered sample size and sampling issues, and quality indicator 5 (Q5) reviews whether statistical analysis assumptions were considered. Finally, quality indicator 6 (Q6) addresses whether authors sufficiently describe included variables with descriptive statistics.
Description of main results	• In the reviewed studies, correlations between ORF scores and reading comprehension scores ranged from 0.58 (n = 30; Beattie, 2018) to 0.82 (n = 102; Nam, 2012). Correlations were generally higher in lower grades (for example, third grade, r = 0.75; Baker. 2007) than in upper grades (for example, 5th grade, r = 0.44; Crosson and Lesaux, 2010). Only 3 articles included students in grades 6 and 7. Correlational and diagnostic evidence showed Sixth grade Winter ORF to spring ISAT: 0.6SE: 0.97; SP: 0.26, n = 59; Sixth grade Fall ORF/spring AIMS: 0.577 n = 90; Sixth grade ORF/ OAKS: 0.75 n = 36 and Seventh grade ORF/ OAKS: 0.51 n = 154. • In the reviewed studies that used broad reading rather than reading comprehension measures, correlations between ORF and reading outcomes did not have a decreasing pattern across grades. Correlations ranged from 0.36 (n = 109; J. S. Kim, 2012) to 0.75 (n = 33; Sáez et al. 2010). Almost all of the correlations between ORF scores and broad reading measures fell within the range of moderate to strong, which mirrors the results found with monolingual English-speaking students. • Based on the reviewed studies, accurate identification of students at risk of failing an outcome assessment would lead to over-identification of ELs, or accurate identification of students passing an outcome assessment would lead to under-identification of ELs. In all studies but one (Scheffel et al. 2012), when SE met desired criteria of 0.8, specificity did not meet criteria, and when specificity met criteria of 0.7, SE did not meet criteria. Gaps: The quality of the reviewed studies was low in critical areas. The majority of studies did not meet criteria for the description of participants and the description of L2 instructional program. First, future research must consider emerging biliterates as multiple populations. Next, future research must define and measure variables relevant to biliteracy. Finally, future research should consider alternatives to a single ORF assessment as a screening tool.

Study	Oxley E, de Cat C. 2021
Review objectives	To synthesise the language and literacy outcomes of recent interventions with a population of EAL pupils undertaking their schooling in English.
Descriptions of interventions/ phenomena of interest	Methods of intervention that are most effective for language and literacy outcomes.
Descriptions of outcomes included in the review	Various vocabulary and reading comprehension measures.
Descriptions of contexts included in the review	Primary or secondary schools in 26 studies in this review originated in the US (k = 26); the United Kingdom (k = 1), in Canada (k = 1) and in Lesotho (k = 1). Language of instruction was English Language of the wider community was English.
Search details	Four databases: British Education Index, Web of Science, Educational Resources Information Centre and Language and Linguistics Behavioural Abstracts. Date range: 2014 to 2017.
Number of studies and participants included	There was a total of 32 languages spoken (in addition to English), including pupils who spoke Spanish (k = 15), 3 of these interventions were with groups of pupils all of whom spoke Spanish. Six interventions did not specify the languages spoken by the pupils. Numbers of participants overall and for included studies (in tables) was not reported.
Appraisal instruments used	Cochrane Risk of Bias tool (Chandler et al. 2014).
Description of main results	• Language focused interventions had small to large effects (d = 0.05 to d = 2.21) on language development of children and adolescents with English as an additional language (EAL). Of the 11 language-focused interventions, 10 had a primary focus on vocabulary, implemented with differing intervention methods, such as academic vocabulary explicitly taught through morphology, vocabulary explicitly taught with sign-language support or vocabulary encountered during shared reading. However, 6 interventions were classified as having a high risk of bias, and therefore large effects must be interpreted with caution. • Interventions focused on the literacy development of children and adolescents with EAL had small to medium effects on language and literacy development. A range of effect size measures and ranges was reported, such as Cohen's d (0.035 to 0.85), Hedges' g (0.06 to 0.69, η2 (0.01 to 0.518), Cramers V (.337 to .372). These fifteen studies used a range of interventions including technology as the primary intervention platform (k = 3), teacher-led reading comprehension (k = 5), continued professional development (CPD) with reading comprehension as a primary outcome (k = 4), and family literacy programmes (k = 2). The risk of bias across studies was mainly high (k = 6) or medium (k = 8), with only one study being classed as low, so results should be interpreted with caution.

Study	Oxley E, de Cat C. 2021
Description of main results	• Interventions focused on improving reading comprehension provided less evidence of intervention gains for children and adolescents with EAL. The studies reported below targeted reading comprehension as the overarching aim of the intervention. Their effects ranged from none to medium. Furthermore, bias assessments were classed as high (n = 4) or medium (n = 2). Gaps: Only 4 of the interventions in this review were carried out on upper secondary school children. This review highlights the paucity of interventions for older children. More interventions taking place over longer time spans are needed to fully investigate how to equip EAL learners with sufficient skills to comprehend texts.

Study	Peterson AK, Fox CB, Israelsen M. 2020
Review objectives	To synthesise the effectiveness of existing narrative and expository discourse interventions for late elementary– and middle school–aged students with language-related learning disabilities (LLD).
Descriptions of interventions/ phenomena of interest	Interventions that target narrative and/or expository discourse for school-aged children ages 9 to 14 years with LLD. Interventions conducted with at least 80% fidelity that produce both statistically and practically significant improvements in discourse comprehension and/or production. Based on the Gillam and Gillam (2006) Critical Appraisal Standards, at what level of quality were these studies conducted.
Descriptions of outcomes included in the review	One or more outcome measures of discourse comprehension and/or production. Outcome measures of comprehension included multiple-choice or true/false questions about literal or inferential information from the text and outcome measures of production included written (such as, essays, short answers) or oral presentations.
Descriptions of contexts included in the review	Academic setting directed by a teacher, clinician, or researcher Grades 4 to 8.
Search details	Electronic database for the American Speech- Language-Hearing Association (ASHA) and the EBSCOhosted databases PsychInfo and ERIC. Ancestral search.
Number of studies and participants included	Total studies = 7, group level experimental designs (comparison/control group) Total participants with LLD = 287. Participants were within the specified age range (such as, ages 9–14 years or Grades 4–8). Participants were students with LLD who had difficulties in language, reading, and writing, qualified for an individualised education plan, or were considered “at risk” for a disability by their school or state criteria.
Appraisal instruments used	Critical Appraisal Standards (Gillam and Gillam, 2006), adapted from Dollaghan (2004).

Study	Peterson AK, Fox CB, Israelsen M. 2020
Description of main results	• Six of the 7 included studies reported that school-aged children with language-related learning disabilities (LLD) made statistically significant improvements on outcome measures following academic discourse interventions. Effect sizes ranged from small to large, with the Ukrainetz (2019) Sketch and Speak intervention producing the largest group difference (g = 1.199; n = 44) at post-test on near transfer tasks. This supports the potential for explicit instruction in expository discourse in the area of written note quality within the treatment setting. • Appraisal points for each study based on the Critical Appraisal Standards (Gillam and Gillam, 2006) indicate that clinicians can have moderate confidence in the results presented in this synthesis. However, of the 7 studies included in this review, 4 of the studies reported fidelity. Of these studies, only 3 reported 80% or higher fidelity, an important consideration for treatment effects due to intervention rather than maturational effects of participants. • No conclusive statements about the impact of moderating variables can be made given the nature of this review [qualitative]. However, there were a number of consistent factors across the study designs and implementation that may have affected outcomes that warrant future investigation. First, studies with participants diagnosed with only one component of LLD (such as, a singular impairment in either reading or language) showed higher post-test gains than those that included participants with overall academic performance difficulties. Second, larger effect sizes for the interventions with a moderate dosage (such as, 180 to 300 min across 6 to 8 weeks). The final potential moderating variable was the intervention administrator; the greatest post-test gains (note quality measured through 5 indices: format, brief, sufficient, and paraphrasing; g = 1.199) was intervention administered by familiar, trained SLPs in a one-on-one or one-on-2 setting. Gaps: The low number of group-level experimental design studies on narrative intervention for this population could identify a need for future research in this area. There was minimal evidence that these skills generalised to comprehension or production measures at post-testing. No study in this review used delayed follow-up testing on student independence and use of strategies in other learning environments.

Study	Reed DK, Cummings KD, Schaper A, Biancarosa G. 2014
Review objectives	To conduct a systematic review of studies to address the question: To what extent is assessment fidelity reported in reading intervention research conducted in elementary and middle schools?
Descriptions of interventions/ phenomena of interest	Features hypothesised to influence the fidelity with which assessments might be administered (for example, tester expertise, training of testers, monitoring of administrations, whether the testers were blind to condition, alternating test forms, number of testing days, distractions in the environment, and verification of scoring). Assessments used were standardised measures of reading with publicly available technical manuals. Instructional interventions were focused on reading skills: word identification, fluency, vocabulary, or comprehension.
Descriptions of outcomes included in the review	Standardised measures of reading with publicly available technical manuals.
Descriptions of contexts included in the review	School settings.
Search details	Three databases: ERIC, Academic Search Complete, and PsycINFO. Search of the references of recent reading intervention syntheses and meta-analyses.
Number of studies and participants included	Total studies = 46 - experimental or quasi-experimental designs with a defined comparison group. Students in Grades K - 8 (for grades 6 to 8, k = 36). At least some of the students were identified with a learning disability (LD), reading disability (RD), dyslexia, or at-risk for reading failure. The total number of students ranged from a low of 20 to a high of 576, with a median of 86 students. The number of treatment students ranged from 11 to 279, with a median of 47 (k = 43).
Appraisal instruments used	Not reported

Study	Reed DK, Cummings KD, Schaper A, Biancarosa G. 2014
Description of main results	• None of the K to 8 reading intervention research studies reviewed reported sufficient information for determining whether all screening, pre-, interim, and post-measures were administered and scored as intended. It was interesting to note that only 11 of the 46 studies (24%) included in this review did not overtly state whether IMPLEMENTATION fidelity was monitored. • There is a potential for construct irrelevant variance in the measurement of study participants' reading outcomes. Construct irrelevant sources of variance may include: use of at least some extant data from state assessments (15 studies reported some data were archival from state assessments, locally mandated tests, special education records, intelligence tests, or an unspecified testing requirement resulting in accessible records); test administrators (the majority of studies [k = 26] did not provide any information on the individuals who administered the assessments), their training, tester to student ratios and time durations; test scoring errors (only 4 studies reported double scoring); tester bias (few studies[k = 2] reported blindness to conditions for all measures); environmental distractions (most studies [k = 45] used individually administered assessments that would necessitate hearing the student clearly in order to score responses but few (k = 4) reported using secluded, distraction-free or quiet locations). Gaps: Findings suggest assessment integrity data are rarely reported. Developing clear standards for reporting testing integrity will improve the reliability of future attempts to examine the presence, quality, and impact of the information. Future research might directly query researchers to determine what kinds of training, monitoring, calibrating, and data checking or double scoring were enacted as well as what kinds of problems, if any, were experienced. Studies might also explore whether aspects of training and administration can be manipulated to improve assessment fidelity in both highly controlled research studies as well as more naturalistic environments where teachers test their own students for instructional purposes.

Study	Roberts CA, Leko MM, Wilkerson KL. 2013
Review objectives	To describe the literature on reading interventions for adolescents with significant cognitive disabilities; to evaluate the characteristics of each intervention to determine whether they represent the comprehensiveness recommended by Erickson et al. (2009) and recent reports focusing on improving reading instruction for adolescents; and, to discuss implications of these results for current reading practices for adolescents with significant cognitive disabilities and discuss potential promising instructional practices that warrant further exploration.
Descriptions of interventions/ phenomena of interest	Reading interventions for adolescents with significant cognitive disabilities.

