Nomanis

Reading | Teaching | Learning | Connecting

Issue 17, June 2024

SPOTLIGHT ON ASSESSMENT: THE WHEN, WHY AND HOW

The Academy
for the Science of Instruction

Essential knowledge for teachers

The Academy for the Science of Instruction draws on a wealth of research and expertise to offer in-depth courses for educators.

Expand your knowledge and skills in evidence-based teaching and the application of best practice in the classroom.

Explore our courses:

- Effective Instruction in Reading and Spelling
- Positive Teaching and Learning

Enrol today

scienceofinstruction.com.au | academy@scienceofinstruction.com.au

Reading | Teaching | Learning | Connecting

ISSN 2207-0478 (Online)

Nomanis is published twice yearly
by MultiLit Pty Ltd
Suite 2, Level 7, Building C
11 Talavera Road
Macquarie Park NSW 2113
Australia
www.multilit.com

Joint Editors
Emeritus Professor Kevin Wheldall AM
e: kevin.wheldall@multilit.com

Dr Robyn Wheldall
e: robyn.wheldall@multilit.com

Assistant Editor
Dr Nicola Bell

Editorial Team
Sarah Arakelian
Dr Mark Carter
Dr Anna Desjardins
Maddy Goto
Dr Alison Madelaine
Dr Ying Sng

Everything old is new again

"Five years ago, there was almost nothing known about how educators can use research well to improve practice."

(Deliberately unattributed, 2024)

Kevin Wheldall

Robyn Wheldall

We'd like to begin by stating rather bluntly that, despite what some may profess, the science of learning (and especially the concept of evidence-based instruction) is not something new. Some of us have been arguing in this vein for many years. For example, Richard Riding and the first author (KW) launched the journal *Educational Psychology: an experimental journal of educational psychology* in 1981, arguing for 'Effective Educational Research' in their first editorial. In the early issues we included articles on Direct Instruction (DI), classroom seating arrangements, classroom behavioural interventions, effects of contextual cues on reading, Precision Teaching, Theory of Instruction, morphographic spelling, etc. It might seem like something bright and shiny that has emerged over the last five years or so, but this is not the case.

Another example is the case of Reading Recovery, recently and at long last officially discontinued in New Zealand, its country of origin. But we showed experimentally that it was of limited efficacy 30 years ago (and took a lot of heat for saying so!). One could argue similarly that the experimental evidence for the efficacy of phonics instruction and Direct Instruction has been known for years. Possibly the biggest educational experiment in history, Project Follow Through was completed in the late seventies. It was largely, and arguably deliberately, ignored. While we celebrate the new-found commitment to these approaches, we must not forget that they come from a long research tradition. This commends them even further. It is not the case that because research is not new that it is to be viewed as outdated or 'back to basics' or 'old school' – which are terms often used pejoratively.

Similarly, voices protesting that 'phonics only' is not enough are nothing new. Whole language enthusiasts have protested this for years in spite of the fact that no one could point to a source claiming the contrary. 'Phonics only' has never been recommended by anyone! The National Reading Panel report of 2000, nearly 25 years ago, emphasised that phonics was only one of the 'Five Big Ideas' of effective reading instruction.

As the battles in the reading wars draw to a close, at least for now, and it has become increasingly obvious and accepted that the Science of Reading Instruction is the victor, it is disappointing to see minor skirmishes breaking out even among the erstwhile allies; fighting over the spoils of war, perhaps! Our point here is not to expose the apparent hubris of individuals but rather to identify unedifying trends in current thought and to reassert our continuing commitment to securing our thinking on what we can learn from empirical research evidence, based on the scientific method. As the late lamented Christopher Hitchens stated ('Hitchens Razor'), and with which we concur, "what can be asserted without evidence can also be dismissed without evidence". Let's look at some examples.

One argument that has been aired recently is that not all specific reading interventions and programs necessarily need specific empirical evidence for their efficacy. If the program/intervention makes conceptual sense, it is argued, and is based soundly on the empirical research supporting its operational principles, then it can be recommended as sound instructional practice. We would demur from this assertion. It is quite possible for a program/intervention to be sound in theory but weak in practice. We cannot be certain of efficacy unless it is empirically tested using the scientific method. This may be inconvenient, but it is necessarily the case. This is the distinction between evidence-informed as against evidence, based practice because "extraordinary claims require extraordinary evidence" (the Sagan standard). As we shall argue later, there are levels of what constitutes acceptable evidence.

Similarly, there are those who argue, contrariwise, that what works in theory, the Science of Reading, will not necessarily work in practice because education is a much more complicated, 'nuanced' process than that, and we cannot control all the relevant variables. There may be some truth in this. But rather than discarding it as unworkable, this simply emphasises the need for further scientific research to identify and isolate these potentially confounding variables.

Yet another source of controversy within the Science of Reading community is the argument regarding the superiority of teaching sounds before letters as against letters before sounds. There are advocates favouring each of these approaches, but many of us, in the absence of empirical evidence to the contrary, would argue that either/or is a false dichotomy and that there is no reason why print-to-speech and speech-to-print should not both be taught together simultaneously.

So where does MultiLit stand in all of this? We reaffirm and hold fast to the need for the scientific method as the basis for understanding what works.

We often hear about the research to practice gap – that challenge of taking what the research tells us and translating it into effective classroom practice. Here are some ways that the gap from research to practice can be closed:

- growing teacher knowledge
- implementing evidence-based policy
- using tested approaches.

But how we can know what approaches we can be confident in using to help close the gap between research and practice?

Some 15 years ago we argued for a simple model of evidence for efficacy comprising five levels. But before rehearsing this we will reprise the research that we follow (and endeavor to create) in the MultiLit Research Unit, and in the MultiLit company.

- We rely (in the main) on research studies that have an empirical focus and that appear in peer-reviewed journals.
- We place our confidence in the research in proportion to the rigour of that research.
- We preference experimental research over correlational research.
- We look for replication of the research findings in subsequent research to evaluate whether consistent findings can be found.

In program design, we also look to the instructional literature for the best way to put programs together for classroom use. And where there are unanswered questions, we need to apply what we do know and align our next steps as closely as possible to approaches of proven effectiveness (an informed 'best guess' if you like).

Let's just take a moment to remind ourselves about the empirical method:

1. Define the purpose of the research.
2. Explore theories and relevant literature supporting or challenging the research proposition.
3. Create a hypothesis (research question/s framed as a hypothesis) and determine measurement.
4. Specify methodology, research design and empirical data collection.
5. Conduct data analysis and compile results.
6. Draw conclusions.
7. And as we said previously, replication is very important.

There are a couple of important questions. First, is all evidence created equal? To that we would say a firm no. Second, how can we assess the strength of the evidence on which we seek to rely? To help us in this, back in 2007, the first author (KW) proposed a five-level scale. Using this scale helps us to weigh the evidence, and in some cases even reject it.

At Level 1, the evidence is research-based and makes conceptual sense in terms of current research and theory plus there are independent, replicated, randomised controlled trials (RCTs) providing strong evidence for specific program efficacy. This is the 'gold standard' to which all programs and interventions aspire, and such programs and interventions may be recommended with confidence. Unfortunately, they are very few in number.

At Level 2, the evidence is research-based and makes conceptual sense in terms of current research and theory, but the empirical evidence for specific program efficacy is more limited and may not include fully randomised controlled trials. This would count as 'very promising', and such programs could be recommended with reasonable confidence. It constitutes a 'silver standard' pending the collection of stronger evidence.

At Level 3, the evidence is research-based and makes conceptual sense in terms of current research and theory, but there is little or no empirical evidence for the specific efficacy of the program. Clearly, there is a need for supportive empirical evidence of specific program efficacy before such a program can be wholeheartedly recommended for wide application, but it may be 'worth a try' because it at least makes conceptual sense. In today's parlance, this is an evidence-informed approach or program. This is the minimum basis for program recommendation and constitutes the 'bronze standard'.

At Level 4, the quality of evidence is not research-based and makes no conceptual sense in the light of current research but may claim empirical evidence for specific program efficacy.

Such programs should not be adopted without further substantial empirical evidence for their efficacy and do not meet even the lowest standard of acceptability. Proponents of such programs should be invited to provide specific evidence, or at the very least cite supporting generic scientific research evidence or desist from making their claims. This is the 'brass standard'. When highly polished it might, at first blush, superficially resemble gold but is soon shown not to be so, on closer examination.

At Level 5, there is no reliable research-based evidence, and it is predicated on assumptions counter to substantial scientific evidence to the contrary such that any empirical evidence offered should be viewed with considerable scepticism. Such programs should not only not be adopted, but the public should be warned that the programs are unlikely to be effective and, rather than meeting any standard, should be regarded as requiring the educational equivalent of a 'health warning'. At best this is the 'tin standard'.

So, we must bear in mind the evidence credentials of the approaches and programs that we use in our classrooms. Instructional time is precious, and everything must earn its keep. We need to throw out the tin cans (not recycle them!), leave the brass ornaments in the attic as they lose their lustre, provisionally settle for bronze medals in the lack of competing or better alternatives, admire and keep burnishing our silver accomplishments, while continuing to strive for gold.

Emeritus Professor Kevin Wheldall AM
and Dr Robyn Wheldall
Joint Editors

Reference

Wheldall. K. (2007). Efficacy of educational programs and interventions. *Learning Difficulties Australia Bulletin, 39*(1), 3–4.

What we've been reading

Nicola Bell

What does it mean to actually 'enjoy' a book? Because at no point during my reading of *Yellowface* did it give me any pleasure, and yet I'm still glad that I finished it. In her acknowledgements, the author Rebecca F. Kuang describes the work as "a horror story about loneliness", and this succinctly captures why her writing manoeuvres the reader into experiencing dread from the very first page. How can a book that elicits such unsettling feelings and has such unlikeable characters who make such awful decisions still be so readable? I'm still trying to work that out, but I know some of my colleagues have also read the book, so maybe they have some clues.

Cain's Jawbone is another book I read recently that really toed the line between 'enjoyable' and 'just too hard'. I'd never heard of it before spotting it in my local bookstore, but based on its price, compactness and premise (in that order) I decided to give it a go. The book was first released in 1934 by 'Torquemada', who was a writer of cryptic crosswords and other puzzles. It comprises 100 short pages, which, if deciphered, provide sufficient information to reveal the details of six murders. The main hurdle for readers is that all the pages are intentionally out of order. *Cute*, I probably thought when I stood in that bookstore, smiling smugly at the 'warning' on the back cover that describes *Cain's Jawbone* as "extremely difficult and not for the faint-hearted". Well, that description is no joke. Cut to about a month later and you'll find me pinning colour-coded notes from various bookmarked Reddit threads to a bulletin board and feverishly explaining why page 84 precedes page 13 to my husband who is smiling, nodding and wondering if I'm losing my mind. Anyway, I'm happy to report that I am now living in a post-CJ era and the nightmares about it are definitely getting less frequent. Five stars.

Gabrielle Brawn

As this is the first time I have been asked to contribute to 'What We've Been Reading', I thought I would start off by noting that I enjoy different forms of reading. While I still sometimes read paper books, I also enjoy using technology to access books. I love the convenience of ebooks as they provide immediate access to a book and have the benefits of an inbuilt dictionary and being able to enlarge the font size (useful if you don't have your glasses to hand)! I also 'read' via audiobooks – although maybe some would not count this as reading. In choosing an audiobook, the voice of the narrator is very important to me. I find that autobiographies read by the author are particularly effective. A recent highlight was *Finding Me: A Memoir* by Viola Davis, for which she won the 2023 Grammy Award for Best Spoken Word or Non-Musical Album for her narration of her personal and powerful story. Other reading 'performances' I have enjoyed include Meryl Streep's narration of *Tom Lake* and Tom Hanks' reading of *The Dutch House*, both by Ann Patchett.

I always return to a good mystery/crime series, and I recently finished *The Running Grave* written by J K Rowling under the pseudonym Robert Galbraith. This is the seventh book in the series about private investigator Cormoran Strike and I now must wait until J K Rowling writes the next instalment. I recently discovered a series by Canadian author Louise Penny set in a fictional remote village in Quebec called Three Pines and featuring Chief Inspector Armand Gamache of the Sûreté du Québec. I am currently on book three, *The Cruelest Month*, and I am delighted to find there are 18 books in this series with another due later this year, so lots of murder and mystery ahead! As much as I enjoy a good mystery, I don't think I will attempt *Cain's Jawbone* as previewed by Nicola – sounds like too much hard work for me!

Mark Carter

My recent reading has been in the area of natural history, inspired by interactions with our local wildlife. We have an increasing number of birds visiting our house, including the mischievous and sometimes 'bitey' rainbow lorikeets, friendly and elegant king parrots, and the strongly discouraged, garbage bin raiding, bonsai decimating, juvenile delinquents of the bird world, the sulphur-crested cockatoos. However, the most interesting encounters are the occasional sightings of an owl on my pre-dawn walk. With forward facing eyes, human like round face and ability to fly without any apparent sound, they are certainly the most enigmatic of birds. *What an Owl Knows* by Jennifer Ackerman is an absorbing exploration of the remarkable variety of these birds and their extraordinary range of capabilities. For example, their disc-like face can act as a giant third ear and at least some species seem to have cross wiring of the vision and hearing parts of their brains, suggesting they can actually see sounds. Equally remarkable are some of the ingenious strategies used by researchers to investigate the talents of these birds. This book is recommended to anybody who encounters an owl on their early morning walk.