Study	Roberts CA, Leko MM, Wilkerson KL. 2013
Descriptions of outcomes included in the review	One or more measures of reading, such as vocabulary, the dependent measure in 12 studies (63%), was the most frequent component; also, reading fluency, comprehension, and phonics. All outcomes were researcher developed (seemingly) informal measures.
Descriptions of contexts included in the review	Specified secondary setting (such as, middle school or high school). Self-contained classroom (k = 12; 63%); segregated school (k = 3; 16%); general education classroom (k = 1; 5%); self-contained classroom and a community grocery store (k = 1; 5%); self-contained classroom and a transition centre at a local university (k = 1; 5%); and, secondary school with no other details were specified (k = 1; 5%). The teacher implemented reading intervention in 8 (42%) cases. Researchers implemented the intervention in 4 (21%) of the studies. In 2 studies (11%), the researcher, who was also the classroom teacher, delivered the intervention. In 1 study (5%), the intervention was delivered via a computer and 1 study (5%) utilised peers as interventionists. Two studies (11%) had more than one interventionist and used the following combinations of individuals or tools to implement the intervention: (a) teacher and peers and (b) researcher and paraprofessional. The remaining study (5%) did not specify who implemented the intervention.
Search details	Four databases: Academic Search Premier, Educational Resource Information Clearinghouse (ERIC), ProQuest and PsychINFO. Used the same databases to search specific journals that focus primarily on significant cognitive disabilities. Searched the databases for recurrent author names. Ancestral searches of relevant existing literature reviews and hand searched issues published from 2008 through 2011. Date range: 1975 and 2011.
Number of studies and participants included	Total = 19 empirical intervention studies. Sixteen of the 19 studies (84%) utilised single-case design; 3 studies (16%) used a group design. All the group designs utilised pre- and post-test measures though some studies were single-group studies and others had multiple groups. Total of 101 student participants; the majority of the participants had a moderate or severe cognitive disability and the average IQ of participants was in the moderate to severe disability range. Fifty-four (53%) of the student participants were male and 39 (39%) were female (k = 18).
Appraisal instruments used	Not reported.
Description of main results	• There is a significant gap in research on the literacy skills of adolescents with significant cognitive disabilities. Only 19 studies have been published since 1975 with a focus on this specific population. The characteristics of the 19 studies suggest literacy interventions lack the comprehensiveness mandated by policy and advocated for by the field. • Only one study targeted academic reading content in the inclusive general education classroom. Out of the 19 studies, 13 (68%) of the interventions targeted functional reading content and 4 (21%) targeted academic reading content from the core curriculum. 2 (11%) targeted academic and functional content. The 13 studies solely focused on functional content did not provide access to the general education curriculum.

Study	Roberts CA, Leko MM, Wilkerson KL. 2013
Description of main results	• The literacy interventions present in the field for adolescents with significant cognitive disabilities do not appear to be comprehensive in nature. Vocabulary, through sight word acquisition was the most frequent component of reading addressed in the reading interventions, such asthat is, 12 studies (63%). Two studies (11%) addressed fluency and only 1 (5%) addressed comprehension only. None of the 19 studies addressed phonemic awareness skills. Three of the studies (16%) addressed vocabulary and comprehension and 1 study (5%) addressed phonics and fluency. Although comprehension was addressed in 4 studies (21%), it was most often addressed through the receptive identification of a previously learned sight word as opposed to comprehension of an authentic text. • While some studies incorporated modified or authentic texts, most studies relied solely on simple materials such as flash cards with target words written on them. The majority of the studies (n = 9, 47%) used flash cards to display target words for instruction. Three studies (16%) utilised commercially made materials (such as, created for use with commercial curricula, including Corrective Reading, Ready to Read, and Story Box) and 3 (16%) utilised a combination of flash cards and actual grocery items or cooking products. Two studies (11%) utilised flash cards paired with a form of technology (such as, computer, video, SMART Board). Finally, one study (5%) utilised teacher-adapted literature and one study (5%) solely utilised computer-delivered instruction. • All studies reported positive outcomes for adolescents with significant cognitive disabilities in terms of improvements in the targeted elements of reading. “Potential Promising Practices” identified were: sight word instruction using time-delay procedures, integrating functional and academic content, incorporating technology and, modifying grade-level authentic text. Gaps: This scarcity of research on comprehension indicates that researchers must continue to identify innovative ways to assess adolescents with significant cognitive disabilities’ understanding of what they read beyond simply demonstrating an understanding of sight words.

Study	Scammacca N, Roberts G, Vaughn S, Edmonds M, Wexler J, Reutebuch CK, et al. 2007
Review objectives	To advance the knowledge of technical assistance providers working with state departments of education and local education agencies concerning reading-related issues for students with reading difficulties and learning disabilities (LD).
Descriptions of interventions/ phenomena of interest	Effectiveness of the reading interventions for adolescent struggling readers that have been examined in research studies. The specific impact of these reading interventions on measures of reading comprehension. The specific impact of these reading interventions on students with learning disabilities. Interventions that were focused on word study, fluency, vocabulary, reading comprehension strategies, or multiple components of reading instruction.

Study	Scammacca N, Roberts G, Vaughn S, Edmonds M, Wexler J, Reutebuch CK, et al. 2007
Descriptions of outcomes included in the review	Various assessments of reading or reading-related variables, including standardised, norm-referenced measures (k = 11), and researcher/teacher developed measurers.
Descriptions of contexts included in the review	Specific contexts not reported; grades 4 – to 12.
Search details	Searches of electronic databases, by perusing reference lists of prior syntheses on related topics, and by researching citations to assure a comprehensive pool of eligible studies. Date range: 1980 and 2006.
Number of studies and participants included	Total studies = 31 - multiple-group experimental or quasi-experimental designs. Total students = 1329; English-speaking struggling readers in grades 4-12; All with learning disabilities (LD) = 720.
Appraisal instruments used	Not reported.
Description of main results	• Adolescence is not too late to intervene; interventions do benefit older struggling readers. The overall estimate of the effect size across all 31 studies was 0.95 (p<.001; 95% CI=.68, 1.22). In the 11 studies that used standardised, norm-referenced measures, the average effect was 0.42 (p=.002, 95% CI=.16, .68). • Older students with reading difficulties benefit from interventions focused at both the word and the text level. Fluency interventions yielded an overall ES of 0.26 (n=4, 95% CI=-.08, .61) and 0.04 (k=2, 95% CI=-.43, .50) for standardised measures; 0.26 all measures of reading comprehension; and, -.07 standardised measures of reading comprehension). Multicomponent interventions yielded an overall ES of 0.56 (k=6, 95% CI=.25, .95) and 0.41 (n=3, 95% CI=-.08, .61) for standardised measures; 0.80 all measures of reading comprehension; and, 0.59 standardised measures of reading comprehension. • Older students with reading difficulties benefit from improved knowledge of word meanings and concepts. Vocabulary interventions had the largest overall effect size, 1.62 (k=5, 95% CI=1.13, 2.10). No vocabulary interventions used standardised measures, and most of them assessed the extent to which students learned the meaning of the words that were taught, rather than directly assessing their impact on reading. • Word-study interventions are appropriate for older students struggling at the word level. Estimated ES for word study interventions was 0.60 (k=4, 95% CI=.25, .95) and 0.68 (k=3, 95% CI=.32, 1.03) for standardised measures; and, 0.40 measures of reading comprehension. • Teachers can provide interventions that are associated with positive effects. However, effect sizes were larger for researcher-implemented interventions (1.49 all measures; 1.08 standardised measures) than for teacher implemented interventions (0.63 all measures; 0.21 standardised).

Study	Scammacca N, Roberts G, Vaughn S, Edmonds M, Wexler J, Reutebuch CK, et al. 2007
Description of main results	• Teaching comprehension strategies to older students with reading difficulties is beneficial. Overall ES for comprehension strategies was 1.23 (k=12, 95% CI=.68, 1.79) and 0.55 (k=2, 95% CI=-.99, 2.09) for standardised measures. The estimate of effect size across 23 studies targeting reading comprehension was 0.97 (95% CI=.61, 1.33). The overall effect-size estimate for the 8 studies using standardised, norm-referenced measures of reading comprehension was 0.35 (95% CI=-.05, .75). Note that the confidence interval does not exclude the possibility of a "true" effect of 0. While less likely than the .35 point estimate, it is nonetheless possible. Older readers' average gains in reading comprehension are somewhat smaller than those in other reading and reading-related areas studied. Effect sizes were larger in studies where participants were middle-grade students (1.05 all measures; 0.56 standardised measures) as opposed to high-school students (0.78 all measures; 0.13 standardised measures). Older students with learning disabilities (LD) benefit from reading intervention when it is appropriately focused. Across all studies, those with only participants with learning disabilities had significantly higher effects than those with no participants with learning disabilities (ES=1.20 for all LD, 0.39 for no LD; p<.05). Gaps: we need studies that provide instruction over longer periods of time and assess outcomes with measures more like those schools use to monitor reading progress of all students.

Study	Scammacca NK, Roberts G, Vaughn S, Stuebing KK. 2015
Review objectives	The effectiveness of interventions for struggling readers in Grades 4 through 12 published between 1980 and 2011.
Descriptions of interventions/ phenomena of interest	1. The overall effectiveness of reading interventions (any type of reading instruction, including word study, fluency, vocabulary, reading comprehension, or multiple components of reading instruction) in English. 2. How the magnitude of the effect varies based on: a. student: i. LD status: All designated learning disabled; some designated learning disabled, some struggling; all struggling, none designated learning disabled, ii. Grade level: 4-5; 6-8; 9-12 b. type of intervention: vocabulary, word study, fluency, comprehension, or multiple components of reading c. research design characteristics: i. Implementer: teacher; researcher, ii. Hours of intervention: 0-5hr; 6-15hr; 16-25hr; 26+hr, iii. Design type: multiple treatment; treatment/comparison. 3. What differences in effectiveness exist between more recent interventions and older ones.

Study	Scammacca NK, Roberts G, Vaughn S, Stuebing KK. 2015
Descriptions of outcomes included in the review	Reading performance as measured by: standardised (norm-referenced) measures of reading; standardised reading comprehension measures.
Descriptions of contexts included in the review	1. Struggling readers (low achievement in reading, unidentified reading difficulties, dyslexia, and/or with reading or LD) in Grades 4 to 12. 2. Research design: Experimental or quasi-experimental treatment-comparison (comparison group received either no intervention or the school's 'business-as-usual' reading intervention) or multiple-treatment comparison (all groups received an intervention designed by researchers that they would not have received if they were not participants in the study) research design.
Search details	1. A computer search of ERIC and PsycINFO was conducted to locate studies published between 2005 and 2011 to add to the studies published between 1980 and 2004 that were included in Scammacca et al. (2007). 2. A search of abstracts from other published research syntheses and meta-analyses was undertaken. 3. Reference lists in seminal studies were reviewed to ensure that all relevant studies were identified. 4. A hand search through all articles published between 2005 through 2011 in 11 major journals was conducted: Annals of Dyslexia, Exceptional Children, Journal of Educational Psychology, Journal of Learning Disabilities, Journal of Special Education, Learning Disabilities Research and Practice, Learning Disability Quarterly, Reading Research Quarterly, Remedial and Special Education, and Scientific Studies of Reading. Date range for meta-analysis: 1980 - 2011 (that is, including those meta-analysed in Scammacca et al. 2007).
Number of studies and participants included	82 independent, study-level effect sizes from 67 published research reports were included in the meta-analyses conducted. Of these, 32 were published between 1980 and 2004 (the 1980–2004 group) and 50 were from studies published between 2005 and 2011 (the 2005–2011 group). 13,360 students from grades 4 to 12.
Appraisal instruments used	Publication bias was evaluated using the trim-and-fill approach.
Description of main results	• For all types of outcome measures: a. Overall, mean ES across the 82 studies was .49 ($p < .001$, 95% CI = 0.38, 0.60) – a moderate positive effect of intervention b. the mean ES for the 1980-2004 studies was .96 ($p < .001$, 95% CI = 0.69, 1.23) c. the mean ES for the 2005-2011 studies was .23 ($p < .001$, 95% CI = 0.15, 0.31) d. the difference in the means for the 2 groups of studies was statistically significant (Q-between = 25.81, df = 1, $p < .001$).