Encounters with other native wildlife are less frequent than for birds who come to us, but we do see the odd echidna, wallaby, snake, native rodent, and the ever-present possums, who party on our tin roof as we attempt to sleep. After a chemical spill in the 1970s resulted in the local extinction of platypus, they have been recently reintroduced in our area. This inspired me to read *Platypus Matters: The Extraordinary Story of Australian Mammals* by Jack Ashby. The book addresses a wide range of Australian mammals, but the platypus gets star billing. This is unsurprising given Ashby states, without a hint of bias, that the platypus is the best animal in the world. They really are quite remarkable, with rubber bodies that would be the envy of a contortionist and electrical sensors in their bills. They also possess Swiss army knife limbs that enable them to walk, dig and swim, and in the case of males, deliver an extraordinarily painful venomous sting.

Among the many interesting facts about other native mammals is that wombats can supposedly out sprint Usain Bolt over a short distance. It is acknowledged that this has never actually been tested, probably to spare Bolt the humiliation. One aspect of the book that was completely unexpected was the political dimension. Apparently, when encountered by Europeans, the monotremes broke the existing taxonomic system. After an extended period of denial, a hierarchy of mammals was constructed with the 'primitive' egg-laying monotremes at the bottom, marsupials in the middle and the 'advanced' placental mammals at the top, with garden variety humans being at the absolute pinnacle, naturally. Ashby clearly points out monotremes are not in any way primitive and simply represent a different branch of mammals that are brilliantly adapted to their environment. Apart from describing many Australian mammals as primitive, there are many additional forms of "othering" (Ashby's term) of Australian wildlife, including inappropriate comparative naming, suggesting they are inferior copies of better-known animals and the patently absurd claim that "everything in Australia is trying to kill you". I live near the bush and have not been killed once – yet. Ashby suggests that "othering" of Australian mammals explains, at least to some extent, our world-leading extinction rates and poor environmental regulation. While this may well be a contributing factor, as with most postmodern mutterings, that which can be asserted without evidence, can be dismissed on the same grounds. Nevertheless, Ashby provides a fascinating insight into the life and recent history of Australian mammals.

Anna Desjardins

Like my colleague Maddy, I spent a long time over these last few months with *Demon Copperhead* by Barbara Kingsolver. While a book by Kingsolver really requires no review – she captures human experience in such an authentic way and always impresses with her ability to voice a diverse cast of characters – this book certainly deserves one.

Kingsolver shines a light on the recent history of an impoverished pocket of the state of Virginia in the 1990s and early 2000s, as she explores the tussle between 'nature' and 'nurture' for her protagonist, Demon, who spends a

large part of his childhood in foster care. She gives us a heartbreaking sense of how the environment children are bequeathed when they come into this world carries the force to twist and crush them entirely, while concurrently imbuing the story with a deep appreciation of place and its importance for identity. Her similes alone always had me sitting up in recognition of a master wordsmith. Just to give one example, when Demon spends time away from his home in the city, he says, "I made my peace with his place, but never went a day without feeling around for things that weren't there, the way your tongue pushes into the holes where you've lost teeth." Nothing short of brilliant.

Demon Copperhead required a lot of processing time, and I definitely needed something lighter to follow it up. *Remarkably Bright Creatures* by Shelby Van Pelt was doing the rounds of our Product Development team and fit the bill. With an octopus as one of the main characters, it was a refreshing and quirky read – it felt a little too light-on at times, but the human characters grew on me and the feel-good ending was satisfying, stopping just short of being too twee. I also dipped into the first of *The Thursday Murder Club* series by Richard Osman. This was a book with another fun choice of characters, as senior citizens turn sleuths in the classic murder mystery genre – *The Thursday Murder Club* members are full of verve, and as an airport purchase for non-demanding reading during a flight, it entertained while offering some touching observations of life in different age brackets.

Feeling ready for something meatier again, I turned to an unusual non-fiction choice for me (I like to think Mark Carter would be proud!) and waded into *Courting: An intimate history of love and the law* by Alecia Simmonds. Alecia is the searingly intelligent friend of a friend, and I had attended her book launch at the State Library of NSW some months prior. I bought the book on the back of her entertaining presentation (and witty title), but then failed to open it for some time. Having now dived in, I have been enjoying a rollicking ride through our courtrooms.

The book reviews a rather niche area of the law, 'breach of promise of marriage' actions in Australia between the time of the early colony up to when the action was abolished in the 1970s (yes, one could sue a partner for breaking an engagement not so long ago, receiving monetary compensation for 'lacerated feelings'). Simmonds selects a number of key cases that allow her to trace parallels between the action, the settlement of the colony (and corresponding desettlement of the original inhabitants) and our societal norms and changing attitudes towards gender and love. She clearly enjoys bringing the cases to vivid life, so that we feel like those crowding into the courtrooms for the salacious entertainment that they provided, and she lends her litigants depth by detailing their lives both before and after their actions. The book has made me think about how much the law overlaps with and illuminates our history.

Finally, from time to time, I consult *365 Poems for Life* compiled by Allie Esiri, one for each day of the year if one so desires, with poems chosen from the greats of yesteryear and into the modern day. It so happens that on writing this, just yesterday, the chosen poem was from *Pippa Passes* by Robert Browning – it was the exact excerpt that my grandmother, Pippa, once voiced on radio and whose scratchy recording I had been trying to decipher for some time.

Maddy Goto

For my first taste of Barbara Kingsolver, *Demon Copperhead* was a good one. I didn't know much about it when I added it to my Kindle library but was intrigued by the title and its prize-winning stickers. At the time of adding it, I hadn't seen it in print and so didn't realise quite how hefty it was. It took me a long time to get through, not only because of its length. There were parts where I slowed down to savour the language, parts that made me laugh, parts that made me wince and parts that took some time to process and remained with me for a long time afterwards. Kingsolver's fictional creation is brutally real, and Demon's narration sucks the reader right into it. I'm glad I discovered this one as a book to read rather than listen to. It gave me the time to 'go slow', reread and ponder things like the seemingly simple but oh-so-clever naming of characters, and how much they add to the narrative.

I balanced the bleak world and written text of *Demon Copperhead* with a couple of audiobooks for the car. I've decided I'm not good at listening to fiction – I get distracted too often and miss bits and after a while seem to find myself getting irritated by the

narrator. So, it's non-fiction for my car journeys. For quite a long time I've thought about how I've changed as a reader, not only my reading habits but how I engage with (all sorts of) texts. I know that having a phone in my pocket has had a frighteningly big impact. In *Reader, Come Home: The Reading Brain in a Digital World*, Maryanne Wolf examines how the digital world has affected our ability to read deeply and the implications this has for our kids. I found it quite hard-hitting and depressing at times, not least because it confirmed my theories on why I'm a different reader now to the reader I was 25 years ago. It's not all doom and gloom though, and Wolf wraps it up with some hopeful ways we can try and mitigate it. Recommended reading (or listening).

Alison Madelaine

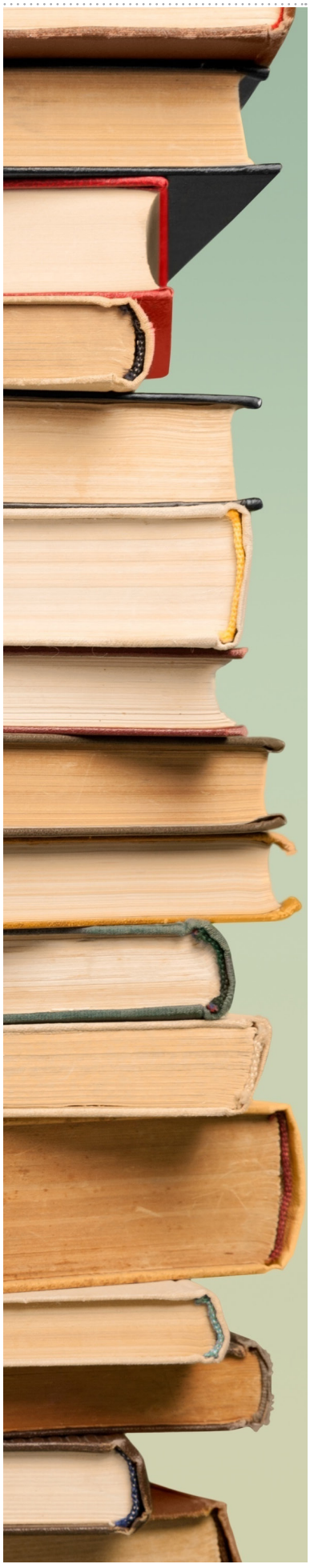
Late last year, I received my first two ARCs (Advanced Reader Copies): *A Shadow at the Door* by Jo Dixon and *The Dinner Party* by Rebecca Heath. Both were domestic thrillers that I really enjoyed over the Christmas/New Year break. *The Dinner Party* was my favourite of the two. During a neighbourhood dinner party in the late 1970s, the couples left their children at home asleep in their beds, checking them every so often as people did back then. But one night, a young baby went missing. Forty years later, the case has still not been solved and her older sister gets a visit from a woman claiming to be the missing baby. The story is told mostly in the present day, with flashbacks to the night of the dinner party and transcripts from episodes of a podcast series on the disappearance of the baby. It is somewhat of a slow burn as the reader gradually finds out that all is not as it seems and many of the characters have secrets they are trying to keep hidden. I could not put this book down as I had to find out the truth of what really happened the night of the dinner party. Rebecca Heath now has my attention, and I am looking forward to reading her previous thriller.

In the past few months, I seem to have read my share of heavy fiction containing difficult and disturbing events: *The Beekeeper of Aleppo* by Christy Lefteri, *Chai Time at Cinnamon Gardens* by Shankari Chandran, *Demon Copperhead* by Barbara Kingsolver, *Yellowface*, by Rebecca F. Kuang, and *Prophet Song* by Paul Lynch. After reading about the difficult journey of Syrian and Sri Lankan refugees, racism in Australia, a boy growing up in less-than-ideal circumstances, plagiarism, cultural appropriation, internet trolls, and family struggles amidst the rise of totalitarianism in contemporary Ireland, I really needed something a bit lighter. It may seem like a strange choice, but for me, that is crime fiction. *Resurrection Walk* by Michael Connelly and *What Happened to Nina?* by Dervla McTiernan were both great page-turners. *The Seven* by Chris Hammer was also an excellent read, incorporating three timelines.

Ying Sng

According to their website, a Street Library is "… a tiny vestibule of literary happiness". They give me an inordinate amount of joy and I cannot resist checking them out. On a jaunt around the neighbourhood, I saw a copy of *The Cuckoo's Calling* by Robert Galbraith. Crime fiction is not one I naturally gravitate towards but I knew the Cormorant Strike series were immensely popular, so I liberated the book to find out what the fuss was about. Once I'd finished reading it, I looked up how many more were in the series and quickly realised the best thing about stumbling upon a new-to-me series is the back catalogue (or is it the front catalogue in this case?). The other books were devoured in quick succession but please don't ask me about the plot. All I remember is they generally start with a character dying under suspicious circumstances which leads to someone close to the victim securing the services of Cormoran Strike and Robin Ellacott, his extremely competent assistant. Needless to say, the duo gets into some strife during the investigation, engage in some romantic shenanigans (not with each other …yet) and the book culminates with a clever resolution. Perfect reading for pleasure fodder!

A run of reading in this genre must have done something to my search algorithm because other crime fiction titles kept coming across my feed, including *Central Park West* by James Comey. If you think that name sounds familiar, you'd be right – it is that very tall man from the FBI. The book was OK. If you see it in a 'tiny vestibule', pick it up but don't expect too much in the way of 'literary happiness'.



I have also gone down a rabbit hole of Japanese fiction. It began with *Before the Coffee Gets Cold* by Toshikazu Kawaguchi and was quickly followed by *Sweet Bean Paste* by Durian Sukegawa, *Days at the Morisaki Bookshop* by Satoshi Yagisawa, *What You Are Looking for is in the Library* by Michiko Aoyama and *The Kamogawa Food Detectives* by Hisashi Kashiwai. I got a real feeling of contentment from reading them. I wonder if this sense comes from the absence of a major story arc and the undercurrent of morality. Whatever it was, they were all very satisfying and relaxing reads. However, I do lament my inability to read in another language – it would be nice to compare the original composition to the translation. If this has piqued an interest, I'd recommend *What You Are Looking for is in the Library*. This book was my favourite! It is a collection of stories where people go into a library and the quirky librarian helps them find something that they didn't know they were looking for. Seriously, how can you resist a book with that title?

I have been doing a bit of academic reading on the topic of reading comprehension and had a comprehension epiphany of my own while reading *The Bee Sting* by Paul Murray. I alternated between reading and listening to the book – *Whispersync for Voice* makes it so easy to switch between audiobook and ebook. Each chapter is told from the point of view of the (mostly irritating) characters and I was getting increasingly frustrated with them as I read or listened. About halfway through the book, I realised that Paul Murray had written all the chapters for Imelda, the wife and mother character, with limited punctuation. I had somehow managed to only listen to her chapters where the lack of punctuation was not evident because the narrator inserted pauses and used appropriate intonation. Once I saw the text laid out in Imelda's chapters, it gave me a completely different mental picture. The meagre use of punctuation provided a richness to her background and almost added a sense of desperation that the audiobook couldn't. I suppose I knew this implicitly but that "Aha!" moment confirmed the edge printed text has over audiobooks.

Although *The Bee Sting* provided me with a moment of self-development, I'm not sure it is a book I'd recommend. It was on the Booker shortlist and has won some other awards. Maybe it is a book that should be read twice. Maybe I'll re-read it one day. Maybe I won't.

Kevin Wheldall

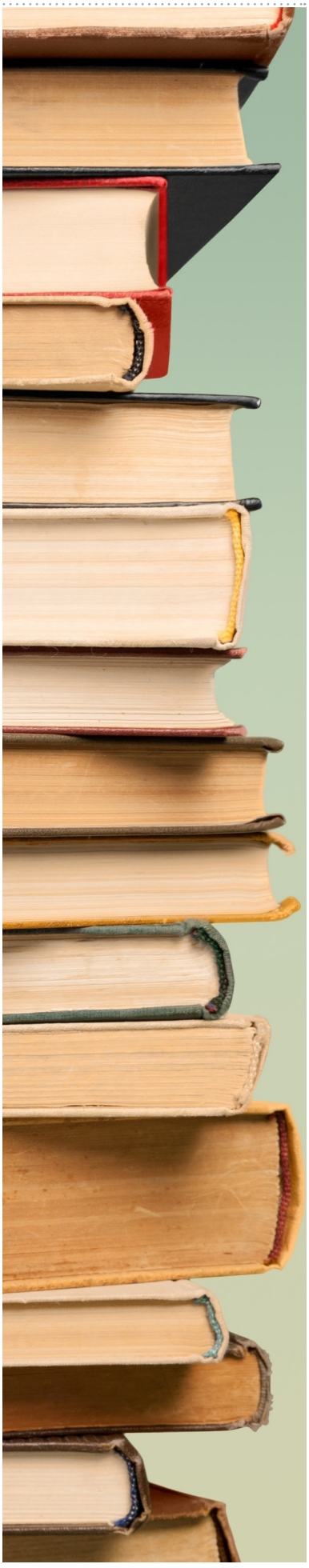
Like some of my other colleagues, I have read and been intrigued by *Yellowface* (by Rebecca F. Kuang); intrigued and discomforted in equal part. I find it hard to actually like books where the principal protagonist, and in this case the narrator, is profoundly unlikeable. It does not help that there is no other character to love either. Doubtless this is rather shallow of me, but I appreciate the author's skill, intelligence, and ingenuity, nonetheless.

I enjoyed Willian Boyd's *The Romantic*, and also *Trio*, but perhaps not as much as usual for this confirmed Boyd fan; but he is a fine writer. *A Heart Full of Headstones* by Ian Rankin and *Treasure and Dirt* by Chris Hammer were both enjoyable enough.

Anne Patchett, however, has delivered for me in spades of late. *These Precious Days* is a delightful collection of essays and her latest novel, *Tom Lake,* was a joy to read. I have written before about my difficulties with *Heart of Darkness* but an earlier book by Patchett, *State of Wonder*, loosely based on Conrad's novella, was a revelation. Pursuing similar colonial themes, her evocation of life on the far reaches of the Amazon is breathtaking. An encounter with an anaconda is particularly riveting.

The Running Grave by Robert Galbraith (aka J K Rowling) signalled a return to form following, for me, her disappointing previous novel, *The Ink Black Heart*. Her latest is a very satisfying account of a deeply sinister cult. Having said that, 960 pages is asking a lot of the reader: author in need of a firm editor. But there is a lot to love about J K not least her brave public stand for women's rights on 'the socials'. She has made a lot of enemies in the process and has been subject to outrageous abuse.

Two standout books for special mention were *Wifedom* by Anna Funder and *Mornings in Jenin* by Susan Abulhawa. The former recounts the role of his first wife, Eileen O' Shaughnessy, in the life and literary works of Eric Blair, better known as George Orwell. Funder makes a strong case for acknowledging Eileen's considerable, and until now relatively little-known influence on Orwell's writings, not to mention



putting up with an insufferable husband! Both *Animal Farm* and *1984* owe a debt to her own brilliance. Abulhawa's novel recounts the travails of growing up as a continually displaced Palestinian girl and woman in post-war Middle East. It is a moving story with a clever, complex plot and, while highlighting the plight of ordinary everyday Palestinians over decades, she manages to treat her Jewish characters sympathetically and with deserved respect. Both of these novels make for rewarding reading and are highly recommended.

Robyn Wheldall

In January this year, I took a month's leave from work. There were plans. And then a nasty virus struck. All plans were unmade. This was disappointing but there was a silver lining. As I was unable to do much on the holiday list, January became my month of reading. You could argue that I would have read books on my holiday anyway. This is true but no other summer has been so productive in terms of digesting books – eight in all. But before 'the bug', I finished the first book. It was a Christmas gift – *The Exquisite Art of Getting Even* by Alexander McCall Smith. A book of four short stories, it was the perfect way to while away the lazy days between Christmas and New Year. Although the title sounds rather mean, the final message of the book is an invocation to mercy and forgiveness, as revenge is never a worthy thing. The introduction to the final story provides a good summary. "Forgiveness heals; it allows us to unclutter our lives with the business of the past; it makes room for human flourishing" (p. 168).

Other books in my summer haul touched on some similar themes, by chance. I finished reading *Bright Shining* by Julia Baird. Excellent as always, Baird firmly grasps life with both hands and exhorts us to show grace and pursue 'moral beauty'. *Liturgy of the Ordinary* by Tish Harrison Warren is a call to find meaning in the everyday tasks that make up the daily grind, or more positively, our daily life. Quoting Dr Johnson at the beginning of the book set the tone, hooked me in and the book delivered. "It must be remembered that life consists not of a series of illustrious actions, or elegant engagement; the greater part of our time is passed in compliance with necessities, in the performance of daily duties, in the removal of small inconveniences, in the procurement of petty pleasures." This book helped me with the sorting of paperwork over the Christmas break (a task I definitely do not relish) and raised hope that I might practise this approach throughout the year, as close I got to a New Year's resolution.

Bittersweet by Susan Cain (another Christmas gift from a thoughtful colleague) is a great read from the insightful writer who brought us *Quiet: The Power of Introverts in a World That Can't Stop Talking* more than 10 years ago. Another non-fiction title, Cain helps us to understand how love, loss and sorrow make us whole; that we should embrace and use the pain in life rather than run from it. Cain dedicates the book to the memory of Leonard Cohen, quoting from his song *Anthem*, "There is a crack, a crack in everything. That's how the light gets in."

Having had my fill of non-fiction, I turned to novels. *Jesustown* by Paul Daley is a skilful, challenging and historically informed piece of fiction about the brutality of the frontier wars and the conflict between Indigenous and European Australia. *No Words* by Maryam Master is a children's book about a mute boy who is traumatised in Iran. It is a story of friendship and finding his voice in his new country. The author, born in Tehran, fled persecution after the Iranian revolution and escaped the country with her family, arriving in Australia at age 9. Master's own experience creates a depth of experience for the reader. Continuing the refugee theme and also set in Australia, I found *Hopeless Kingdom* by Kgshak Akec a compelling story of the lives of Sudanese refugees, told simultaneously from the perspective of a child and her mother.

Leaving the best till last, I was enthralled by Holly Ringland's *The Lost Flowers of Alice Hart*. This book is an epic tale of the life of a young, orphaned girl born into a family marred by domestic violence, anger and fear. It is a modern Australian masterpiece in my opinion – resonating with Tim Winton's *Cloudstreet* to my mind. The thematic use of Australian wildflowers that have their own language is inspired and Ringland captures the vast and varied Australian landscape with mastery. An amazing debut novel. And so my summer of reading concluded, with five of the eight titles being from Australian authors.

Which should we use, nonsense word tests or word ID tests?

Tim Shanahan

The answer depends on what you are trying to learn about your students' reading skills.

Teacher question:
I am an Assistant School Superintendent. We are moving towards explicit phonics instruction this year and are debating between using the nonsense words assessment or the decodable words assessment. Do you have thoughts about this? I have consulted with several people who I respect, and opinions are varied and passionate.

Answer:
I feel your pain.
Recently, a colleague asked me to make a similar recommendation to help figure out something about a grandchild's reading. I suggested the use of DIBELS Nonsense Word test, given the specific purpose and its easy availability.

You'd have thought I'd recommended drowning kittens or banning the Barbie movie!

People do get passionate about the strangest things.

I try to save my passion for non-empirical questions (Go Cubs, go!). If we have data that will allow us to make a sound determination, I'd turn the heat down and try to follow the numbers. Remember, this is about trying to do what's best for kids. It is not an opportunity to vent your spleen or espouse your philosophy.

There are two different kinds of tests used to determine student progress in decoding. Both kinds have a proven ability to evaluate how well students are learning their phonics and both can predict later success with oral/text reading fluency and reading comprehension.

Word identification tests have been around for a long time – more than 100 years. Nonsense word or pseudoword tests are a newer development.

Researchers were concerned about the validity of word identification tests for determining the effectiveness of decoding instruction. Word identification tests often focus on irregular spellings (e.g. 'the', 'of', 'done'), the kinds of words that are inconsistent with the spelling patterns usually stressed in phonics. Such tests couldn't tell you much about the effectiveness of phonics instruction. Even word tests with more common spellings were suspect. With such tests it was impossible to know if a student decoded a word or just remembered it from previous exposures.

The solution to the problem was the creation of nonsense word or pseudoword tests. Because the researcher (and, later, the test designer) constructs the words by mimicking English spelling patterns, there are no exceptional spellings, one offs, accidents of morphological history, and the like. Whether teachers are leading the kids to memorise Dolch or Fry list words or are just providing them with repeated exposure to certain words through phonics instruction, it was certain that the students wouldn't have previously seen letter combinations like 'dop', 'lan' or 'sepe'.

The idea was that a nonsense word measure would provide a purer look at how well students can decode, and their performance on such a test should reveal their decoding progress.

As is often the case, scientists may identify a real problem, but solving it is not always so easy.

At first blush, the nonsense test appeared to do a terrific job of assessing decoding ability, perhaps more valid than the traditional word identification test.

Over time, their faults became evident.

Often, if teachers know that their students are to be evaluated with nonsense words, they start teaching them to the students. This teaching is a waste of time for producing readers and renders useless the intended improvement in test design. Researchers and school district administrators must be

vigilant in discouraging teachers from fraudulently enhancing their students' test performance. (I don't think most teachers are intentionally trying to defraud – they just want to make sure their kids do well on the test, and teaching the specific test items seems logically to be the most direct route to that outcome. Well-meaning but unfortunate.)

A more important issue has to do with the nature of decoding. There is more to decoding than pronouncing letter patterns. Pseudoword tests provide a useful assessment of that part of the process, but not of the rest.

As Richard L Venezky so aptly described the process:

> *A third function of phonics is to generate a pronunciation for a word … This function is problematic, in that the imperfections in English orthography make such generation uncertain. If a word is totally unknown, the reader has little basis for deciding whether any particular pronunciation is correct or not. (*Venezky, 1999*, p. 202)*

Phonics is a tool for helping readers to decode the words in a text. But that is a necessarily imperfect process due to the complexity of the English spelling system. Some 'experts' throw up their hands, ready to surrender. For them, phonics would be useless because of the complexity of our spelling system. But as Venezky points out, readers don't need to arrive at exact pronunciations. Reasonable approximations are good enough, and then the readers make adjustments and consider alternatives based on their knowledge of the English language.

Nonsense tests, by their very design, can tell us whether students have managed to master particular spelling patterns, but they prevent students from any kind of self-evaluation and adjustment of pronunciation, which are key aspects of decoding. As such, these tests may do a good job of evaluating student learning from a decoding program, but they are unlikely to do equally well in predicting later reading achievement, as measured by oral reading tests or reading comprehension tests.

What do the research studies have to say about the usefulness of these measures?

For the most part, word identification tests and nonsense word reading tests tend to be interchangeable early on. There are copious amounts of validation data showing the value of both (e.g. Fien et al., 2008; Vanderwood et al., 2008). They both work reasonably well (i.e. there are high correlations between these measures and other reading tests).

However, in direct comparisons in which students are taking both tests so that they can be evaluated head-to-head, the word identification tests tend to do a bit better. For example, in one well-done study it was found that word ID tests provided a "clearer index of reading growth" (Clemens et al., 2014). Early in first grade, the tests were indistinguishable, but by second semester the word identification tests inched ahead.

Similarly, in a very large study of first graders (n = 3506, from 50 schools), it was reported that the nonsense word fluency tests did the best job of predicting end of year reading fluency and comprehension for most kids (Fien et al., 2010). There are other studies of this with similar results (e.g. Fuchs et al., 2004). However, this was not true for the higher achieving students. As kids' reading advanced, leaving out those word identification skills that Venezky noted becames a real problem.

By third grade, the correlations between nonsense word reading and word ID separate to a greater degree with the real word performance becoming the best predictor of oral reading fluency (ORF) for most kids (Doty et al., 2015).

Finally, a recent meta-analysis of data shows that across many studies, word ID tends to have the best relationship with various reading outcomes (January & Klingbeil, 2020).