Study	Scammacca NK, Roberts G, Vaughn S, Stuebing KK. 2015
Description of main results	• For all standardised, norm-referenced outcome measures: a. the mean ES for all studies was .21 (p < .001, 95% CI = 0.12, 0.30), indicating a small positive effect of intervention b. for the 1980-2004 studies, the mean ES was .42 (p = .006, 95% CI = 0.25, 0.59) c. for the 2005-2100 studies, the mean ES was .13 (p < .001, 95% CI = 0.07, 0.18) d. the difference in means for the 2 groups of studies was statistically significant (Q-between = 4.83, df = 1, p < .028). • For all reading comprehension measures: a. the mean ES for all studies was .45 (p < .001, 95% CI = 0.34, 0.57) - a moderate positive effect of intervention on students' reading comprehension skills b. for the 1980–2004 studies, the mean ES was .91 (p < .001, 95% CI = 0.59, 1.24) - a large effect c. for the 2005-2011 studies, the mean ES was .24 (p < .001, 95% CI = 0.16, 0.33) - a small effect d. the difference in means for the 2 groups of studies was statistically significant (Q-between = 15.49, df = 1, p < .001). • For standardised reading comprehension outcome measures: a. the mean ES from standardised reading comprehension measures for all studies was .24 (p < .001, 95% CI = 0.14, 0.34) - a small positive effect of intervention on students' reading comprehension skills b. for the 1980-2004 studies, the mean ES from standardised reading comprehension measures was .65 (p = .03, 95% CI = 0.06, 1.19) - a moderate effect c. for the 2005-2011 studies, the mean ES from standardised reading comprehension measures was .19 (p = .03, 95% CI = 0.11, 0.27) d. The difference in the mean effect sizes for the 2 groups of studies was not statistically significant (Q-between = 2.24, df = 1, p = .13). • Year of publication was a statistically significant predictor of ES for all types of outcome measures (β = –.04, SE = 0.01, Q-model = 40.95, df = 1, p < .001, T2 = .15), and all measures of reading comprehension (β = –.04, SE = 0.01, Q-model = 18.47, df = 1, p < .001, T2 = .18).

Study	Scammacca NK, Roberts G, Vaughn S, Stuebing KK. 2015
Description of main results	MODERATOR VARIABLES (findings from all studies for all outcome measures). • Type of intervention: For all outcomes across all studies: Reading comprehension (ES=.74); Fluency (ES=.30); Word study (ES=.33); Vocabulary (ES=1.58); Multiple components (ES=.20): Vocabulary interventions had a significantly larger mean effect size than all other types of interventions (p<.05). NB. For reading comprehension measures, the ES for reading comprehension interventions (ES=.78) was significantly higher (p<.005) than fluency interventions (ES=.31). • LD status of participants: Mean ES for studies that included only students with LD (ES=.95) was significantly greater (p<.017) than the mean ES for studies that included some students with LD and some struggling readers (ES=.32), and studies with no LD students (ES=.24). • Duration of intervention: Shorter interventions had a significantly larger mean ES (p<.008): 5hr or less (ES=1.0) and 6 to 15hr (ES=.66) compared to +26hr (ES=.18). • Type of implementer. Researcher implemented interventions (ES=.68) had a significantly larger (p<.05) mean ES than teacher implemented interventions (ES=.35). • Grade level and design type were not significant. The most recent research (2005-2011) suggests that all struggling readers benefit from intervention regardless of their diagnosed LD status.

Study	Shin J. 2017
Review objectives	Providing a comprehensive examination of the research base to support the utility of CBM oral reading and maze as indicators of reading comprehension as measured by state achievement tests.
Descriptions of interventions/ phenomena of interest	CBM oral reading and maze as indicators of reading comprehension.
Descriptions of outcomes included in the review	1. The estimated average correlations between CBM tasks (oral reading and maze) and reading comprehension on state achievement tests. 2. Potential moderating factors (grade level, participant characteristics, characteristics of CBM and criterion measures, and time interval between the administration of CBM and state test).
Descriptions of contexts included in the review	Study conducted in the US Grade range of included studies: Primary (Grades 1–3), intermediate (Grade 4–6), secondary (middle- and high-school).
Search details	Three databases: ERIC PsycINFO, and Google Scholar Ancestral plus hand search. Commercial CBM sites plus research institutes. No date range specified. 45% studies published after 2010.
Number of studies and participants included	61 studies of correlation (132 correlations). 80% of the 132 correlations were from studies conducted with elementary students. Mean proportions of ELLs and students receiving special education were 13% and 14%, respectively. The mean proportion of students receiving free or reduced lunch was 49.5%.

Study	Shin J. 2017
Appraisal instruments used	No formal critical appraisal was undertaken – no instrument named. Bias was accounted for by: use of random effects model. A 22-level meta-analysis was employed to yield the overall mean correlation coefficient and to examine whether variances between studies were statistically significant. Calculated heterogeneity coefficient I2.
Description of main results	• CBM reading tasks (oral reading and maze) function as valid indicators of reading comprehension on state tests across grades. • Oral reading may be more predictive of reading comprehension than maze for students after grade. • Results of the present study strengthen the empirical evidence that oral reading fluency is a strong indicator of reading comprehension across grades. Furthermore, findings of this meta-analysis suggest that oral reading might be a slightly better indicator of reading comprehension than maze even for intermediate or secondary grade levels, which contradicts previous findings. • Correlations from state-developed tests (for example, Iowa Basic Skills Test, Stanford Achievement Test)) were significantly higher than those from commercially prepared tests of reading comprehension. • Researcher-developed CBM and commercial CBM passages (for example, DIBELS) produce similar predictive validity coefficients. • Correlations did not significantly differ by the number of CBM passages used. This result suggests that the data used to comprise the final score (such as, a single score from one passage, a mean score from 2, or a mean or median score from 3 passages) did not influence the relations between CBM and reading comprehension on state tests. • Type of administrator of CBM (researchers or school personnel) did not influence the relations between CBM and reading comprehension on state tests. Correlations for both type of administrators were very similar, meaning that the validity of CBM (oral reading and maze) as an indicator of reading comprehension on state tests would be similar regardless of the type of administrator. These results support the argument that CBM reading tasks can be used flexibly across different conditions, such as materials and administrators. • Correlations did not significantly differ by response format. (open ended questions versus multiple choice). • This result suggests that the 2 CBM reading tasks may be less predictive of reading comprehension on state.

Study	Shin M, Bryant DP, Powell SR, Jung P-G, Ok MW, Hou F. 2021
Review objectives	To investigate the overall effect of WP instruction and the moderating effects of study quality based on the Council for Exceptional Children (CEC;2014) QIs and Common Core State Standards for Mathematics (CCSSM) content and practice standards (NGA Center and CCSSO, 2010) for students with learning disabilities (LDs) in Grades 1 through 12.
Descriptions of interventions/ phenomena of interest	The effectiveness of WP-solving instruction (independent variable), targeting the teaching of WP solving and moderating variables, especially study quality indicators.
Descriptions of outcomes included in the review	Students with LDs who were identified by school districts as having LDs in Grades 3 through 12. Total = 82 students, with an average of 4 students per study (min = 3, max = 10, median = 3, SD = 1.83), of whom 55 were males and 27 were females. All participants had difficulty in mathematical problem solving; 16 students were reported to have additional difficulty in reading or writing.
Descriptions of contexts included in the review	Not reported.
Search details	Followed PRISMA guidelines. 3 databases: Education Resources Information Center (ERIC; n = 1,181), PsycINFO (n = 1,982), and Academic Search Complete (n = 1,899) Ancestral search.
Number of studies and participants included	20 studies SCD either multiple-baseline or multiple-probe across subjects, ABk (k = number of phase contrasts) withdrawal designs targeting, or alternating treatment designs with baseline Minimum of 3 participants in each study, and a minimum of 3 data points per baseline and intervention phase. 1975 to 2020.
Appraisal instruments used	CEC QIs (CCSSM) content and practice standards.

Study	Shin M, Bryant DP, Powell SR, Jung P-G, Ok MW, Hou F. 2021
Description of main results	• There were strong overall effects of WP instruction for students with LDs who experience difficulty solving WPs. Specifically, in all 20 studies, students with LDs displayed significant improvement from baseline to intervention phases, yielding a significantly large overall weighted effect size estimate (BC-SMD = 4.52). • In general, the weighted average effect size estimate of WP instruction for students with LDs was large and significant for all eight CEC QI categories. Meeting all the 3 fidelity QIs (5.1. adherence, 5.2. dosage, and 5.3. regularity) resulted in the greatest effect (BC-SMD = 5.54). • The average weighted effect sizes of WP instruction for students with LDs were large across different CCSSM content standards and, in particular, were greater in the domains of Geometry (BC-SMD = 11.80) and Expressions and Equations (BC-SMD = 6.21) than in Operations and Algebraic Thinking (BC-SMD = 2.98). • WP-solving instruction for students with LDs was largely effective across CCSSM practice standards in general (BC-SMD = 3.98–6.21). The effect sizes were particularly large to: Look for and make use of the structure (Mathematical Practice 7, BC-SMD = 6.21). In this study, the greater the number of MPs (greater than 3) provided, the larger the effect size. Gaps: a limited number of studies focused on the advanced level of WPs within the domains of G (n = 2), Number System (n = 2), and Ratios Proportions (n =2). Considering that many secondary students with LDs consistently experience difficulty with topics such as ratios, proportions, and percentages at the secondary school level (Soares et al. 2018), more studies should fill this gap in our knowledge in this field.

Study	Shin M, Bryant DP. 2015
Review objectives	To identify intervention studies that focused on fraction interventions for students with mathematics learning difficulties.
Descriptions of interventions/ phenomena of interest	The features (for example, participants) and instructional components (for example, explicit, systematic instruction) of targeted intervention studies to determine the nature of fraction instruction for students who struggle with learning mathematics. The effect of the interventions on performance in fractions. The extent of research on fractions as it relates to the Common Core State Standards for Mathematics (CCSSM).
Descriptions of outcomes included in the review	Researcher developed measures (n = 16 studies); both standardised and researcher-developed measures (n = 4 studies).
Descriptions of contexts included in the review	US schools; students in Grades 3 - 12. Whole-group, small-group, and independent practice.
Search details	3 databases: ERIC, JSTOR, PsychInfo, and ProQuest Dissertations and Theses Full Text. Hand search. Ancestral search. Date range: 1975-2014.

Study	Shin M, Bryant DP. 2015
Number of studies and participants included	Total = 17 studies. RCT, quasi-experimental or single subject designs. Total number of students = 805. Of these students, 153 were low achieving students, 301 were students with LD, 321 had other disabilities, and 30 were typically achieving students. Specifically, of the 301 students with LD, 175 had mathematics goals on their Individualised Education Programs (IEPs) indicating that they needed special education assistance in mathematics. 2 studies included elementary school students (89 third graders). The remaining 15 studies included middle or high-school students: 5 students in Grade 5, 32 in Grade 6, 8 in Grade 7, 26 in Grade 8, 36 in Grade 9, 477 in Grades 6 to 8, 56 in Grades 9 to 11, and 76 in Grades 9 to 12.
Appraisal instruments used	Not reported.
Description of main results	• Intensifying instruction by increasing time (duration of the study) can be beneficial for struggling students to better learn fraction concepts. Bottge et al. (2014) devoted approximately 4,230 to 5,640 min to instruction. Not surprising, Bottge et al.'s (2014) study showed higher positive outcomes for students in Grades 6 - 8 than the previous study by Bottge et al. (2010), in which 1,200 to 1,320 min were devoted to instructional time. • Many studies did not focus on multiplication and division of fractions; none of the studies included negative fractions; and none addressed how to represent fractions on a number line by linking conceptual and procedural knowledge (NMAP, 2008; Siegler et al. 2010). The lack of number line use in the included studies is problematic in terms of helping students make conceptual connections. • Reform-based mathematics instruction, which highlights whole-class discussion, may be challenging and impose a high degree of cognitive load for struggling learners. Thus, mathematics instruction for struggling learners should be explicit coupled with using representations and student dialogue even for developing conceptual understanding and student led mathematical reasoning activity in class. • The features of explicit, systematic instruction – including explicit modelling of step-by-step strategies, sufficient practice, and cumulative reviews with feedback – enhance instruction in combination with visuals. Ten of the 17 studies (for example, Bottge et al. 2014; Bottge et al. 2010; Hunt, 2014; Watt, 2013) that used concrete and visual representations in combination with explicit, systematic instruction showed highly positive outcomes on fraction concepts and skills.

Study	Shin M, Bryant DP. 2015
Description of main results	• Five studies implemented heuristic strategies by promoting mathematical verbalization with a 'think-aloud' strategy or in combination with virtual manipulatives, leading to improved problem-solving and fraction computation skills. Other studies supported using a heuristic strategy (Gersten et al. 2009), noting a significant effect for students with LD in emphasizing student verbalizations and the reflective process of their mathematical problem solving. • Using real-world problems, connecting fraction skills to contextualised problems via video-based anchored instruction (AI) or enhanced anchored instruction (EAI), resulted in improvements on contextualised tests for students struggling to learn mathematics but not on fraction computation and problem-solving tests. However, when EAI was provided in combination with explicit, computer-based instruction, students with disabilities outperformed their control group on fraction computation and problem solving (Bottge et al. 2014; Bottge et al. 2010). These findings highlight the significant role of explicit instruction in delivering mathematics instruction to struggling learners.

Study	Solis M, Ciullo S, Vaughn S, Pyle N, Hassaram B, Leroux A. 2012
Review objectives	To summarise the findings of reading comprehension intervention studies specifically designed for middle-school students (Grades 6–8) with LD and assist in identifying both strengths and weaknesses within this body of research.
Descriptions of interventions/ phenomena of interest	The effectiveness of reading comprehension interventions based on experimental, quasi-experimental, and single-participant research studies in improving reading comprehension outcomes for middle school students with LD.
Descriptions of outcomes included in the review	Measures of reading comprehension. The majority of outcome measures were researcher developed (n = 17), whereas 4 studies used standardised measures of reading comprehension.
Descriptions of contexts included in the review	The language of instruction was English. Researchers implemented treatments in the majority of studies (n = 11). No other contextual details were reported.
Search details	Electronic searches of the ERIC and PsycINFO databases. References from previously published syntheses were reviewed. A hand search of 4 major journals. Date range: 1979 and 2009.
Number of studies and participants included	Total studies= 12; 9 experimental, 3 quasi-experimental, and 2 single participant studies. Students were in Grades 6 through 8 (ages 12–14); identified with an LD. Total students = 491; 410 students identified as LD based on state or federal guidelines as the criteria (k = 2), a "discrepancy" model, comparing reading achievement to IQ or ability (k = 8) or criteria unreported (k = 4).
Appraisal instruments used	Coding included elements specified in the What Works Clearinghouse Design and Implementation Assessment Device (Institute of Education Sciences, 2003).

Study	Solis M, Ciullo S, Vaughn S, Pyle N, Hassaram B, Leroux A. 2012
Description of main results	• Outcomes for middle school students with learning disabilities (LD) following reading comprehension interventions were largely characterised by medium to large effect sizes derived primarily from researcher-developed comprehension measures. Few studies (n = 4) reported standardised measures of reading comprehension, which yielded on average lower effect sizes (ESs = -1.42, 0.40, 0.33, 0.97) than researcher-developed measures (ES range = -0.35 to 6.66). • The findings from this synthesis support using summarisation or main idea strategy instruction to improve understanding of text. The studies reviewed indicated several strategies to accomplish summarization including a sequential process, self-questioning, mnemonics, and graphic organisers. Further support may be provided to students through self-monitoring tools such as a checklist or a prompt card. Gaps: Motivation and engagement, cooperative grouping, use of discourse, and reader response are all recommended practices from research syntheses with older students that have been inadequately validated for students with LD.

Study	Steinle PKS, Stevens E, Vaughn S
Review objectives	The purpose of this synthesis is to extend the findings of Wexler et al. (2008), adding to the knowledge base by summarizing the results of reading fluency interventions for struggling readers in Grades 6 through 12 published since 2006.
Descriptions of interventions/ phenomena of interest	The impact of reading fluency interventions conducted in English on reading fluency and comprehension outcomes for struggling readers in secondary schools.
Descriptions of outcomes included in the review	Reading fluency (for example, WCPM and decreased errors per minute (EPM) and comprehension. NOTE: of the 11 studies included in this synthesis, only 5 used standardised measures of reading comprehension. The remaining studies used proximal, unstandardised measures of reading comprehension typically consisted of literal or recall comprehension questions from that the passage.
Descriptions of contexts included in the review	Secondary schools, interventions run for struggling students including those with reading difficulties, learning disabilities, developmental disabilities and emotional/behavioural disorders.
Search details	Four databases: Educational Resources Information Clearinghouse, PsycINFO, Education Source, and Academic Search Complete. Included studies ranged from 2006–2020.
Number of studies and participants included	Seventeen studies met the inclusion criteria. Four studies used a group design. The remaining 13 studies used a single-case design. There were 337 participants in grades 6 - 12, 27% of which were in the early years of secondary school (below year 9).
Appraisal instruments used	WWC design standards for single-case research were used. Six SCD studies were appraised as 'does not meet design standards' and were deemed ineligible for further review of evidence of an effect. No critical appraisal was reported for the quality of the group design (treatment-comparison) studies.

Study	Steinle PKS, Stevens E, Vaughn S
Description of main results	Impact on fluency: • in the current review, there was conflicting support for Repeated Reading to improve reading fluency as there was inconsistency between the results of single-case designs and group designs with regard to positive effects on fluency outcomes • group-designed studies showed no effect while single-case designed studies showed a positive impact on fluency. Impact on comprehension: • repeated Reading interventions did not result in improved comprehension for students.

Study	Stevens EA, Park S, Vaughn S. 2019
Review objectives	To examine the effects of summarising and main idea interventions on the reading comprehension outcomes of struggling readers in Grades 3 to 12.
Descriptions of interventions/ phenomena of interest	The effects of summarizing and main idea interventions on the reading comprehension outcomes of struggling readers in Grades 3 through 12. Instructional practices that exist in the literature to address summarizing and main idea. Type of text, narrative, or expository is used in summarisation and main idea instruction.
Descriptions of outcomes included in the review	Reading comprehension outcomes.
Descriptions of contexts included in the review	Instruction, in English provided as part of the school programming.
Search details	Three electronic databases: Education Source, Educational Resources Information Clearinghouse (ERIC), and PsycINFO. Date range: 1978–2016.
Number of studies and participants included	Total studies = 30 studies - 24 group design studies (such as, 20 treatment–control experiments, 3 multiple treatment experiments, one quasi-experiment) and 6 SCD studies. Total participants = 983 (range: 2 - 81 per study) struggling readers in Grades 3 through 12. Struggling readers were defined as: 1. students identified with learning disabilities 2. students identified with reading disabilities 3. students with reading difficulty or at-risk status as determined by low performance on a reading measure (for example, Comprehensive Test of Basic Skills), placement in a remedial reading class, or district or school identification (for example, the principal and teachers selected students struggling with reading comprehension).
Appraisal instruments used	Coding procedures were based on study features described in the What Works Clearinghouse (WWC) Design and Implementation Assessment Device (Valentine and Cooper, 2008).

Study	Stevens EA, Park S, Vaughn S. 2019
Description of main results	• Students' performance on more proximal measures of main idea and summarisation was consistently better than on more generalised measures of reading comprehension for struggling readers in Grades 3 to 12. The mean estimate (ES = 0.97) reflected primarily unstandardised measures. Of the 103 ES estimates reported in 22 experimental studies, only 77 effects resulted from standardised measures used in 3 experimental studies and yielded primarily negative, small effects and positive, small effects. • Students in the upper elementary, middle, and high school grades benefit from summarising and main idea practices. While not statistically different, the mean estimate for middle- and high-school students resulted in a large effect of 1.12 compared to a moderate effect for elementary students of 0.68. • Struggling readers benefit from main idea and summarising interventions regardless of the group size (one to 4 vs. 5 or more) or number of sessions (12 or fewer vs. 13 or more). Group size and number of sessions were not statistically significant predictors of ES. • Paraphrasing with self-monitoring may enhance struggling readers' comprehension across the grade levels. Review of the SCD studies, resulted in strong evidence in favour of main idea and summarizing interventions on measures of oral and written retell. However, results were less conclusive on short-answer comprehension measures (such as, WWC evidence standard ratings ranged from no evidence to strong evidence). Most SCD studies used unstandardised measures (such as, research-developed open-ended, short-answer comprehension questions). Gaps: further research of high quality and rigor (such assuch as, fidelity of implementation, use of standardised measures, nature of the counterfactual's instruction) is needed to inform future practice and research by identifying high-impact instructional practices.

Study	Stevens EA, Rodgers MA, Powell SR. 2018
Review objectives	To conduct a meta-analysis of 25 years of mathematics interventions for students with mathematics difficulty (educator determined or mathematics scores below 40th percentile) or disability (school identified) in Grades 4 to 12.
Descriptions of interventions/ phenomena of interest	The effects of mathematics interventions on mathematics outcomes for students with MD in Grades 4 through 12. Potential moderating factors on intervention outcomes, specifically (a) treatment duration or dosage, (b) mathematics content, (c) study characteristics, and (d) study quality.
Descriptions of outcomes included in the review	Researcher-developed measures; standardised measures; NAEP = National Assessment of Educational Progress; WRAT4 = Wide Range Achievement Test.
Descriptions of contexts included in the review	Instruction provided at school in small group or individual settings by educators.
Search details	Four databases: Academic Search Complete, Education Source, Educational Resources Information Clearinghouse (ERIC), and PsycINFO. Ancestral search. Hand search. No grey literature. Date range: January 1990 and December 2015.