None of these differences just noted are especially large, though they are often statistically significant. Nevertheless, some authorities suggest including both in early reading inventories, and that makes a certain kind of sense since they tap a slightly different array of skills.

I certainly have no problem with ongoing monitoring of decoding skills with nonsense words, alongside a word reading check to determine how well kids can read those most frequent words.

If you are only going to give one, and your specific interest is monitoring phonics progress in grade K–2, I'd go for a real word reading test – especially second semester of first grade or later and with my highest achieving schools. Those tests should do a slightly better job of revealing student progress towards success in reading. Just make sure, given your purpose, that the word ID test that you choose includes many words with regular spelling patterns.

But remember the differences here aren't large. In a different situation (e.g. I'm a school psychologist and a student has been referred to me due to a concern about their phonics ability), I would likely give you a different answer. You really can't go too far wrong in this case.

This article originally appeared on the author's blog, Shannon on Literacy.

Timothy Shanahan [@ReadingShanahan on X] is Distinguished Professor Emeritus at the University of Illinois at Chicago and was formerly Director of Reading for the Chicago Public Schools, and President of the International Literacy Association. He is a former first grade teacher and is a parent and grandparent. His website www.shanahanonliteracy.com *is popular with parents and teachers.*

Does ChatGPT mean we have to change how we assess?

Daisy Christodoulou

ChatGPT is capable of producing original and high-quality essays with minimal effort.

What does this mean for educational assessment?

Many people argue that it shows the paucity of our current assessment tasks. If we are setting assessments that a robot can complete, surely that shows the assessments are not good enough or hard enough or just plain 'human' enough and that they therefore need reworking.

I have previously appeared on Good Morning Britain where Dan Fitzpatrick made exactly this argument, and Susannah Reid agreed with what she called a 'profound' point. (Unfortunately, the debate ended there, and I did not get the chance to respond.)

They are not the only ones. In the *Financial Times*, the columnist Camilla Cavendish suggested the following: "Rather than banning ChatGPT, teachers should ask pupils to give it an assignment and critique its response."

Marc Andreessen has weighed in with a similar argument on X. "'ChatGPT plagiarism' is a complete non-issue. If you can't out-write a machine, what are you doing writing?"

And even before ChatGPT existed, people were making this argument in the context of other technologies. Here is the economist Daniel Susskind in his book A World Without Work.

> *Think of the way that we teach and test mathematics, for instance. Many of the problems we set students in secondary school, if not university, can now be solved by apps like PhotoMath and Socratic: take a photo of the problem, printed or handwritten, with a smartphone, and these apps will scan it, interpret it, and give you an instant answer. It is not a good sign that we still teach and test mathematical material in such a routine way that free off-the-shelf systems like these can handle lots of it with ease.*

As well as those examples, I have heard similar sentiments expressed to me by many, many people over the last month or so, to the extent that it almost seems as though this is the prevailing opinion. ChatGPT can write essays? We will have to set harder/different essays or assessments then!

As popular as this argument is, I disagree with it. There are three big flaws with the 'we should set assessments that computers can't complete' argument. It's an argument that misunderstands basic principles about (1) technology, (2) education and (3) assessment.

1 It's quite hard to find educational assessments computers can't do

Even before ChatGPT, this was true, as Daniel Susskind's point inadvertently makes clear. Now ChatGPT has come along: as well as writing essays, it can get passing marks in a number of prestigious professional qualifications. The solution of 'getting kids to critique a ChatGPT essay' is not going to work either,

as ChatGPT is rather good at critiquing its own responses. I suspect it would also be good at critiquing critiques of its responses, and critiquing critiques of those, and so on *ad infinitum*.

The basic technological principle here is Moravec's paradox, first developed in 1988, which is that computers find the types of academic skills we teach and assess in schools trivially easy. In Moravec's (1988) words: "It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility" (p. 15).[1]

Hans Moravec suggested a biological basis for his paradox: humans as a species have been evolving visual and spatial skills for millions of years, but abstract thought only for about a hundred thousand or so.

OK, you might say. This is all very interesting, but doesn't it just prove the point that we need to change the way we teach and assess? If computers really are so brilliant at these typical academic skills that are taught in schools, maybe we should stop teaching them completely or only teach the particularly advanced, specialist and niche ones that computers can't do?

No. First of all, we will always want to teach academic skills for personal development. It's good to be able to read, write and count, even if a computer is faster and quicker. We didn't stop teaching PE because of the invention of the car or drawing because of the invention of the camera.

It is true that in order to develop, understand and use advanced technologies, humans are going to need advanced skills, and it is true that these more advanced skills should be one of the *ultimate* aims of education.

But this does not mean that we can forget about the more fundamental skills, because they are what allow us to develop the more advanced skills. This brings us to the second flaw in the 'set more complex assessments' argument.

2 If we want students to have advanced skills, they can't leapfrog fundamental skills

Of course, we want our students to develop the higher order skills of being able to critique writing produced by AI chatbots and to direct the outputs of new technologies. But those skills *depend* on more fundamental skills and there is no way we can jump ahead to the more advanced skills without acquiring the more basic skills first. In order for students to successfully grapple with problems computers *cannot* do, they must work through problems that computers *can* do. If schools could only teach maths using problems that computers cannot solve, we would have to teach six-year-olds maths using problems even top mathematicians find difficult!

So, it doesn't matter if we set our students tasks that can be easily solved by computers. It doesn't matter if they produce writing that is weaker than that of ChatGPT. The easy problems and the weak writing are milestones on their journey to mastery which cannot be skipped or outsourced.

The basic educational principle here is to do with the relationship between working memory and long-term memory. We have limited working memories, so we need to make up for that weakness by storing lots of information in long-term memory. You can't outsource that information to Google or ChatGPT: it needs to be in long-term memory so it can be effortlessly and frictionlessly summoned to working memory when needed, and combined with information in the environment, where it will produce what we typically call 'skill'. Here's a quotation from Daniel Willingham, Professor of Psychology at the University of Virginia, which expresses this well.

We will always want to teach academic skills for personal development. It's good to be able to read, write and count, even if a computer is faster and quicker.

[1] Of course computers are getting better at perception and mobility – but Moravec's paradox still holds in that the computers are requiring a lot more computing power to achieve them.

> *Data from the last 40 years lead to a conclusion that is not scientifically challengeable: thinking well requires knowing facts, and that's true not simply because you need something to think about. The very processes that teachers care about most – critical thinking processes such as reasoning and problem solving – are intimately intertwined with factual knowledge that is stored in long-term memory (not just found in the environment). (*Willingham, 2021*, p. 28)*

There is an analogy here with chess. Chess computers have been better than the very best humans at chess for decades now. What do we do if a child wants to learn chess? Do we say, well, there is no value in teaching or assessing any content that a chess machine can do? Do we say, we need to set them problems that Alpha Go cannot solve? Of course not! We teach them how the pieces move, what the basic openings are, and some of the common patterns to look out for, even though computers find all these tasks trivially easy. Interestingly, we can and do use technology to help students *acquire* these basics, but at no point do we assume technology means they *never have to learn* those basics. The same is true of other skills.

3 The point of an assessment is not the product but the process

The value of the work students produce in an assessment is not in the work itself but in understanding it represents and the thinking that went into creating it. Imagine two students write an essay. One struggles hard, reads a lot, writes and redrafts it, and produces something that is OK but not great. Another produces something perfect by pasting the prompt into ChatGPT and copying the output into a Word doc. Who has done better? If all we cared about was the product, then it would be the second student. But we don't. We care about the process. The first student's response has led to them learning more and their essay represents greater understanding of the topic than the second student's essay. Fundamentally, it does not matter that a computer can answer this question better than the student. What matters is what the student has learned from answering it, and what it tells us about their understanding.

The basic assessment principle here is the difference between the sample and the domain. The sample is the test itself. The domain is the student's wider understanding. The sample only matters if it tells us something valuable about the domain – otherwise it is worthless. This might seem like a fairly straightforward distinction, but even before ChatGPT it was widely misunderstood. Here's what Daniel Koretz, Professor of Educational Assessment at Harvard, has to say about it.

> *This might be called the sampling principle of testing: test scores reflect a small sample of behaviour and are valuable only insofar as they support conclusions about the larger domains of interest. This is perhaps the most fundamental principle of achievement testing. A failure to grasp this principle is at the root of widespread misunderstandings of test scores. (*Koretz, 2008*, pp. 21–22)*

The reaction to ChatGPT bears out Koretz's point about how poorly understood this principle is. I also think this misunderstanding could cause real problems with student motivation. If a student struggles for an hour over an extended piece of writing and then finds that a computer has surpassed it in seconds, it is entirely possible they will feel demotivated. What they need to hear from adults is "Don't worry, your work is of value, you're on a journey and you are developing your own writing skills." What they don't need to hear is "Well there's no point in even bothering, the computer is so much better than you. Try this assessment which is even harder instead!"

So, if we are setting assessments that a robot can complete, what does that say about our assessments? It doesn't tell us very much at all. Maybe it's a good assessment, maybe it's not. Whether a robot can complete it or not is largely irrelevant when judging its quality.

The one way in which the answer to this question will matter is in terms of the conditions that the assessment should be taken in, something I'll consider in future articles.

This article originally appeared on the No More Marking Substack. *Subscribe for more updates:* https://substack.nomoremarking.com/subscribe.

Daisy Christodoulou [@daisychristo on X] is Director of Education at No More Marking, *a provider of online Comparative Judgement software for schools. She is also the author of three books:* Teachers vs Tech, Making Good Progress?, *and* Seven Myths about Education. *In 2019 Daisy was recognised in the New Year's Honours List with an MBE for services to education.*

When tests get old: A simple model for rehabilitation

Nicola Bell

Kevin Wheldall

What happens to tests that no longer produce accurate scores?

Background

Classroom-based assessments of reading have many purposes, one of which is to indicate to educators how a student's decoding skills compare with those of their peers. To gather such information, teachers and literacy specialists need tests that:

- are norm-referenced (i.e. they have norms that allow for a given score to be compared with a reference group of students in the same year level or age group)
- provide an accurate indication of ability (i.e. they measure the underlying construct of interest and provide a score that estimates the student's level of skill well)
- measure nonword reading performance (i.e. they contain items that must be decoded using knowledge of letter–sound correspondences).

At MultiLit, a test that we often use to assess nonword reading accuracy is the 'Martin and Pratt Nonword Reading Test'. In fact, the Martin and Pratt has been a staple within our program trial test batteries over many years, such that, even when it was no longer available for purchase, the MultiLit Research Unit (MRU) sought permission to continue using the materials.

In 2021, the MRU decided to embark on a 'check norming' study of the Martin and Pratt, since, at that point, 25 years had passed since the original test norms were collected. The aim was to confirm whether the assessment was valid and had standardised norms that accurately represented students' decoding skills. Specifically, the research questions of interest were:

1 How valid is the Martin and Pratt as a measure of nonword reading proficiency?

2 How well do Martin and Pratt standardised scores estimate primary school-aged students' nonword reading proficiency?

The two research questions sound alike but involve different methodologies and analyses.

The first question relates to *validity* – that is, the degree to which the test measures the underlying skill it is purported to measure. This is typically examined by analysing the correlations between the test of interest and other similar tests. The test may be said to have validity if the scores derived from it correlate strongly with other similar measures.

The second question relates to *norm representativeness*. Even if a test is valid, that doesn't mean its norms are up-to-date or representative. It may still measure what you want to measure but spit out a score that over- or

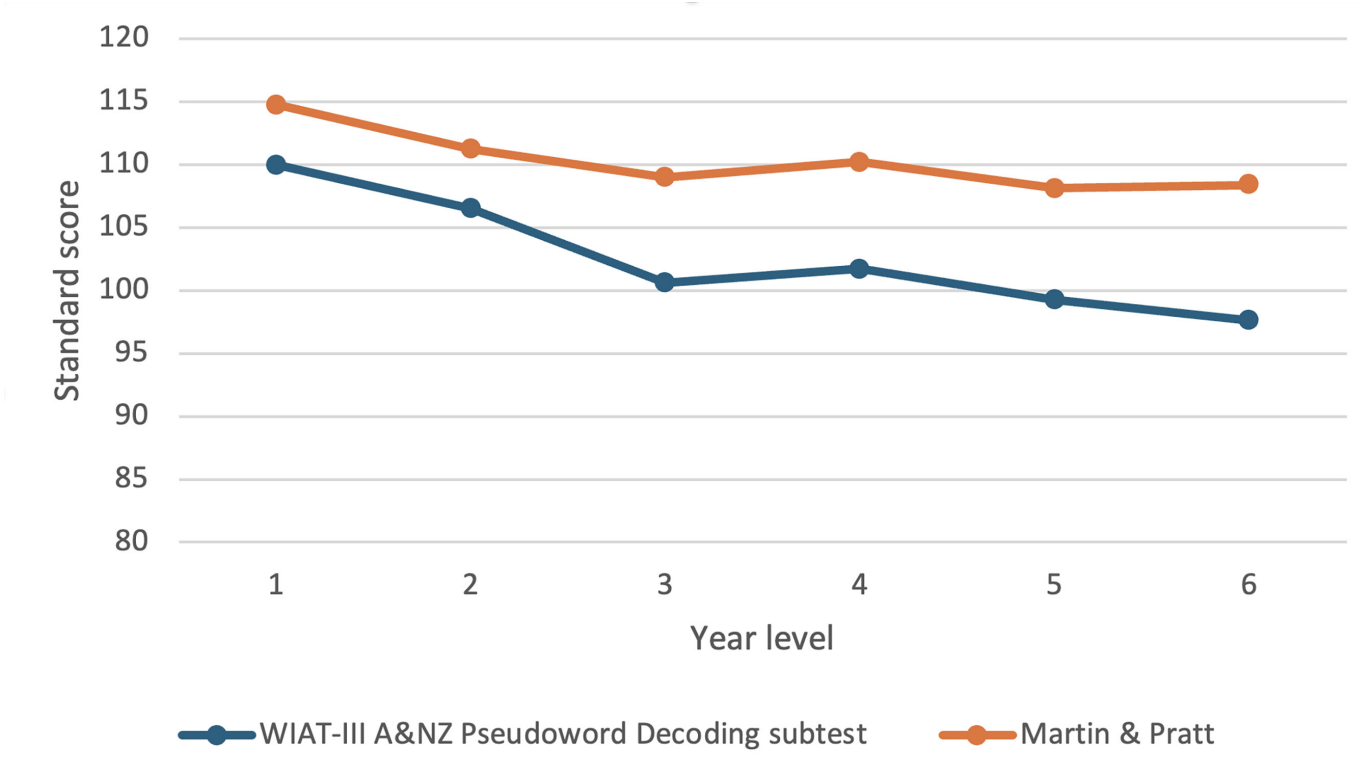


Figure 1. Mean standard scores achieved by each year level on the M&P and WIAT-III A&NZ PD.

underestimates the examinee's actual skill in that area.