Study	Stevens EA, Rodgers MA, Powell SR. 2018
Number of studies and participants included	Total experiments = 25. Total number of students = 1948; range: 6 - 259.
Appraisal instruments used	Council of Exceptional Children's quality standards for classifying evidence-based practices in special education (Cook et al. 2015)
Description of main results	• The estimated mean effect of interventions was moderate and not statistically significant. Even though the Precision Estimate Test (PET) estimate was not significant, the ES suggests practical significance of half a standard deviation. The PET estimate of 0.49 was lower than the mean effects found in prior reviews for interventions targeting primarily elementary students. This suggests that remediating MD for older students may be more challenging than remediation for younger students. • Older students may require interventions of greater intensity and duration to make substantial improvements in mathematics performance. The moderator analysis supported this conclusion, as student performance increased significantly (p = .013) with interventions providing more than a total of 15hr of intervention. • Fractions interventions significantly improved students' mathematics outcomes more than interventions in operations (p = .04). Only one study investigated the effects of generalised mathematics skills (Ketterlin-Geller, Chard, and Fien, 2008); no studies singularly targeted geometry, measurement and data, statistics and probability, or algebraic expressions and equations, all of which are necessary skills to succeed in middle and high school courses. Gaps: It remains unknown if students respond more favourably to a multicomponent mathematics intervention, targeting multiple, yet related domains, or an intervention that builds in complexity to address gaps in students' prior understanding. Overall, there is a need for high-quality mathematics intervention research that employs rigorous designs, larger sample sizes, and using rigorous, standardised mathematics assessments with better reliability and validity.

Study	Stewart AA, Austin CR. 2020
Review objectives	To investigate reading studies for participants identified with or at risk of ADHD.
Descriptions of interventions/ phenomena of interest	The effects of reading interventions on reading outcomes of students with or at risk of ADHD in Grades 4 to 12. Reading interventions that use an alphabetic language that focused on reading fluency, vocabulary, or reading comprehension.
Descriptions of outcomes included in the review	Measures of reading outcomes in fluency, vocabulary, or comprehension.
Descriptions of contexts included in the review	Separate classroom (k = 15); general education setting with special education services provided via push-in supports (k = 1).

Study	Stewart AA, Austin CR. 2020
Search details	Three databases: (a) EducationSource, (b) ERIC, and (c) PsychINFO. Peer-reviewed studies published before June 2017; no grey literature. Hand search to examine studies published from January 2015 to August 2017.
Number of studies and participants included	Total studies = 16. 414 single-case studies and 22 group design studies. Total students = 65 (55 males and 10 females). All participants were identified as having a medical diagnosis of ADHD. All participants received special education services and many were identified as having co-occurring diagnoses such as learning disabilities (LDs), emotional and behavioral disorders EBD), speech and language impairments (SLI), or autism spectrum disorders (ASDs). Eight studies included students in upper-elementary grades (Grades 4 and 5), and eight studies included students in middle and high school (Grades 6–12).
Appraisal instruments used	Council for Exceptional Children (CEC) QIs described by Cook and colleagues (2015)
Description of main results	• Only one intervention (direct instruction) included in the current review met the criterion necessary to be classified as a potential evidence-based practice (EBP) according to CEC criteria. • Two studies utilised DI to target students' ability to comprehend text (Flores and Ganz, 2007, 2009). In both studies, researchers taught DI using Corrective Reading Thinking Basics: Comprehension Level A (Engelmann, Haddox, Hanner, and Osborn, 2002). The program included scripted lessons, opportunities for students to respond chorally, using signals to stimulate student response, corrective feedback procedures, explicit modelling of skills, guided practice, and independent practice. Researcher-developed and standardised measures were administered to investigate effects of the intervention. Tau-U calculations for both studies ranged from 0.75 to 1.04. Gaps: future studies utilizing interventions included in this review (for example, computer-based instruction, SRSD + TWA + PLANS, GO FASTER) with rigorous study design and a minimum of 3 students with or at risk of ADHD (if using a single-case design) are necessary to document sufficient evidence of EBPs in reading for this population. None of the included studies investigated the intervention effects on specific ADHD presentations (such as, hyperactivity, inattentive, combined). Although 3 studies noted differences in presentation among participants, conclusions were not connected to specific presentations when reporting results. Future research is needed to determine which interventions may be more effective for specific ADHD presentations, particularly those that are strongly associated with lower academic outcomes: inattentive and combined presentations.

Study	Swanson E, Hairrell A, Kent S, Ciullo S, Wanzek JA, Vaughn S. 2014
Review objectives	To analyse reading interventions designed for students with learning disabilities in kindergarten through Grade 12 that are implemented using social studies text.
Descriptions of interventions/ phenomena of interest	Reading interventions implemented within the context of social studies have employed using graphic organisers, mnemonics, reading and answering questions, guided notes, and multicomponent comprehension instruction. The features of research reflected in the literature – the outcomes result from reading interventions delivered to students with learning disabilities using social studies content – the reading intervention types most effective for students with learning disabilities when delivered within the social studies content area. Studies that employed treatment-comparison, multiple treatment, single-case, and/or single-group designs. Intervention included reading instruction within the context of social studies text or social studies content acquisition. Interventions that were part of typical school programming graphic organisers, mnemonics, read and answer questions, guided notes, multicomponent comprehension instruction.
Descriptions of outcomes included in the review	At least one outcome measure assessed social studies content learning or reading comprehension related to social studies concepts.
Descriptions of contexts included in the review	Typical school programming high school; junior high: upper elementary through middle school (3rd through 8th grade), high school (9th through 12th grade), both middle and high school. The context of social studies.
Search details	2 databases: ERIC and PsycINFO Ancestral search, hand-search.
Number of studies and participants included	A total of 24 articles, comprising 27 studies, with 16 studies providing sufficient data for inclusion in the meta-analysis. A total of 19 studies reported sample sizes ranging from 4 to 103 students, with a mean of 36.9, a median of 29.0, and a mode of 17.0. Student grade level was reported in 25 studies and ranged from 3rd to 12th grade. Also, 13 studies investigated interventions in upper elementary through middle school (3rd through 8th grade), and an additional 8 studies investigated interventions in high school (9th through 12th grade) 4 studies included students in both middle and high school.
Appraisal instruments used	An extensive coding document was created to organise pertinent information from each study. The code sheet addressed elements specified in the What Works Clearinghouse Design and Implementation Assessment Device (What Works Clearinghouse, 2008).

Study	Swanson E, Hairrell A, Kent S, Ciullo S, Wanzek JA, Vaughn S. 2014
Description of main results	• Social studies content may be a valuable text type for improving comprehension and knowledge of students with learning disabilities. Reading interventions utilizing social studies content are effective for students with LD; results indicate that for older students (Grades 7–12), the effect size may be greater. • Social studies reading interventions using social studies content have a large overall positive effect on content and comprehension outcomes among students with learning disabilities. Overall effect sizes were high (1.00 for mnemonics vs. other intervention types and 1.06 for graphic organisers vs. other intervention types); however, none of the intervention types appear to be significantly more effective than the others. • Intervention type per se is not as important as applying an organised, structured, text-based intervention to assist students with comprehending and learning social studies content. • Intensity of intervention: Interventions lasting less than 60 minutes total were associated with a mean effect size of 1.20 and interventions that lasted more than 500 minutes total were associated with a mean effect size of 1.07. • Duration of intervention: different amounts of treatment time resulted in similar outcomes, the duration of the intervention is closely tied to the intervention type 22 studies reported the number of intervention sessions, with a range of 1 to 25 and a mean of 8.5 sessions. In addition, 21 studies reported the length of each session, with a range of 8.5 to 90 minutes and a mean of 39.6 minutes. Among experimental studies, the number of intervention sessions averaged 9.4, whereas the length of sessions averaged 34.75 minutes. • Larger groups (groups with 11 or more students) were associated with a larger effect size (ES = 1.16) than smaller groups (ES = 0.87), therefore unable to confirm the most effective group size for using social studies intervention for students with learning disabilities, though for younger students' one-on-one interventions and those with very small groups are associated with higher reading outcomes than are interventions utilizing moderate to large group sizes. Gaps: high-quality research study standards, including using random assignment and standardised measures and the reporting of fidelity of treatment procedures - Studies utilizing standardised measures that are reliable and valid are needed to evaluate the efficacy of reading interventions in social studies for students with learning disabilities. Valid and reliable measures of social studies content learning. Studies require adequate information to calculate standard mean effect sizes and their error values. Studies on the effectiveness of these intervention types would be valuable – two-thirds of the studies were in Grades 7 to 12. Additional studies in the lower grades (3–6) are needed as well. - high-quality studies that include standardised outcome measures.

Study	Valasa LL, Mason LH, Hughes C. 2014
Review objectives	To synthesise the research on essay-writing interventions for adolescents with high incidence disabilities and to discuss the extent to which the research base is reflective of high CCSS (2013) initiatives.
Descriptions of interventions/ phenomena of interest	Interventions for improving narrative, persuasive, and expository essays of students in sixth through 12th grade with high-incidence disabilities.
Descriptions of outcomes included in the review	To determine outcomes of essay-writing interventions for adolescents with high-incidence disabilities across classroom setting, participant characteristics, and narrative, persuasive, and expository writing genre; to determine the quality of research for each writing genre; and, to determine how well the research base within each writing genre addresses CCSS (2013).
Descriptions of contexts included in the review	Middle or high school (such as, secondary) school settings in the US. Twenty-seven percent (n=7) of studies took place in an inclusion classroom, 23% (n=6) in the resource room, 8% (n=2) in an alternative program for students with emotional/behavioral difficulties, and 42% (n=11) administered treatment individually or outside of the regular daily setting. In the 23 studies that documented school district classification, 57% (n=13) of schools were considered suburban school districts, 17% (n=4) urban, 17% (n=4) rural, and 9% (n=2) urban/suburban.
Search details	Three academic databases were searched: ERIC, ProQuest Education Journals, and PsycINFO. An ancestral search was completed. No date range of the search was specified.
Number of studies and participants included	417 students in sixth to 12th grade in secondary school. Eighty-five percent (n=354) of students were diagnosed with LD, 12% (n=51) with EBD, 2% (n=9) with ADHD, less than 1% (n=1) with ADHD/SLI, less than 1% (n=1) with ADHD/DD, and less than 1% (n=1) with SLI. In the 21 studies that specified participant gender, 74% (n=226) were male and 26 % (n=80) were female.
Appraisal instruments used	Nine indicators adapted from Gersten et al. (2005) by Graham and Perin (2007a), each holding a value of 1 point, were used to evaluate group studies. Additionally, if no confound for the Hawthorne effect was documented, the Hawthorne effect was assumed controlled. Quality indicators for single subject studies (Horner et al. 2005) were adapted by Mason and Graham (2008) and were reported for groups of studies based on persuasive, narrative, and expository genre.

Study	Valasa LL, Mason LH, Hughes C. 2014
Description of main results	SRSD is a commonly used instructional approach in persuasive essay writing that increases the number of essay strategy specific elements in writing, as well as essay quality. • Goal setting is an effective means of improving the quality of persuasive writing, whether general goal setting or elaborated. • Strategy instruction may greatly improve the clarity and cogency of persuasive essay writing though the quality of evidence is limited. • SRSD has a large effect on essay strategy specific elements and essay quality in narrative essay writing that can be somewhat maintained over time. • Strategy instruction has a large effect on expository essay writing clarity, aptness, and organisation. • SRSD in expository essay writing increase essay strategy specific elements, quality, and expository composition. • SIM has a medium effect on quality of expository essay writing and a large effect for number of strategy specific elements. • Concept mapping (hand-mapping or computer mapping) can improve the quality of expository essays. • In order to ensure results are applicable to students in urban and rural districts, studies should be replicated across urban and rural settings, as students in these settings could possess different characteristics/cultural differences that may impact outcomes. • To address middle and high school CCSS standards, future persuasive writing research should incorporate instruction in writing persuasive essays across content areas using accurate data and credible sources. • More research is needed to conclusively recommend SRSD and strategy instruction without explicit instruction in self-regulation as best-practice narrative essay intervention techniques for adolescents with disabilities. • Future intervention studies should specify expository text-structure prompt type, document whether or not the intervention can be generalised across text-structures, and should include instruction for including discipline specific content and vocabulary.

Study	Wanzek J, Vaughn S, Scammacca NK, Metz K, Murray CS, Roberts G, et al. 2013
Review objectives	To determine the effectiveness of extensive interventions in improving reading outcomes for struggling readers (those with reading difficulties or disabilities) in Grades 4 through 12. To determine features of extensive interventions (for example, group size, duration, grade level) that are associated with improved outcomes for students.
Descriptions of interventions/ phenomena of interest	Extensive interventions (at least 75 sessions) for struggling readers in grades 4-12 Interventions targeted reading in an alphabetic language. Features of interventions: group size, duration of interventions, grade level.

Study	Wanzek J, Vaughn S, Scammacca NK, Metz K, Murray CS, Roberts G, et al. 2013
Descriptions of outcomes included in the review	Reading outcomes. Types: reading comprehension, reading fluency, word reading, word reading fluency, and spelling. a. Reading comprehension included measures that required students to read text and answer questions or complete sentences in the text. b. Reading fluency included measures that assessed students' rate and accuracy in reading text. c. Word reading included measures of students' word reading accuracy, with words or nonwords presented in isolation (not text). d. Word reading fluency included word reading measures with a rate component. e. Spelling included measures that assessed students' ability to encode.
Descriptions of contexts included in the review	Interventions were provided as part of the school day programming (not home, clinic, or camp programs) Interventions were provided for 75 or more sessions, and were not part of the general education curriculum provided to all students. Research design was experimental, quasi-experimental, single group, or single case.
Search details	1. A computer search of ERIC and PsycINFO to locate studies published between 1995 and 2011. 2. A hand search of eight major journals commonly reporting reading intervention research for students with reading difficulties or disabilities (Exceptional Children, Journal of Educational Psychology, Journal of Learning Disabilities, Journal of Special Education, Learning Disabilities Research and Practice, Reading Research Quarterly, School Psychology Review, Scientific Studies of Reading): 2010 and 2011. 3. Search for studies conducted by IES and posted on its website: www.ies.ed.gov.
Number of studies and participants included	19 studies: a. 11 treatment and comparison, 1 multiple-treatment, 1 single-subject, and 6 single-group design b. 10 studies included in metanalysis. For the other 9 studies that did not provide sufficient data for inclusion in the meta analysis, a synthesis of findings is provided. Implementer of intervention: c. school staff members (12 studies – teachers in 8, teachers and paraprofessionals in 3, and paraprofessionals in 1) d. trained research staff members in 6; 3. 1 study - no information 9,371 students 9 of the 19 samples included only students with identified disabilities, largely students with LD. The majority of the studies were conducted in Grades 6 to 8, with 10 studies including one or more of these grades. 6 studies included students below Grade 6, and 3 studies included students in ninth grade. No studies examined extensive interventions in Grades 10 through 12.

Study	Wanzek J, Vaughn S, Scammacca NK, Metz K, Murray CS, Roberts G, et al. 2013
Appraisal instruments used	Code sheet was developed based on elements specified in the What Works Clearinghouse Design and Implementation Assessment Device. Publication bias was evaluated by using the trim-and-fill approach.
Description of main results	• Overall, the findings for students in Grades 4 through 12 indicated a small effect for extensive interventions on reading comprehension, reading fluency, word reading, word reading fluency, and spelling outcomes. For studies included in the meta-analysis: • reading comprehension outcomes. The estimate of the mean effect size across the 22 reading comprehension effects included in the analysis was 0.10 ($p < .001$; 95% confidence interval [CI] [0.06, 0.19]), indicating a small positive effect of intervention on students' reading comprehension. No statistically significant differences were found between groups based on any moderator variable (group size, relative number of hours of intervention, or grade level of intervention) • reading fluency outcomes. The mean effect size estimate for the 9 effect sizes from fluency outcome measures was 0.16 ($p = .004$; 95% CI [0.05, 0.26]), indicating a small positive effect of intervention on students' reading fluency ability • word reading outcomes. The 12 effect sizes from word reading outcome measures had a mean effect size estimate of 0.15 ($p = 0.003$; 95% CI [0.05, 0.24]), indicating a small positive effect of intervention on students' word reading outcome scores • word reading fluency outcomes. Eleven effect sizes were analysed from word reading fluency outcome measures. The mean effect size estimate was 0.16 ($p = 0.001$; 95% CI [0.06, 0.26]), indicating a small positive effect of intervention on students' word reading fluency ability • spelling outcomes. 5 effect sizes were available from spelling outcome measures. Their mean effect size estimate was 0.15 ($p = 0.014$; 95% CI [0.03, 0.27]), indicating that the interventions had a small positive effect on students' spelling ability. Similar findings for the 9 studies not included in the meta-analysis (described qualitatively).

Study	Watt SJ, Watkins JR, Abbitt J. 2016
Review objectives	To identify effective interventions for teaching algebra to students with LD. Also to evaluate the complexity and alignment of skills with the Common Core State Standards in mathematics.
Descriptions of interventions/ phenomena of interest	5 interventions that were the primary focus of the study: (a) CRA, (b) cognitive strategy instruction, (c) EAI, (d) tutoring, and (e) graphic organisers.
Descriptions of outcomes included in the review	Mainly researcher-developed (for example, problem-solving tasks).
Descriptions of contexts included in the review	Grades 3 to 12; about 55% were participants in grades 7 to 9. Only 5 studies in general education settings.

Study	Watt SJ, Watkins JR, Abbitt J. 2016
Search details	1. Database search of Academic Search Complete, PsycINFO, and EBSCOhost from 1980 to 2014. 2. Manual search of journals likely to publish studies involving math interventions for algebra students. 3. Ancestral search of the located studies yielding no new articles.
Number of studies and participants included	15 studies; 10 quantitative, 5 single-subject designs 827 participants of which 451 were in grades 7-9. 224 students were LD and 10 had behavioural/ emotional disorders.
Appraisal instruments used	Inter-rater reliabilities of ES calculations.
Description of main results	• Moderate overall effect size (g = 0.48). • Most common intervention was the CRA instructional sequence. High effects (g = 0.53, tau-U = 1.00) on student achievement in algebra. • ES for studies using EAI (g = 0.80) and ES for cognitive strategies (g = 0.83, tau-U = 1.00) – highly effective practices for teaching algebra content to students with LD. • Studies using the CRA instructional sequence also produced large effects (g = 0.53, tau-U =1.00) – 7 studies indicated increased student performance in mathematics. CRA approach to replicated across the largest range of instructional content: from solving equations with unknown variables to multiplying linear expressions. 2/7 studies using CRA sequence used a cognitive strategy to support the retention and generalization of student learning. • Studies investigating using graphic organisers (g = 0.57) and individualised tutoring (g = 0.40) produced moderate to large effects on student learning in algebra (only 2 studies in each instructional category).

Study	Wexler J, Vaughn S, Edmonds M, Reutebuch CK. 2008
Review objectives	The efficacy of fluency interventions in enhancing the fluency and comprehension outcomes of students with reading difficulties in grades 6 through 12.
Descriptions of interventions/ phenomena of interest	1. Any type of fluency intervention with a comprehension and/or fluency outcome. 2. Elements that influence performance in repeated reading on fluency and comprehension: repeated reading with a model: audiotape model, modelling by an adult or more proficient peer; repeated reading without a model; repeated reading without a model compared to non-repetitive reading. 3. Period of time of intervention: 19–25, 10–18, and 6–9 months. 4. Effects by type of study: treatment/comparison, single group, single subject. 5. Number of repetitions versus reaching a specified criterion. 6. Text difficulty. 7. Effects by exceptionality.

Study	Wexler J, Vaughn S, Edmonds M, Reutebuch CK. 2008
Descriptions of outcomes included in the review	Comprehension and fluency measures: standardised tests and others. Reading: 1. words read correctly per minute, errors per minute, mix of both: or mean numbers 2. target words read correctly 3. reading efficiency scores. Comprehension: factual/inferential comprehension questions answered correctly.
Descriptions of contexts included in the review	Settings unclear.
Search details	1. A computer search of ERIC and PsycINFO was conducted to locate studies published between 1980 and 2005. 2. A hand search of 7 major journals that publish research on secondary students with reading difficulties/ disabilities including: Annals of Dyslexia, Exceptional Children, Journal of Educational Psychology, Journal of Learning Disabilities, Journal of Special Education, Learning Disability Quarterly and Scientific Studies of Reading. 1980 to 2005.
Number of studies and participants included	19 intervention studies: 11 single subject design, 2 interventions with a single group of students, and 6 using a treatment and comparison design. 340 grade 6-12 students who struggled with reading. Struggling readers included: (exceptionality labels) low achievers, students with unidentified reading difficulties, dyslexia and/or with reading, learning, or speech language disabilities.
Appraisal instruments used	What Works Clearinghouse (WWC) was used to differentiate high-quality scientifically based research from weaker research and improve confidence in the findings.
Description of main results	• Effect sizes for fluency and comprehension outcomes in studies using random assignment of students ranged from small to large (ES range = .23 to 1.02). • Effects sizes from standardised tests were very small. • Repeated reading interventions that incorporate the opportunity for students to preview the text with a model of good reading (for example, an adult reader or audiotape reading of the text) or someone to provide corrective feedback, make more gains in rate then students who do not preview the text or preview the text silently or on their own. (Most notable are listening passage previewing interventions in which students preview a passage prior to reading it while following along silently with an audiotape or adult model of good reading.) • The improvements made in reading rates may not generalize to unpractised passages, or to word reading accuracy or comprehension.

Study	Wexler J, Vaughn S, Edmonds M, Reutebuch CK. 2008
Description of main results	• Students participating in the repeated reading condition showed significant effects of treatment on oral reading speed scores (ES = .97) and silent reading speed scores (ES = 1.02). • Unclear whether setting a specified number of times to read repeatedly or reaching a certain criterion was more beneficial. • Outcomes of fluency interventions for secondary students did not seem to vary according to exceptionality. While repeated reading generally increased reading rate, this did not necessarily generalize to unpractised passages or comprehension skills.

Study	Williams KJ, Austin CR, Vaughn S. 2018
Review objectives	To examine the effects of spelling interventions on spelling outcomes for students with disabilities in Grades 6 through 12.
Descriptions of interventions/ phenomena of interest	Effectiveness of interventions on spelling outcomes.
Descriptions of outcomes included in the review	Spelling tests of various lengths (9 studies); spelling in written composition (3 Studies); Written spelling test plus follow-up written spelling test (1).
Descriptions of contexts included in the review	Secondary schools – non-specified.
Search details	1. Computer search of the databases of ERIC, Education Source, and PsycINFO for studies from January 2004 to April 2016. 2. Hand search of cross-categorical special education research journals considered representative of the field. The hand search examined studies published from April 2014 to April 2016. 3. Ancestral search of reference lists in articles included in this synthesis and the syntheses by Weiser and Mathes (2011) and Graham and Santangelo (2014).
Number of studies and participants included	13 studies: all single-case design. Numbers of studies involving various disabilities: Emotional and behavioural disorder (2); Specific learning disability (6); Orthopaedic disability (2); Intellectual disability (1); Autism (1); Intellectual disability and autism (1). Total number of participants: 29. By grade levels: grade 6 (8); grade 7 to 9 (8); grade 10-12 (13).
Appraisal instruments used	Authors indicated that quality of study was undertaken but instrument was not specified.