Data collection

The process for conducting the study started with obtaining the prerequisite ethics and research approvals. Getting institutional ethics approval from Macquarie University was a straightforward enough process. However, we did not receive approval from the Queensland Department of Education, which meant we could not recruit any government schools in Queensland into the study (and we could not assess any students within the grounds of government schools that approached us). As such, we only approached independent schools in the greater Brisbane area with information about the study, and we were fortunate to have three sign on to participate. In terms of sample size, we got close to the anticipated number of students – 176 in total, with close to 30 in each year level. The student populations at all schools had average socio-educational backgrounds and produced approximately average NAPLAN Reading results.

In addition to the Martin and Pratt, students were assessed on two other nonword reading accuracy measures: the Castles and Coltheart 2 (CC2) and the Wechsler Individual Achievement Test 3rd ed Australian & New Zealand Pseudoword Decoding (WIAT-III A&NZ PD) subtest. Other aspects of reading proficiency were also assessed, such as fluency and comprehension. Each assessment session lasted approximately an hour. One key correspondent from each school kept our testing team informed about the best days and times for withdrawing students from their classes for testing.

Results from the study

The results from our analyses showed that Martin and Pratt scores correlated significantly with all other reading measures. Importantly, the Martin and Pratt was most strongly correlated with the other two measures of nonword reading accuracy: CC2 Nonwords (r = .92) and WIAT-III A&NZ PD subtest (r = .91).[1] These findings provide good evidence for the Martin and Pratt's validity. As such, the test's raw scores can still reliably be used to measure a student's progress.

However, the *standard* scores derived from the Martin and Pratt were consistently higher than those derived from the other two nonword reading accuracy assessments (on average by 7 standard score points). This means that, despite the test's solid design and observed validity, the standard score and age equivalent values that the norms

These findings provide good evidence for the Martin and Pratt's validity.

[1] Generally, r-values above .80 are considered to represent a 'strong' relationship.

allow users to compute, significantly over-estimate students' skills. Figure 1 shows the mean standard score for each of the Martin and Pratt and WIAT-III A&NZ PD across year levels. The gap between scores that *should* assess the same underlying skills is obvious.

Recalibration of Martin and Pratt norms

Having found that the test norms over-estimated students' skills, we decided to try and recalibrate them. This involved a novel analytical process, which was based on the assumption that the WIAT-III A&NZ PD standard score represented students' *actual* level of nonword reading proficiency. (The WIAT-III A&NZ was selected for this purpose because its normative data were collected quite recently from students across all Australian states.) Using a few different techniques, we sought to close the gap between those WIAT-III A&NZ PD scores and the Martin and Pratt scores.

The technique we landed on involved conducting a regression analysis with Martin and Pratt standard scores, WIAT-III A&NZ PD standard scores and age as variables. The resulting equation from that analysis was used to update all values in the original norms table to 'recalibrated' values.

There are a couple of important differences between our recalibrated norms and the original ones. Firstly, we only used Form A of the test, whereas the original manual contains norms for both Form A and Form B. Secondly, the recalibrated norms extend only to 11 years, 11 months, whereas the original norms extend all the way to 16 years, 11 months. We intentionally limited the age range of our sample in this way because we have doubts about the meaningfulness of assessing nonword reading in typically developing students beyond the primary years.

Final thoughts

The study described here provided the impetus to attempt a 'rehabilitation' of the Martin and Pratt. Happily, and with the generous support of the original test authors, Frances Martin and Chris Pratt, preparations are now underway to republish the test alongside the recalibrated norms.

But one question worth asking is: Why were the Martin and Pratt norms no longer accurate in the first place? Obviously, a considerable amount of time has passed since data for the original test norms were collected, but what exactly happened during that time to change students' nonword reading ability at a population level?

We think the most likely answer is that Australian teacher knowledge around the importance of phonics has increased. More broadly, reading skills in general – as measured in primary school-aged students – have improved, according to national and international testing.

Based on those observed improvements alone, test developers should think more about how norms can be updated in response to widespread shifts in instructional practices – particularly when those practices have such a direct relationship with the skills being assessed (e.g. phonics instruction and nonword reading proficiency; see article by Shanahan in this issue of *Nomanis*). Perhaps the recalibration undertaken as part of our study could be useful as a model for 'rehabilitating' other assessments that are outdated.

Test developers should think more about how norms can be updated in response to widespread shifts in instructional practices – particularly when those practices have such a direct relationship with the skills being assessed.

This article is an edited version of a presentation delivered at the DSF Language, Literacy & Learning 2024 conference.

Nicola Bell [@NicolaBellSP on X] is a Research Fellow in the MultiLit Research Unit.

Kevin Wheldall AM [@KevinWheldall on X] is Chairman and Director of MultiLit and Emeritus Professor of Macquarie University.

Has cognitive load theory been dealt a devastating blow?

Greg Ashman

Alfie Kohn has taken recent aim at cognitive load theory, but a closer inspection of his arguments shows they may not hold much water.

Alfie Kohn is an unapologetic supporter of progressive education, that long educational tradition that has promised so much and delivered so little. While others sympathetic to the cause play word games and smuggle their ideas into practice under a bewildering variety of ever-changing names, Kohn has remained admirably willing to hoist the standard, unapologetically. Now, in a new blog post with many footnotes, he has taken aim at cognitive load theory and the research of John Sweller.

It is interesting to speculate on why Kohn has chosen to do this. American education academics and pundits have mostly ignored cognitive load theory. This may be because it is not an American creation, even if a remarkably similar theory – the cognitive theory of multimedia learning – is. Perhaps Kohn's attempt at a rebuttal of cognitive load theory indicates that it is starting to percolate through to American education's collective consciousness and that's why it is time for progressive educators to knock it down.

It is tempting to try to chase down all of Kohn's references. I won't do this because I don't really want to, and it would result in a long blog post with an obsessive tone. Instead, I will look at some examples and wider points.

Kohn makes a pre-emptive strike against accusations of 'cherry-picking' – the widely wielded academic criticism that a writer has selected only the sources that support their contention and ignored the ones that do not:

> *I've cited several metaanalyses and other research reviews in the extensive endnotes to this essay precisely so that sceptics can't claim that I've cherry-picked unrepresentative studies to make the case in favour of what is sometimes called progressive education.*

I tend to find accusations of cherry-picking a little tedious. If a writer has missed important sources, point it out – something I now ironically intend to do to Kohn's piece.

For example, Kohn completely ignores the widespread evidence from PISA that inquiry learning is associated with worse learning outcomes. Why is this important? Well, there are three broad types of study that can potentially answer the question of which teaching method works best – small educational psychology experiments, large education experiments and correlational studies. Typically, researchers only consider large education experiments, but these are often the most confounded. They tend to vary more than one factor at a time and often compare a cool new intervention with business as usual.

To control for the expectations of subjects in a study, we should either compare interventions with interventions or business as usual with business as usual. Large correlational studies like PISA do the latter and so are an important part of the picture. This is a key component of the argument of one of the papers Kohn cites but he never mentions this. He does mention a rebuttal of

Guidance is important, but so is student involvement. The trouble is that these two do not operate on a continuum.

that paper but he does not mention the response to the rebuttal. This is all highly selective.

What does Kohn focus on? Well, to bolster his point that explicit instruction is less effective than "some variant of student-centred learning", from early years to college, he cites *two* sources. The first is a Slate article by Alison Gopnik that relates one of her famous experiments with toys: A 'teacher' taught children how to use a toy whereas another researcher showed children the toy but expressed surprise at what it did. Children in the second group were more likely to discover features of the toy that had not been demonstrated.

What does this prove? Not much. We are talking about toys that are designed to be fun and not maths or reading. And it seems reasonable for the children to assume the 'teacher' had shown all the features and not looked for more. It certainly does not prove that direct instruction somehow destroys creativity.

The second source pits 'active learning' against 'traditional lecturing' and summarises the findings of 225 studies. The active learning conditions involved students who listened to lectures but also completed tasks such as worksheets, discussed ideas with a partner or responded to multiple-choice prompts via clickers. This group did better than those who just listened to the lectures.

I am not surprised by this finding. Barring the worksheets, the listed activities in the active learning condition sound like rudimentary versions of the activities described in Explicit Direct Instruction by Hollingsworth and Ybarra. The kind of explicit teaching advocated for by proponents is highly interactive. If you walked around my school during lessons, you would see students answering a question on a mini whiteboard, giving a thumbs up or thumbs down or turning and talking to their partner every couple of minutes or so.

However, Kohn thinks this argument is unfair.

> *On the one hand, [proponents of explicit teaching are] apt to set up inquiry learning for failure by using a caricatured version of it, a kind of pure discovery rarely found in real-world classrooms, with teachers providing no guidance at all so that students are left to their own devices. On the other hand, the version of DI [explicit teaching] they test sometimes sneaks in a fair amount of active student involvement – to the point that the two conditions may just amount to different forms of constructivist instruction.* [references removed]

There is something approaching a point here. Surely, suggests the pragmatist, we should be looking for a compromise between the two extremes. Guidance is important, but so is student involvement. The trouble is that these two do not operate on a continuum. One is a key distinction between inquiry and explicit teaching, whereas the other is not.

If guidance is important, why not agree to have lots of it, at least when learning new things? What about *full* guidance? And if student involvement is important, why not have lots of that, too? Why not have as much as possible of both? This then becomes the kind of explicit teaching proponents advocate.

Wait, what have I done there? Is it a trick? No. The defining feature of explicit teaching is that concepts are fully explained and procedures are fully demonstrated before we ask novices to apply those concepts or procedures. You either do that or you do not. You cannot do both. Inquiry learning requires students to figure something out for themselves, so it is necessarily antagonistic to full guidance – an antagonism that is present throughout Kohn's piece.

However, the definition of explicit teaching says nothing about the amount of student interaction. Unlike inquiry learning and guidance, we can max this out without stopping it being explicit teaching. This is a distinction that Kohn misses.

A key finding of cognitive load theory is that the effectiveness of full guidance does not apply to relative experts. This is something known as the 'expertise reversal effect'. Relative experts already have relevant schemas in long-term memory to draw upon, so they need more practice solving different problem types.

Kohn quotes a 2007 article by Schnotz and Kurschner that is critical of the cognitive load theory of the time to demonstrate that "Reducing cognitive load isn't always desirable … That's

because 'learning can be impeded … when too much help is provided.'" Kohn does not make clear that this quote relates to *relative experts* and certain uses of animations. Here's the full context of the related footnote:

> *This is demonstrated, for example, by the expertise reversal effect, when performance aids (such as worked-out examples) turn out to be disadvantageous for individuals with higher expertise, or when animations prevent learners from running their own mental simulations (cf. Kalguya et al., 1998, 2003; Schnotz and Rasch 2005). Aids are then beneficial for task performance, but not for learning. In other words: Making a task easier does not necessarily result in better learning.*

Cognitive load theory *predicts* the expertise reversal effect. The animation effect seems more complex and it's not obvious to me whether it aligns with the predictions of the theory.

For the final citation rabbit hole, let's find the source for the quote that for reading instruction, "The more a teacher was coded as telling children information, the less [they] grew in reading achievement."

It comes from a process-product study by Michael Rodriguez. 'Telling' is an odd word and sounds like a pejorative description of explicit teaching. We can find out what Rodriguez means by it by looking at an example:

> *During making words activities, the children manipulated their own set of letters as Ginger [the teacher] coached:*
>
> Let's do tub. Listen to the middle sound. It's not tab, it's not tob. It's / ttt-uuu-bbb/. You need a letter for /uuu/.
>
> *While reading leveled books, students tracked with their fingers as they read independently from their own copies. If they got stuck on a word, Ginger coached by providing hints instead of telling them the word.*

So, 'telling' is telling a student what a word is rather than asking them to sound it out. Ginger's teaching seems pretty explicit to me.