Study	Williams KJ, Austin CR, Vaughn S. 2018
Description of main results	• Overall, the studies included in the synthesis reported increased spelling accuracy for taught words following spelling interventions, or increased spelling accuracy in written compositions. • Specifically, systematic study strategies and technological assistance in interventions led to increased spelling accuracy for taught words and in written compositions. • Even though there were positive effects in spelling and large or very large degrees of nonoverlap as reported by Tau-U for most participants in systematic study strategies and technological assistance interventions, there are still many gaps in this literature. • There was some variability for individual participants, and some of the interventions could be interpreted as having minimal practical significance despite improved performance from baseline to intervention phases. Gaps: studies tended to measure improvement through proximal (researcher-developed) measures. Most studies did not assess the maintenance of spelling performance and generalisation of spelling skills to untaught words. Studies did not examine the impact of spelling instruction but rather systematic study strategies, such as cover-copy-compare, as well as technological assistance such as word prediction software. Noticeably absent from these interventions were the direct teaching of phoneme to grapheme correspondences and morphemic approaches, which are often used with students in the elementary grades. Future research for secondary students with disabilities should examine alternative, explicit approaches to spelling instruction, as well as the impact of those approaches for different types of disabilities.

Study	Xin YP, Jitendra AK. 1999
Review objectives	To use a meta-analytic technique to aggregate intervention studies on solving word problems and examine the relationship between study characteristics and intervention outcomes for students with learning difficulties.
Descriptions of interventions/ phenomena of interest	The effectiveness of word-problem-solving instruction across student characteristics (for example, grade, IQ); instructional features (for example, intervention approach, treatment, length); methodological features; skill maintenance; and generalization components. Separate analyses were performed for group-design studies and single-subject studies using standardised mean change and percentage of nonoverlapping data (PND), respectively.
Descriptions of outcomes included in the review	Measures of mathematical word problem solving. For Group Design Studies: The most commonly employed measures were criterion-referenced tests (86%) developed by researchers. Only 2 studies (14%; Gleason, Car9, and Boriero, 1990, and Hutchinson, 1993) employed items from standardised tests in addition to criterion-referenced tests. For Single-Subject Design Studies: Only one study (Hutchinson, 1993) employed items from standardised tests in addition to criterion-referenced tests developed by researchers.

Study	Xin YP, Jitendra AK. 1999
Descriptions of contexts included in the review	1. Settings included: classroom (resource, self-contained, or remedial) and pull-out (outside of the special education or remedial classroom). 2. Length of intervention treatment – 3 levels: short (equal to or less than 1 week, 1 to 7 training sessions); intermediate (equal to or less than 1 month, more than 7 sessions); and long-term (more than 1 month). 3. Implementor of intervention: teacher, researcher, or both. 4. Instructional arrangement: in groups or individually.
Search details	1. Database search: Educational Resources Information Clearinghouse (ERIC), Psychological Abstracts (PsycLIT), Dissertation Abstracts International (DAI). 2. Ancestral search. 3. Hand-search of issues of particular journals – Exceptional Children, Exceptionality, Journal of Learning Disabilities, Learning Disabilities Research and Practice, Learning Disability Quarterly, Remedial and Special Education, and The Journal of Special Education. Date range: All published studies located in journals were between 1986 and 1996, whereas unpublished doctoral dissertations identified were from 1980 to 1996. Overall: 1980 - 1996.
Number of studies and participants included	25 studies (15 published, 10 unpublished doctoral dissertations). Of the 25 studies, 14 were group design studies and 12 were single-subject studies - one study included both designs. Studies included participants from elementary (34%), secondary (49%), and postsecondary (17%) grades.
Appraisal instruments used	None were described.
Description of main results	• Results of the meta-analysis indicated that word-problem solving instruction improved the performance of students with learning problems and promoted the maintenance and generalization of the skill. The overall mean effect size after correction for sample size in the group design studies was +.89, while the treatment effect for single-subject studies was 89%. • All intervention approaches, apart from the other approach, yielded moderate to large mean effect sizes (range = .74 to 1.80). – The largest effect size was obtained from the samples coded as CAI intervention (ES = 1.80) – Representation technique was seen to be the next most effective approach in facilitating word problem-solving performance (ES = 1.77). – Strategy training was found to be moderately effective in facilitating acquisition of problem-solving skills (ES = 0.74). • Intervention effectiveness and student characteristics – Due to inconsistent findings, grade/age did not mediate the effect size of the intervention, although there was a large effect size for the postsecondary group (d = +1.68) followed by a moderate effect size for the secondary group (d = +.78), and a low effect size for the elementary group (d = +.47). – IQ had a mediating influence on the effect size. Students with IQ scores below 85 (ES = 1.87) scored higher than students with IQ scores above 85 (ES = 0.51).

Study	Xin YP, Jitendra AK. 1999
Description of main results	– Label had a mediating influence on the effect size. Students with LD (ES = 0.68) seemed to benefit less from the intervention than students with mixed disabilities (ES = 1.96) or those at risk (ES = 2.22). • Treatment outcomes and relationship to instructional features: setting, length of treatment, instructional arrangement, implementation of instruction, word-problem task, and student-directed intervention. – Due to inconsistencies, no conclusive findings for setting (classroom/ pull-out) nor for implementor of intervention (teacher/researcher/both). – Length of treatment did mediate the effect size. Long-term interventions (ES = 2.51) were more effective than short-term interventions (ES = 1.72) which were more effective than the intermediate-term interventions (0.73). – Instruction involving student self-regulation of strategy did not have a mediating influence on the effect size for word problem-solving interventions. – Instructional arrangement had a mediating influence on the effect size of the intervention. The effect size for individually provided instruction (ES = 2.18) was greater than that for small-group instruction (ES= 0.78). The difference was statistically significant ($p < .01$). – Word-problem task did mediate the effect size. Interventions involving simple one-step problems (ES = 1.89) yielded larger effect sizes than multistep word problems (ES = 0.38) or mixed problem types (ES = 0.63). f. Instructional grouping did mediate the effect size. Individually provided instruction (ES = 2.18) was more effective than group instruction (ES = 0.54). • Relationship between methodological features (publication bias, group assignment) and effect size. – Publication bias did moderate the effect size. Unpublished studies in this meta-analysis yielded larger overall effect sizes (ES = 1.81) than published studies (ES = 0.71). – Group assignment did moderate the effect size. Significantly higher effect size scores for matched (ES = 6.01) and randomly assigned (ES = 1.55) groups than for intact groups (ES = 0.11). • Effectiveness of word-problem-solving instruction in fostering skill maintenance and generalization, and whether skill maintenance and generalization are functions of instructional features (for example, treatment length). – wWrd-problem-solving instruction seemed to positively affect skill maintenance (d = 0.78) and generalization (d = 0.84). » CAI was found to be most effective in promoting maintenance of the skill (ES = 1.53) compared to strategy training (ES = 0.77). » CAI was not as effective (ES = 0.44) when compared to strategy training (ES = 1.09) in promoting generalisation.

A.9. Summary table of effective and ineffective strategies for instruction and intervention

The effects of the instructional practices that were identified in the reviewed studies are summarised in the table below. Whether or not each instructional practice was beneficial in improving the outcomes for struggling Years 7 to 9 students in literacy and/or numeracy are indicated. The contents of the table have been deliberately minimised. The table should be read in conjunction with the results section of the report.

Strategy or Skill	Example/Description	Studies
Reading		
Phonemic Awareness and Decoding	Note: There was no mention of interventions for older students that focussed on the strategy of phonics instruction in the systematic reviews included in our umbrella review. This does not suggest that these interventions are ineffective or should be avoided. Joseph et al. (2009) (included in the review) did state that students should be taught basic reading skills when screening indicated deficits in students' phonetic skills.	
Morphological instruction	Morphological instruction generally involves 4 instructional components: 1. improving awareness of the morphological structure of words 2. increasing knowledge of meanings of affixes and roots 3. supporting morphological problem solving 4. developing hypotheses about meanings of unfamiliar words through morphemic analysis. Morphological interventions **can be beneficial** for improving decoding and phonemic awareness outcomes.	Can be beneficial: Goodwin and Ahn (2013)
Vocabulary		
Explicit instruction	Explicit instruction is teaching that is structured and systematic, breaking down complex information or academic skills into small chunks or steps, providing scaffolds, modelling, offering opportunities to practice, and checking for understanding frequently. Explicitly teaching students the meaning of the vocabulary found in subject specific (for example, science) text **can be beneficial**.	Can be beneficial: Kuder (2017) Kaldenberg et al. (2015)

Strategy or Skill	Example/Description	Studies
Morphemic analysis	Morphemic analysis involves students being taught to predict the meaning of words based on their structural components. This is often done in 4 steps: 1. breaking words into their morphemic parts, that is, prefix, suffix, root 2. attaching meaning to each word part 3. making a prediction about the meaning of the unknown word based on the meaning of the parts 4. checking the dictionary for the definition This strategy **can be beneficia**l at improving vocabulary knowledge for secondary students.	Can be beneficial: Kuder (2017)
Mnemonic instruction	Mnemonic instruction involves using strategies such as acronyms, rhymes, or pictures to assist students with memory retrieval of taught content. Using mnemonics **can be beneficial** for improving vocabulary knowledge.	Can be beneficial: Kuder (2017) Kaldenberg et al. (2015)
Multimedia augmented instruction	There was evidence that content instruction, augmented using podcasts that were created using universal design of learning (UDL) principles, **improved** students' vocabulary knowledge.	Beneficial: Kuder (2017)
Reading fluency		
Repeated reading (RR)	Repeated reading is a teaching strategy that involves having students read short (50-200) word passages aloud to a teacher several times until they achieve a satisfactory fluency level. Repeated readings (RR) **were beneficial** for improving reading fluency. The impact of repeated reading can be further enhanced when combined with other instructional practices including: • explicit reading instruction • having good reading modelled (for example, by an adult or peer) • corrective feedback.	Beneficial: Joseph et al. (2009) Lee et al. (2017) Steinle et al. (2022) (mixed findings) Wexler et al. (2008)
Morphological instruction	Morphological instruction interventions (see definition above) **were not beneficial** on fluency outcomes.	NOT Beneficial: Goodwin et al. (2013)

Strategy or Skill	Example/Description	Studies
Reading comprehension		
Explicit instruction	Explicit instruction involves precision teaching of specific content or skills. The following **were beneficial**: 1. explicit text instruction targeting the linguistic structure and organization of expository text 2. explicitly teaching the definitions of science vocabulary words 3. explicitly teaching students a combination of reading comprehension skills and strategies (for example, previewing, clarifying, generating questions, and summarising) 4. explicitly teaching fundamental reading skills (for example, word study).	Beneficial: Hall-Mills et al. (2022) (Point 1) Kaldenberg et al. (2015) (Point 2) Edmonds et al. (2009) (Point 3) El Zein et al. (2014) (Point 3) Burke et al. (2015) (Point 3) Berkeley et al. (2010) (Point 4) Flynn et al. (2012) (Point 4) Joseph et al. (2009) (Point 4)
Strategy instruction	There was wide variation in the application of strategy instruction across studies. A description and examples are provided below. More detail about specific applications will be found in the individual studies that were included in each synthesis study reviewed. Strategy instruction involves the explicit teaching (see above) of meta-cognitive (for example, students actively reflecting on their thinking) or cognitive processes such as those used to solve problems, regulate attention, organise thoughts and materials, and monitor one's own thinking. Strategy instruction **was beneficial** for improving reading comprehension, and can be beneficially taught using the processes below. • SRSD instructional sequence: self-questioning, comprehension strategy training, self-regulated comprehension, and explicit writing frameworks • Main idea/summarisation: explicit teaching of instructional practices for reading text and identifying the most critical information (such as, main idea) and then how to link these main ideas across paragraphs to create summaries.	Beneficial: Kang et al. (2015) Berkeley et al. (2010) Daniel et al. (2021) Filderman et al. (2022) Kim et al. (2012) Solis et al. (2012) Stevens et al. (2019)