Kohn has two criticisms that do land. He doesn't like the separate type of load known as 'germane load'. This makes cognitive load theory unfalsifiable. John Sweller agrees, which is why he has stopped classifying it as a separate kind of load. And this leads to a second criticism – that when their predictions are proved wrong, cognitive load theory researchers review and change the theory. Which doesn't sound like a bad thing to do when you write it down like that.

This only seems like a criticism if we assume that, to be credible, cognitive load theory needs to be some kind of timeless, revealed truth and not a messy, real-world theory still in the process of being developed. Only time will tell whether its adaptions in the face of disconfirming evidence make it more robust or, like the Ptolemaic system's epicycles, are a sign of a need for fundamental revision. At this stage, pundits can take their pick.

I've probably already spent too much time on this. Most of those familiar with cognitive load theory will not be convinced by Kohn's post. Its audience is more likely to be those who are based in the US and have had cognitive load theory cited as evidence against their preferred teaching methods. These folks can then post Kohn's piece into their replies, avoid having to think too much about it and get on with their day.

However, I will just add a short coda before we move on. It was interesting to see a couple of blog posts by Sue Gerrard from 2014 cited by Kohn. It took me back to the heyday of education blogging and some of the to-and-fro of the time. Gerrard's citations sit in a note about 'CLT's simplified view of cognition' that also cites David Jonassen's chapter from *Constructivist Instruction: Success or Failure*? This book is unusual in that it is framed as a debate and allows the opposing side to ask questions at the end of each chapter to the chapter author. It is therefore a great opportunity to post this comment from John Sweller at the end of Jonassen's chapter:

> *I asked whether there was any evidence from randomised controlled experiments indicating that the cognitive distinctions you make have instructional implications. The answer presented is unambiguously 'no', an answer I agree with. You go on to suggest that lack of evidence from randomised, controlled experiments is unimportant because such experiments are themselves unimportant or perhaps impossible, based on atomic physics. We'll have to agree to disagree on that, but there are serious consequences of this position.*
>
> *Is there any technique that could be used to provide evidence that constructivist teaching is a relatively poor method of teaching?*

It's a good question and one I would be interested in Kohn answering.

This article originally appeared on the author's blog, Filling the Pail.

Greg Ashman [@greg_ashman on X] is Deputy Principal at Ballarat Clarendon College, Victoria. He is a prolific blogger and has written three books: The truth about teaching: An evidence-informed guide for new teachers, The power of explicit teaching and direct instruction, *and* A little guide for teachers: Cognitive load theory. *Greg is an honorary fellow at the Australian Centre for the Advancement of Literacy, Australian Catholic University, and a part-time professor at Academica University of Applied Sciences in Amsterdam. Prior to moving to Australia, Greg worked at several comprehensive schools in London.*

Reading Pledge

The Reading Pledge is an evidence-based framework for schools to reduce the number of children who finish primary school unable to read proficiently. The entire Reading Pledge publication can be accessed via the link at the end of this excerpt.

Pledge
To reduce to near zero the number of children who finish primary school unable to read, or who struggle with reading in secondary school, by providing both primary and secondary school teachers with the training and resources they need to deliver targeted assessments and effectively address the needs of those students who are struggling, through the provision of effective intervention.

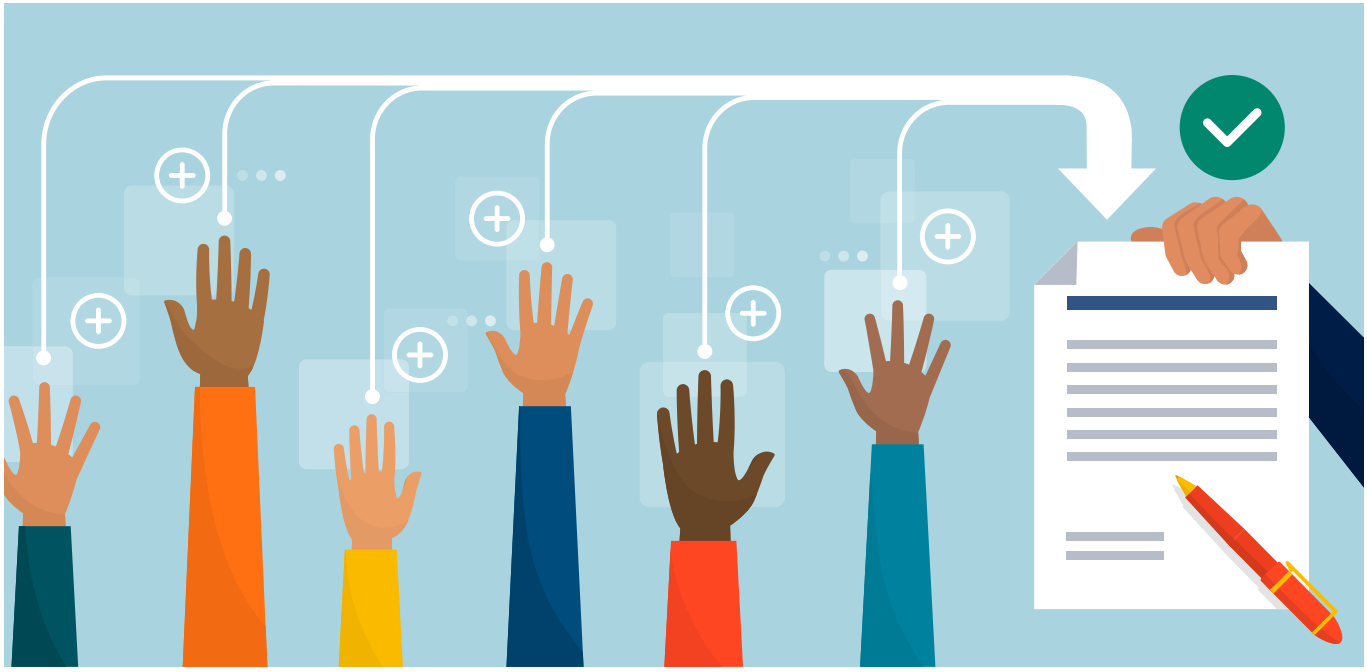
Rationale
Every year since the National Assessment Program for Literacy and Numeracy (NAPLAN) was implemented in 2008, a substantial number of students have not met the literacy standards necessary to make good progress in education. In 2023, almost 90,000 Year 7 students were placed in the lowest two standards, indicating that they did not meet the 'Strong' proficiency standard of 'challenging but reasonable expectations' in reading. Of these, close to 27,000 were in the lowest proficiency standard and identified as 'needs additional support'. There are too many students leaving primary school not meeting proficiency standards in reading.

Action
Every child who does not meet the designated achievement benchmark in the Year 1 Phonics Check (or similar assessment) or the NAPLAN Reading assessments is referred for standardised reading assessments and, based on those results, provided with appropriate evidence-informed interventions.

Background to the Reading Pledge
The Primary Reading Pledge published in 2020 highlighted the number of students who were unable to read well after seven years of primary school and that these students should have been provided with support much earlier in their education. It provided an evidence-based framework for schools and systems to use to reduce this number. The Primary Reading Pledge said:

> *Thousands of children each year are being denied this basic right, most of whom are casualties of a system that has become accustomed to an unacceptable rate of failure.*

Many schools adopted the Pledge and have been working towards a goal of 100% literacy. This updated version, called the Reading Pledge, also extends the framework and guidance on intervention and assessment to secondary schools.



The Reading Pledge is both a call to action for all those involved in education, and a practical and useful tool for schools. It once again draws on the combined expertise and experience of two organisations (Five from Five and Learning Difficulties Australia) who have been supporting teachers to help struggling readers for, in some cases, decades.

NAPLAN results for previous cohorts have shown that every year, similar numbers of students begin their secondary education without the necessary literacy skills to enable them to succeed in education and beyond.

Nor is it the case that these students are identified for the first time in Year 7. Analysis of longitudinal data has shown that 72% of students identified as struggling readers in NAPLAN Year 3 were still struggling in Year 5, and 88% of students identified as struggling readers in Year 7 were still struggling in Year 9 (Productivity Commission, 2022).

The NAPLAN assessments provide an opportunity to identify every child in need of support and for a systematic response to be implemented. Currently, there is little guidance and support for schools to respond to NAPLAN results for low achieving students.

Response to Intervention (RtI) is a tiered model of instruction and intervention for students experiencing difficulties in acquiring basic skills and appropriate social behaviours (Fletcher & Vaughn, 2009). The goal of RtI is to provide screening and/or assessment, deliver effective intervention, monitor student progress and then use the students' responses to the intervention provided to determine 'next steps'.

RtI typically has three 'tiers' of instruction and intervention. With initial whole-class reading instruction based on evidence-based best practice (Tier 1), most students will get off to a good start in learning to read. Those students who begin to fall behind, often operationally defined as those in the bottom 25% of what might be expected for the age cohort, are then offered Tier 2 instruction.

Tier 3 intervention is even more intensive, tailored to the specific needs of the individual student, and preferably provided by a reading expert. If RtI is implemented well, only a very small number of children are likely to require this level of support on a continuing basis, but they may need it for several years.

Students should receive the instruction and intervention they need

The best setting for students to learn to read is primary school. Ideally, all students will receive exemplary Tier 1 reading instruction. This should be the expectation for all schools. Application of the RtI model will identify students who need additional support and provide intervention early. This will result in fewer students progressing to secondary school without adequate reading skills.

Once students reach secondary school, it is much more difficult to catch them up for several reasons. First, the skills gap is often very wide so it can

The Reading Pledge is both a call to action for all those involved in education, and a practical and useful tool for schools.

take a long time for them to reach the level of their peers (Colleu Terradas, 2023). Second, many students have developed anxiety or low self-concept around reading, well before they reach adolescence (McArthur, 2022). Third, finding time in the secondary school timetable for intervention is challenging. Finally, few secondary schools have teachers with specialist literacy skills (de Haan, 2021).

The costs associated with intervention on this scale are considerable but not unrealistic, especially if managed efficiently at a systemic or sector level. The costs of not intervening through intergenerational impacts on employment, income, health, welfare and crime are far greater. It has been estimated that illiteracy costs the economy up to $44 billion each year. With adequate investment to ensure appropriate interventions reach the students who need it and evidence-based reading instruction accessible to all Australian students, it is estimated that a 13-fold return on this investment is possible (Del Rio & Jones, 2023).

While NAPLAN can and should be used as a source of information to identify students who need intervention in Years 3 and 5, evidence-based intervention should be provided in schools as a matter of course much earlier than Year 3. All students in Foundation to Year 2 should be given valid and reliable screening and progress monitoring assessments in reading subskills, including the Year 1 Phonics Screening Check. This is a recommendation of the expert panel report as well as a recent report from the Grattan Institute (Hunter et al., 2023; O'Brien et al., 2023).

The Year 1 Phonics Screening Check is already being used or will be used as a systemic assessment in South Australia, New South Wales, Tasmania and Queensland. Data from South Australia in 2023 show that 71% of Year 1 students achieved the benchmark score of 28/40, up from 43% in 2018 (Government of South Australia, 2018, 2023). This indicates that significant improvement has already occurred in South Australia, but further improvements are required to classroom reading instruction (Tier 1 in the RtI model). In New South Wales, data for 2023 showed 59% of Year 1 students met the benchmark, an increase of two percentage points since 2021 (NSW Government, 2023). Tasmania's results are not yet available, and Queensland will implement the assessment from 2024.

Provision of early intervention can represent a significant investment, but effective intervention at this stage will reduce the number of children requiring intervention in Year 3 and Year 5, at which stage their difficulties will be harder and more expensive to remediate.

Numerous reading interventions are available and are currently being used by schools. Almost all schools offer reading support in some form. However, reading intervention is not consistently evidence-based and targeted, and is often limited to the first few years of school – sometimes due to lack of knowledge of evidence-based intervention and sometimes due to lack of resources and support. Children who continue to struggle with reading after receiving some (but not enough) early support will be among the children identified as 'needing additional support' or 'developing' in NAPLAN.

Ideally, government and system policy would enable and facilitate the provision of evidence-based supports for struggling readers, but schools do not have to wait for this to happen. There are actions that schools can take using existing resources and processes. The Reading Pledge provides a framework and lists of valid assessments and evidence-based and evidence-informed interventions.

The entire Reading Pledge can be downloaded for free from https://fivefromfive.com.au/reading-pledge-2024/.

The best setting for students to learn to read is primary school.

Can children be taught to comprehend what they read?

Daniel Willingham

Some simple comprehension strategies need only be taught for a short time. Others are more advanced and may require continued practice to yield deeper reading comprehension.

Just how much does it help to teach children to use strategies when they read – strategies like creating a graphic organiser of the passage, or summarising as they read, or asking themselves questions and answering them?

I've just published an article in *Educational Leadership* summarising the research on this question, and I'll summarise it here.

In 2006, I argued that there was lots of evidence that comprehension strategy instruction worked, and in fact, yielded a big boost to comprehension. I was in good company– The National Reading Panel had drawn the same conclusion five years earlier.

But I also argued that there was no evidence that practise of these strategies provided any additional benefit. I based that conclusion on two meta-analyses – research that synthesises the results of different studies. Meta-analysis allows one to compare relatively brief exposure to strategy instruction (a total of, say, five hours) versus more practice with strategies (20 hours). Both meta-analyses suggested that there was no benefit to more practice.

There's been a good deal of research since then. In my recent article, I report that the number of meta-analyses is now up to 12, and all are in accord. Practice has no impact on the effectiveness of comprehension strategy instruction.

That observation matters for two reasons. First, and most obviously, it suggests that although it's well worth the time to teach students comprehension strategies, there's no reason to devote a lot of time to practising them. A total of five or 10 hours of instruction yields the same advantage as 20 or 30 hours.

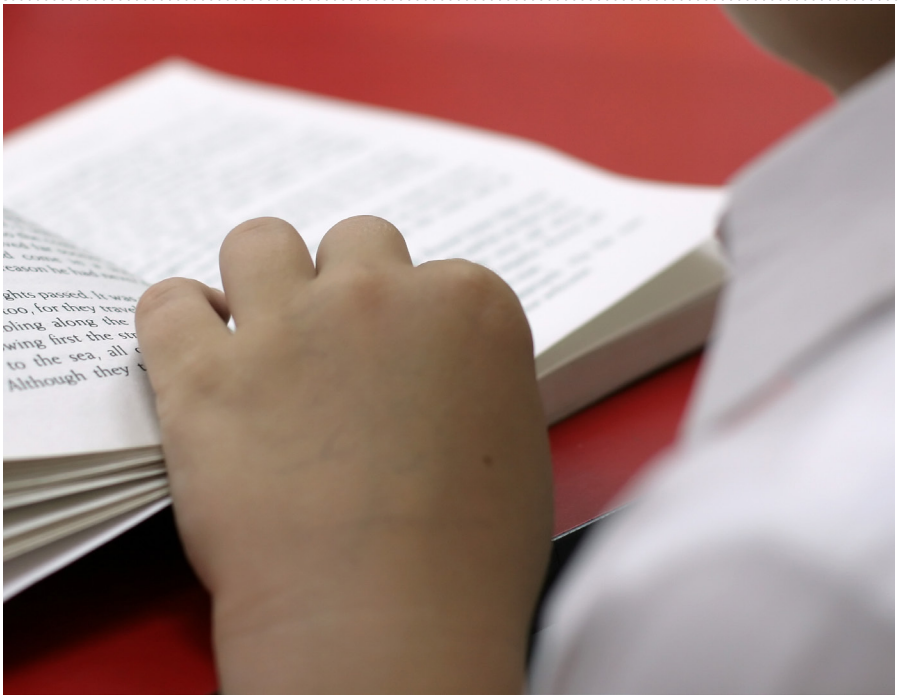
Second, this finding suggests that strategy instruction works for a different reason than I suspect many people believe.

It's tempting to think of comprehension strategy instruction as analogous to coaching in baseball. If you're a poor hitter, a coach shows you how a good hitter swings. You practise that swing and, in time, it becomes automatic and replaces the older, less effective habit. Likewise, we might think that comprehension strategies show less competent readers the way that more competent readers approach texts.

But this hypothesised 'coaching' mechanism doesn't make any sense because it depends on practice, and the data indicate that practice doesn't help.

Here's an alternative interpretation. When a typically developing child starts school, they can use oral language to make inferences, connect sentences and understand the overall gist of a message. These same mental processes are put to work to support reading comprehension. Indeed, it would be odd if the brain created specialised reading comprehension processes from scratch, rather than applying to reading the mental processes that are already in place to support oral language.

The mental processes of reading comprehension don't require or benefit from practice because children are already quite good at them when they start school.



Strategy instruction is comparable to a strategy like 'check your work' in maths. It doesn't improve the processes that actually do maths. It's a useful way of controlling those processes.

According to this account, strategy instruction is comparable to a strategy like 'check your work' in maths. It doesn't improve the processes that actually do maths. It's a useful way of controlling those processes.

In the same way, comprehension strategy instruction probably has no impact on the processes of comprehension per se, but it reminds students that they are *supposed* to coordinate meaning across sentences and paragraphs, and to get the gist of the passage. In short, it reminds them that reading is not simply a matter of decoding each word until you reach the last one.

But that's not quite the end of the story.

My description of comprehension strategy instruction could be interpreted as implying that reading instruction should end around fourth grade. Schooling should include phonics instruction, some work to support fluency, and then perhaps two weeks of comprehension strategy instruction. What's the point of anything else if comprehension can't be taught? (I hadn't thought of this implication of my account until Tim Shanahan pointed it out.)

Surely that implication can't be right. Explaining why calls for differentiating types of comprehension.

I've suggested that strategies prompt children to apply already present oral language comprehension processes.

An example would be anaphora resolution, as when a listener finds the referent for 'he' in 'he went to church'. Another example would be inferences supporting causality or explanation; seeking to understand why things happened seems to be a core aspect of cognition. And indeed, we know a four-year-old has no difficulty in making causal bridging inferences in everyday conversation, as when a parent says, "You seem bored. Shall we go outside?"

Exactly what prompts inferences in oral language or reading has been difficult to pin down, and there are surely individual differences. I think it's uncontroversial that the two examples I've offered are universal.

It's also uncontroversial that students are asked to do things with texts that go beyond comprehension supported by oral language processes. They learn sophisticated ways of evaluating arguments; for example, to appreciate that correlation is not equivalent to causation. They learn to evaluate the quality of writing, as when they come to understand how a good paragraph is structured. They also learn tools of analysis that are discipline-specific: why a novelist uses foreshadowing, for example, or how to interpret source information when reading historical documents.

Clearly, these skills must be taught, and there is every reason to think that they are subject to practice effects.

So we should differentiate kinds of comprehension. Some comprehension is supported by processes initially acquired for oral language, and presumably these processes yield a fairly basic understanding of the who, what, where, why, and how of the text. Other comprehension processes offer more sophisticated analysis, and these need to be explicitly taught.

An implication of this hypothesis is that the comprehension tests used in strategy research lean heavily on the first type of process; comprehension tests demand a basic understanding, not a more complex analysis. That prediction has not been tested, so far as I know.

I've long argued for the critical importance of knowledge in reading comprehension, but knowledge isn't everything – teaching students certain types of analysis is critical as well. Understanding how each applies to instruction can help us maximise student enjoyment of, and achievement in, reading.

This article originally appeared on the author's blog, Science and Education.

Daniel Willingham [@DTWillingham on X] is a Professor of Psychology at the University of Virginia, where he has taught since 1992. Until about 2000, his research focused solely on the brain basis of learning and memory. Today, all of his research concerns the application of cognitive psychology to K–16 education. He is the author of several books, including the best-selling Why Don't Students Like School?, *and most recently,* Outsmart Your Brain. *His writing on education has appeared in 23 languages. In 2017 he was appointed by President Obama to serve as a Member of the National Board for Education Sciences.*

Reading fluency and the Science of Reading

Nathaniel Hansford

Classroom-based reading fluency instruction takes many forms, but some methods are supported by more evidence than others.

What is fluency?

Reading fluency refers to the ability to read quickly, accurately and with prosody (smoothness and expression that reflects the meaning of the text). Fluency instruction is instruction that specifically tries to improve these skills. Some of the most common forms of fluency instruction include:

1. Choral reading: Have an entire class read the same text aloud, at the same time.
2. Varied reading: Created by the Iowa Reading Research Centre and is based on repeated reading. However, instead of reading the same text, they read a text that is 80% similar. This used to require the purchase of special varied reading texts, but with modern generative AI software like ChatGPT, this can be easily done with programs.
3. Readers theatre: Assign students characters from a play script and then have them rehearse the play.
4. Guided reading: Have students read, alone or in small groups, with a teacher to help with errors and unknown words.
5. Partnered reading: Partner students based on ability and have them take turns reading a text.
6. Silent reading: Have students practise reading independently.

Is fluency instruction part of the Science of Reading?

Yes! There is a large body of evidence suggesting fluency instruction benefits students. In fact, the National Reading Panel (2000) report, which in many ways was a founding research paper for the Science of Reading, listed it as one of the five pillars of reading instruction.

What types of fluency instruction work best?

In my opinion, there is probably a time and place for most types of fluency instruction. However, the strongest body of scientific research exists for repeated reading. Indeed, to the best of my knowledge, there have been three separate meta-analyses that looked at this topic.

Meta-analyses are important to rely on when evaluating efficacy because they show the mean result of all experimental research on a topic. (Meta-studies are studies of studies, which seek to systematically quantify the results of experimental research on a topic and synthesise those results into standardised metrics [typically effect sizes].)

Effect sizes are particularly useful because they allow us to compare the findings of different studies. Typically, effect sizes are interpreted as follows:

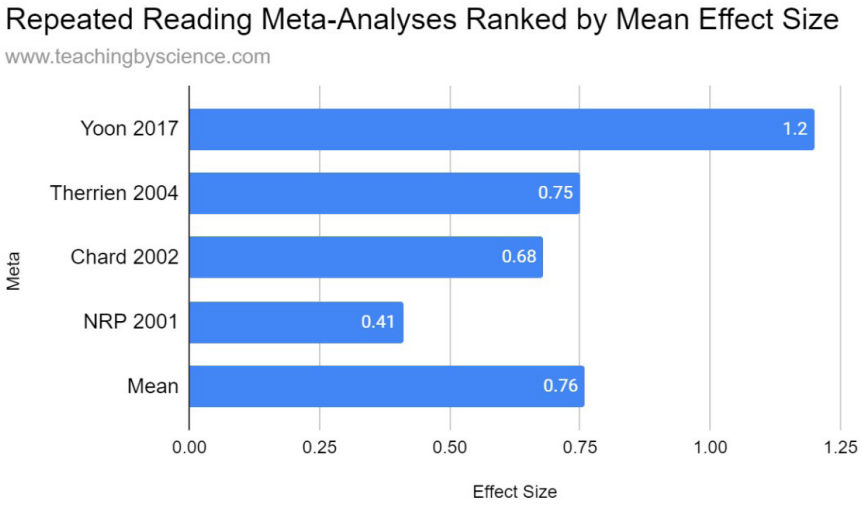


Figure 1.

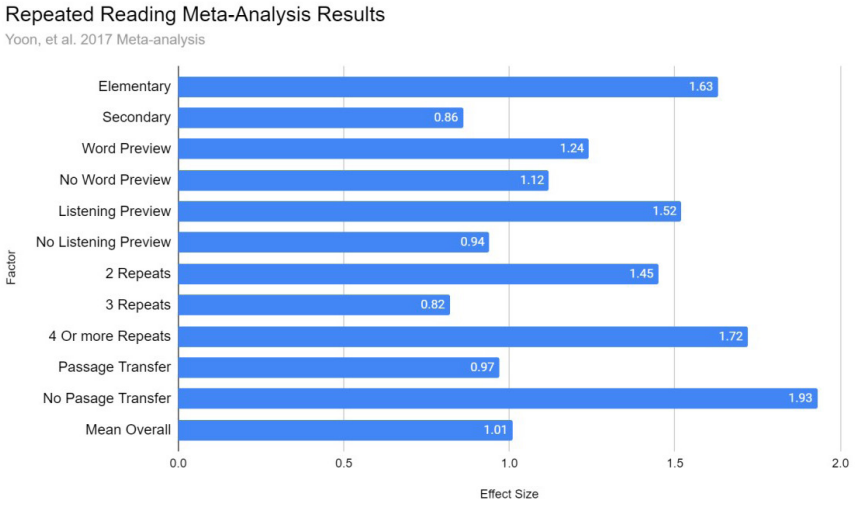


Figure 2.

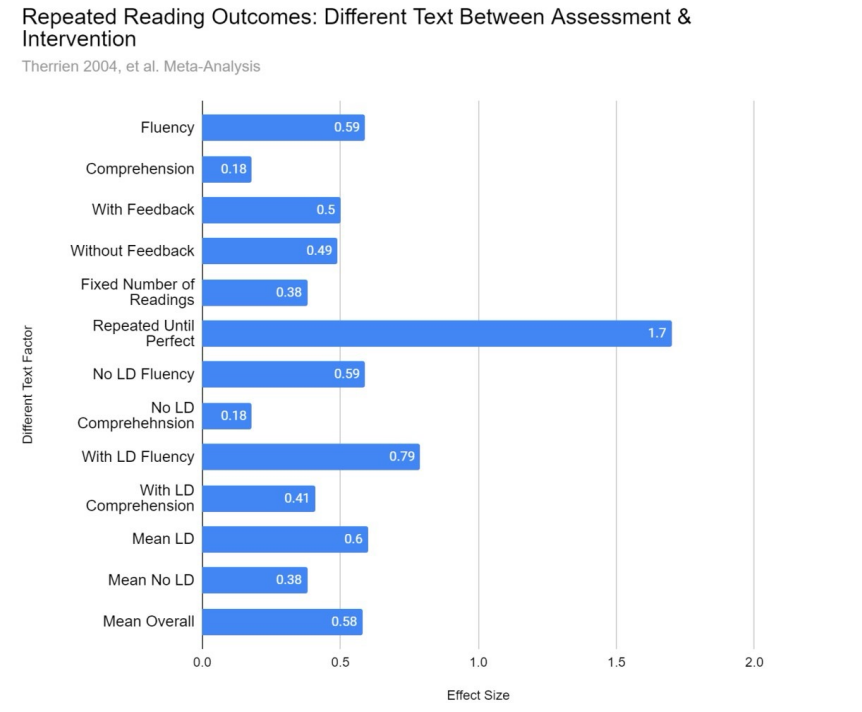


Figure 3.

below 0.20 is negligible, between 0.20 and 0.39 is small, between 0.40 and 0.79 is moderate, and over 0.80 is large. That said, effect sizes are typically lower in reading research. (In my opinion, an effect size above 0.40 should be considered large for reading instruction research.) The results of the three meta-analyses on repeated reading can be seen in Figure 1.