Strategy or Skill	Example/Description	Studies
Strategy instruction	• Questioning: interventions that involve direct questioning of students (with corrective feedback). For example, after each sentence, questions are asked to illicit reflection and clarify meaning, such as: 'Why does that make sense? Alternatively, modelling and thinking aloud how to self-question and reflect during and after reading, for example, 'Who or what is the paragraph about?' and 'What is happening to them?' • Targeting underlying structures: students are taught how to identify structures of different types of passages (main idea, list, and order passages) and organise the passages. • Background knowledge: explicit teaching of information needed to understand the text, including building students' content knowledge, (for example, science, social studies) or vocabulary knowledge prior to, or during, reading and strategy instruction. • Learning strategy instruction: students are taught skills to manage their learning. For example, emphasis on effort; accessing peer and parental support; managing the learning environment (for example, regulating study and break time, dealing with distractions); motivational aspects (such as, self-efficacy, task value, goal orientation); and metacognitive knowledge (such as, teachers model different strategies and explain their rationale to the students).	Berkeley et al. (2010) Solis et al. (2012) El Zein et al. (2014) Burke et al. (2015) Kim et al. (2012) Filderman et al. (2022) Donker et al. (2014)
Data-based instruction (DBI)	DBI is a systematic data-based approach to intensifying instruction. For students needing tier 2 or 3 intervention, this entails implementing an intervention to achieve a long-term goal in skill or knowledge acquisition and using data-based rules to determine when instructional changes are needed. It incorporates frequent progress monitoring and **is beneficial** for reading comprehension.	Beneficial: Jung et al. (2018) Dietrichson et al. (2020)
Graphic organisers	Graphic organisers are simple visual and spatial displays that use lines, arrows and spatial arrangements to describe text content, structure and key conceptual relationships. They **can be beneficial** for supporting reading comprehension. Examples include: • semantic feature analysis • cognitive maps • story maps • venn diagrams.	Can be beneficial: Filderman et al. (2022) Kim et al. (2004) Kim et al. (2012) Solis et al. (2012)
Semantic maps	Semantic maps (a specific type of graphic organiser) highlight the relationship between the words and content, and **can be beneficial** for supporting students' comprehension of science texts.	Can be beneficial: Kaldenberg et al. (2015) Solis et al. (2012)

Strategy or Skill	Example/Description	Studies
Morphological instruction	Morphological instruction interventions (see definition above) were **not beneficial** for improving reading comprehension outcomes.	NOT beneficial: Goodwin and Ahn (2013)
Repeated readings	Interventions targeting reading fluency, (for example, repeated readings) were **not beneficial** for improving reading comprehension outcomes.	NOT beneficial: Edmonds et al. (2009) Scammacca et al. (2007) Steinle et al. (2022) Wexler et al. (2008)
Writing		
Writing fluency	Writing fluency involves the following: • transcription (handwriting, typing, spelling) • text generation (phrases, sentences and compositions) with ease and automaticity. • executive functions (coordinating transcription and text generation through planning, translating and revising).	
Acquisition interventions	Writing acquisition interventions were found to be beneficial for students with low proficiency in the aspects of writing fluency listed above. These interventions included: • untimed tasks • explicit teaching through modelling, prompting, initial practice and error correction were beneficial. • graphic organisers for text generation.	Beneficial Datchuk et al. (2022)
Fluency interventions	Fluency interventions should be offered to students who have acquired the writing fluency skills listed above, but who need support to execute them with ease and automaticity. Instructional practices among fluency interventions include: • timed writing practice • performance feedback • goal setting • graphing performance. These interventions, combined with explicit instruction or SRSD (see above), **were beneficial** for increasing students' writing fluency.	Beneficial: Datchuk et al. (2022) Datchuk et al. (2020)

Strategy or Skill	Example/Description	Studies
Text writing		
Explicit/systematic instruction	Explicit instruction involving the modelling of strategies for planning, writing, revising and/or editing text, together with student practice of the strategies **were beneficial** at improving writing. Explicit instruction, coupled with other instructional strategies, were also included in other interventions (for example, see Learning strategy instruction below).	Beneficial: Gillespie et al. (2014) Kang et al. (2015) Peterson et al. (2020) Valasa et al. (2014)
Learning strategy instruction	Learning strategy instruction that is focused on improving self-regulated learning (including cognitive, metacognitive and management strategy skills), motivational aspects and metacognitive knowledge (see Reading above) **was beneficial** in improving writing. The explicit teaching of these strategies through SRSD, that is, the explicit teaching of writing strategies and the development of self-regulation skills, as well as the Strategic Instruction Model (explicit/ direct instruction of writing strategies) **were also beneficial**.	Beneficial: Asaro-Saddler et al. (2021) Donker et al. (2014) Gillespie et al. (2014) Kaldenberg et al. (2016) Kang et al. (2015) Valasa et al. (2014)
Graphic organisers	Using graphic organisers to plan ideas before writing **were beneficial** in improving text writing. Examples of helpful graphic organisers included: • computer-based graphic organisers • paper-and-pencil graphic organisers • teacher generated graphic organisers • student- generated graphic organisers.	Beneficial: Datchuk et al. (2022)
	Graphic organisers should be used as a supplementary tool in conjunction with effective writing instruction such as: • explicit instruction in how to use the graphic organisers • guided practice in using the graphic organisers • feedback to support learning.	Boon et al. (2018) Ciullo and Reutebuch (2013)
Spelling		
Morphological instruction	Morphological instruction that used morphological features to help students deal with complex spelling patterns and exceptions to grapheme–phoneme correspondence **was beneficial** for improving spelling outcomes.	Beneficial: Goodwin and Ahn (2013)
Systematic study procedures	Students used a specific procedure (such as., cover-copy-compare) to study words. There was **limited evidence** that interventions emphasising repeated practice with words and self-correction of misspelled words were beneficial for improving spelling outcomes.	Limited evidence: Williams et al. (2018)
Technology-based interventions	There was **limited evidence** to support using technology-based interventions. These interventions involved students learning to use word prediction software (for example, Co:Writer) to reduce the number of misspellings in written compositions.	Limited evidence: Williams et al. (2018)

Strategy or Skill	Example/Description	Studies
Mathematics		
Note: Some studies targeted improvements in overall mathematics skills. However, most studies targeted specific skills such as worded problems, algebra, fractions, basic number facts and arithmetic skills. Beneficial strategies were consistent across the targeted improvement areas of general mathematical skills and specific mathematics skills (word-problem solving, algebra, fractions, basic number facts and arithmetic skill) and so are presented together in this section of the table.		
Explicit, systematic instruction	Explicit instruction typically involves modelling the steps for solving problems, providing sufficient practice opportunities and ongoing feedback (see also below). This **was beneficial** for improving overall mathematics skills, algebra, fractions and arithmetic skills.	Beneficial: Bouck et al. (2018) Lee et al. (2020) Jitendra et al. (2018) Maccini et al. (2007) Shin and Bryant (2015) Watt et al. (2016)
	The explicit teaching of strategies to identify underlying structures and relationships, as well as cognitive strategies **were beneficial** for improving word problem solving skills.	Beneficial: Dietrichson et al. (2020) Donker et al. (2014) Lein et al. (2020) Shin et al. (2021) Xin and Jitendra (1999)
Visual models or physical representations	Using visual models or physical representations of mathematical concepts (for example, fraction circles or squares, manipulative objects, diagrams, mental representations) **was beneficial** for improving overall mathematics, fractions and word problem solving skills.	Beneficial: Maccini et al. (2007) Shin and Bryant (2015) Shin et al. (2021) Xin and Jitendra (1999)
	Multiple representations (for example, concrete and virtual manipulatives, pictures) **were beneficial** for improvements in algebra skills.	Beneficial: Jitendra et al. (2018) Lee et al. (2020) Watt et al. (2016)

Strategy or Skill	Example/Description	Studies
Guided practice and independent practice	Guided practice involves the teacher and students practicing strategy steps to solve problems together. Independent practice involves students practicing strategy steps and solving problems on their own. Both guided practice and independent practice **were beneficial** for improving overall mathematics, fractions, algebra, word problem solving and arithmetic skills.	Beneficial: Bouck et al. (2018) Maccini et al. (2007) Shin et al. (2021)
Logically sequencing examples	Examples presented to students that were arranged in a logical order (for example, addition, subtraction, multiplication, division) **were beneficial** for improving algebra skills.	Beneficial: Lee et al. (2020) Watt et al. (2016)
Monitoring student performance and/or corrective feedback	Increased monitoring of students' performances, as well as the provision of corrective feedback by teachers **were beneficial** for improving overall mathematics, fractions and arithmetic skills.	Beneficial: Bouck et al. (2018) Maccini et al. (2007) Shin et al. (2015)
Concrete-Representational-Abstract (CRA) approach	CRA involves explicit, sequential instruction in which mathematical concepts are presented using concrete manipulatives, then using pictorial representations (for example, diagrams), and finally in the abstract using digits (for example, equations). A CRA approach **was beneficial** for improving basic number facts, overall mathematics, algebra and arithmetic skills.	Beneficial: Bouck et al. (2018) Lee et al. (2020) Maccini et al. (2007) Watt et al. (2016)
Schema-based instruction (SBI)	SBI involves explicit instruction to identify and represent the underlying mathematical structures of problems. SBI **was beneficial** for improving word problem-solving skills.	Beneficial: Maccini et al. (2007) Lein et al (2020)
Other strategies		
Computer assisted instruction (CAI)	Interventions that employed CAI tutorial or interactive videodisc programs **were beneficial** for improving word problem-solving skills.	Beneficial: Shin et al. (2021) Xin and Jitendra (1999)
Mnemonic strategy instruction	Mnemonic strategy instruction involves teaching a word, sentence, picture, or technique for improving or strengthening memory. There was **limited evidence** that this strategy was beneficial for improving word problem-solving skills.	Limited evidence: Maccini et al. (2007)
Contextualised instruction	Contextualised instruction involves problems set in real-world scenarios. There was **limited evidence** that this strategy was beneficial for improving overall mathematics and word problem-solving skills.	Limited evidence: Jitendra et al. (2018) Maccini et al. (2007)

A.10. Useful resources

Branching Minds
https://www.branchingminds.com/

Center on Multi-Tiered Stem of Supports
https://mtss4success.org/

Iris Center
https://iris.peabody.vanderbilt.edu/

Kansas Technical Assistance System Ne2rk (TASN)
https://www.ksdetasn.org/

Keys to Literacy
https://keystoliteracy.com/

Macquarie Online Test Interface (MOTIf) assessments
https://www.motif.org.au/home/tests

Multi-Tiered Systems of Support for Mathematics
https://ceedar.education.ufl.edu/cems/math/

National Centre for Intensive Intervention
https://intensiveintervention.org/

National Center on Improving Literacy
https://improvingliteracy.org/

Academic Progress Monitoring Tools Chart
https://charts.intensiveintervention.org/aprogressmonitoring?_ga=2.246648708.1848310776.1664457303-392614326.1664163453

Academic Screening Tools Chart
https://intensiveintervention.org/resource/academic-screening-tools-chart

Office of Special Education Programs IDEAs That Work Website
https://osepideasthatwork.org/

PaTTAN Pennsylvania Training and Technical Assistance Network
https://www.pattan.net/

PROGRESS Center
https://promotingprogress.org/

Reading Rockets
https://www.readingrockets.org/

Self-Regulated Strategy Development
https://srsdonline.org/

Swift Education Center
https://swiftschools.org/

The Science of Reading Podcast
https://amplify.com/science-of-reading-the-podcast/

The TIES Center
https://tiescenter.org/

For more information visit
edresearch.edu.au | **monash.edu**

120523.CH.AA.V1