The mean effect size for meta-analyses on repeated reading is 0.76, which is quite large. To put the above results in context, the National Reading Panel found a mean effect size of 0.44 for systematic phonics and that finding has been used as the main scientific evidence for systematic phonics instruction ever since.

One common criticism of repeated reading is that it only improves fluency for the text read and not for new texts. In other words, many claim there is no transfer effect. However, this claim is verifiably false. Both Lee and Yoon (2017) and Therrien (2004) used meta-analytic methods to systematically examine the transfer effect of repeated reading and found strong transfer benefits, as can be seen in Figure 2 and 3.

Not only has scientific research suggested a strong benefit for repeated reading, it has been shown to be specifically beneficial for learning-disabled students (Therrien, 2004).

How best can we use repeated reading?

Personally, I like to do repeated reading for 5–10 minutes a day with my class. Too much more runs the risk of being boring due to the unavoidably repetitive nature. However, I like to embed my vocabulary and comprehension instruction within this fluency instruction. I typically read the text aloud first to the students. Next, I review any difficult vocabulary or background knowledge. Then I read the text chorally with my students, repeatedly, until I can hear that they sound fluent. I have the students read to perfection, because Therrien (2004) showed more than 4x the benefit for fluency outcomes when students had to read to perfection, compared with a fixed number of readings.

In my experience, using a fluid number of repetitions is superior, because it forces the students to actively

One common criticism of repeated reading is that it only improves fluency for the text read and not for new texts … However, this claim is verifiably false.

participate in the process. Once my students have completed the repeated reading, I typically ask comprehension questions and discuss the content.

What type of text should I use?

I have not found meaningful research on this topic. However, I do see unique benefits for both poetry and cross-curricular texts. Poetry can be great for building prosody because there is a natural rhythm; it helps students to learn to read with intonation and expression. However, lately I have been using cross-curricular texts, because it helps me to review curriculum material for other subjects and theoretically provides a long-term comprehension benefit (Hansford, 2023).

When selecting a text, the only thing that really matters is that it is appropriately challenging. That does not mean we need to use a benchmark assessment to find the right instructional level, as such assessments are typically not valid (Burns et al., 2015). However, in my opinion, we do want the text to be difficult enough that they might need to sound out some words or ask for support, but not so difficult that the student spends the entire exercise decoding unfamiliar words.

When should I teach fluency?

Previous meta-analyses have shown a strong benefit for fluency instruction in both elementary (primary) school and secondary school (Lee & Yoon, 2017). The National Reading Panel (2000) found a strong benefit for repeated reading starting in the second half of Grade 1. Most scholars tend to support the idea that teaching fluency alongside decoding and comprehension will provide a synergistic effect.

In my research with Dr Rachel Schechter on reading legislation, we found that reading laws that mandated the use of all five pillars (phonemic awareness, systematic phonics, vocabulary, fluency, and comprehension) yielded the highest improvements in reading scores (Hansford & Schechter, 2023).

That said, in my own experience, fluency instruction should shift over time, both in how it is conducted and how much time is spent on it. While kids are in the early emerging stage of reading (ages 3–6), meaning they are still learning the basics of decoding, I think it makes sense to use limited repeated reading with decodable texts. However, as kids enter the decoding stage of reading (ages 7–10), I think it makes sense to both increase the amount of fluency instruction and the types of fluency instruction.

Many assume fluency drills like repeated reading are meant to help students memorise words, similar to whole language. However, if teachers help students decode and segment unfamiliar words, it can, in my opinion, help students better orthographically map new words and create automaticity, both with decoding and word identification.

In my opinion, fluency instruction is most important when students can decode some words but cannot yet read fluently independently. Once students can read a variety of complex texts, without any support, I think fluency instruction becomes less important.

Are there other kinds of effective fluency instruction?

Yes! Varied reading has a couple of RCTs, conducted by the Iowa Reading Research Centre, showing similar fluency outcomes to repeated reading (IRRC, 2018). Recently, Mastrothanasis et al. (2023) also conducted a meta-analysis on readers theatre. This meta-analysis examined 10 experimental or quasi-experimental studies on students aged 6–12 years. The study found a mean effect size of 1.23. One study included in the meta-analysis was an outlier (Huang & Luo, 2017), with a mean effect size of 5.19. That said, even with the outlier removed, the result is an unweighted mean effect size of 0.94, which is large.

Personally, I think what really matters is that we are providing students with plenty of opportunities to read rich texts aloud, with opportunities to have an adult support their learning.

Final thoughts

Fluency instruction is one of the five pillars of literacy instruction and should be included as part of any literacy program. There is likely a synergistic effect for teaching fluency alongside other forms of instruction such as decoding, vocabulary and comprehension. However, in my opinion, fluency instruction might be most impactful when students are out of the emergent stage (can decode unfamiliar words), but not yet in the fluent reader stage (still lack automaticity with word and sound identification).

Repeated reading is the most evidence-based form of fluency instruction. There have been multiple meta-analyses showing a strong benefit for repeated reading on fluency outcomes, both for the text read and for new texts. That said, repeated reading runs the risk of being boring.

I would recommend limiting repeated reading to no more than 10 minutes a day. Therefore, I think it is best to also include other fluency exercises within daily instruction such as varied reading, readers theatre or partnered reading.

This article originally appeared on Tim Rasinski's blog.

Nathaniel Hansford [@NateJoseph19 on X] has taught every grade from Pre-K to 12 in many interesting and diverse locations – from South Korea to the subarctic of Quebec. He specialises in using meta-analysis research to help teachers implement proven methodologies.

On mathematics anxiety

Rebecca Birch

The jury is in – maths anxiety is a problem, and it seems to be worse for girls.

Mathematics anxiety seems so prevalent that it's begging for its own DSM-5 entry. I spoke to YouTube superstar Eddie Woo a while back, who gave the discussion the nuance it rarely gets. The Centre for Independent Studies has issued a report, which gives some clarity about what mathematics anxiety is and isn't. It gives some answers about its prevalence and the potential consequences, but not so much on what to do about it.

The report was written by the legendary David Geary, who argues that there are two types of knowledge acquisition, biologically primary and biologically secondary. Primary knowledge requires no direct teaching, like basic means–end problem solving or the acquisition of oral language. Babies acquire more words than would even be possible with direct teaching in the first few years of life. On the other hand, they don't spontaneously pick up solving equations. The kind of learning that happens in schools is unnatural and formal education is a recent invention in the evolutionary scheme of things.

Geary also knows a lot about how we learn mathematics. His analysis paper does a great job of summarising the issues, and the prognosis is not great. The good news is that this is a learned or conditioned fear. It's a response to a situation, and in that sense similar to examination anxiety, which so far also doesn't have its own DSM-5 classification. It's not particularly domain specific. It can be treated like any other anxiety response. Exposure therapy through teacher-led tutoring can work. Just as with other inquiry or student-led methods, a lack of structure and guidance can actually exacerbate student anxiety. However, it's concerning that many families would be unable to afford intensive therapy for their child, not to mention the shortage of teachers and therapists available to do this work.

What's more concerning than the coverage to date is that girls appear to be more prone than boys, even when controlling for overall test anxiety and anxiety traits. Generally, early struggle in mathematics leads to anxiety, but high performing girls experience anxiety in a similar way to those who genuinely struggle. The resulting performance avoidance cycles can eventually impact achievement. Possibly as a result of anxiety, girls express lower utility beliefs about the subject of mathematics, and this may go some way to explain why fewer girls pursue STEM careers.

There seems to be little consensus on how to practically treat mathematics anxiety. I can say for certain that schools are too stretched to provide the kind of CBT and exposure therapy recommended. Adults – including teachers and parents – are sometimes known to express their own anxieties about mathematics. I've done this in front of my own children, who thankfully ignore me most of the time. I recently completed an online introduction to statistics and felt a wild panic come up every time I became cognitively overloaded. Perhaps automaticity of basic maths fact recall is key. But, like reading, if it is not embedded from a young age, it may be very difficult to steer students away from a lifetime of avoidance and fear.

What's more concerning than the coverage to date is that girls appear to be more prone than boys, even when controlling for overall test anxiety and anxiety traits.

This article originally appeared on the author's blog, On Education.

Rebecca Birch [@msrebeccabirch on X] is Director of Research and Practice at a K–12 independent school in Sydney. Rebecca has provided consultation and content services for Ochre Education, Catholic Education Canberra Goulburn, and Catholic Education Tasmania. She has also appeared on panels for researchED and the Gonski Institute for Education. Her current masters research is on the links between explicit instruction, self-regulated learning and student wellbeing.

Controversies within the Science of Reading

Jill Barshay, The Hechinger Report

This story was produced by The Hechinger Report, a non-profit, non-partisan news outlet focused on education.

A growing number of researchers are criticising an overemphasis on auditory skills. Educators around the US have embraced the 'Science of Reading' in their classrooms, but that doesn't mean there's a truce in the reading wars. In fact, controversies are emerging about an important but less understood aspect of learning to read: phonemic awareness.

That's the technical name for showing children how to break down words into their component letter sounds and then fuse the sounds together. In a phonemic awareness lesson, a teacher might ask how many sounds are in the word (cat). The answer is three: 'k', 'a' and 't'. Then the class blends the sounds back into the familiar sounding word: from 'kuh-aah-tuh' to 'kat'. The 26 letters of the English alphabet produce 44 phonemes, which include unique sounds made from combinations of letters, such as 'ch' and 'oo'.

Many schools have purchased scripted oral phonemic awareness lessons that do not include the visual display of letters. The oral lessons are popular because they are easy to teach and fun for students. And that's the source of the current debate. Should kids in kindergarten or first grade be spending so much time on sounds without understanding how those sounds correspond to letters?

A new meta-analysis confirms that the answer is no. In January 2024, five researchers from Texas A&M University published their findings online in the journal *Scientific Studies of Reading*. They found that struggling readers, ages 4 to 6, no longer benefited after 10.2 hours of auditory instruction in small-group or tutoring sessions, but continued to make progress if visual displays of the letters were combined with the sounds. That means that instead of just asking students to repeat sounds, a teacher might hold up cards with the letters C, A and T printed on them as students isolate and blend the sounds.

Meta-analyses sweep up all the best research on a topic and use statistics to tell us where the preponderance of the evidence lies. This newest 2024 synthesis follows three previous meta-analyses on phonemic awareness in the past 25 years. While there are sometimes shortcomings in the underlying studies, the conclusions from all the phonemic meta-analyses appear to be pointing in the same direction.

"If you teach phonemic awareness, students will learn phonemic awareness," which isn't the goal, said Tiffany Peltier, a learning scientist who consults on literacy training for teachers at NWEA, an assessment company. "If you teach blending and segmenting using letters, students are learning to read and spell."

Phonemic awareness has a complicated history. In the 1970s, researchers discovered that good readers also had a good sense of the sounds that constitute words. This sound awareness helps students map the written alphabet to the sounds, an important step in learning to read and write. Researchers proved that these auditory skills could be taught, and early studies showed that they

There isn't a bright line between phonemic awareness and phonics, and they can be taught in tandem.

could be taught as a purely oral exercise without letters.

But science evolved. In 2000, the National Reading Panel outlined the five pillars of evidence-based reading instruction: phonemic awareness, phonics, fluency, vocabulary and comprehension. This has come to be known as the Science of Reading. By then, more studies on phonemic awareness had been conducted and oral lessons alone were not as successful. The Reading Panel's meta-analysis of 52 studies showed that phonemic awareness instruction was almost twice as effective when letters were presented along with the sounds.

Many schools ignored the Reading Panel's recommendations and chose different approaches that didn't systematically teach phonics or phonemic awareness. But as the Science of Reading grew in popularity in the past decade, phonemic awareness lessons also exploded. Teacher training programs in the Science of Reading emphasised the importance of phonemic awareness. Companies sold phonemic programs to schools and told teachers to teach it every day. Many of these lessons were auditory, including chants and songs without letters.

Researchers worried that educators were overemphasising auditory training. A 2021 article, 'They Say You Can Do Phonemic Awareness Instruction "In the Dark", But Should You?' by nine prominent reading researchers criticised how phonemic awareness was being taught in schools.

Twenty years after the Reading Panel's report, a second meta-analysis came out in 2022 with even fresher studies but arrived at the same conclusion. Researchers from Baylor University analysed over 130 studies and found twice the benefits for phonemic awareness when it was taught with letters. A third meta-analysis was presented at a poster session of the 2022 annual meeting of the Society for the Scientific Study of Reading. It also found that instruction was more effective when sounds and letters were combined.

On the surface, adding letters to sounds might seem identical to teaching phonics. But some reading experts say phonemic awareness with letters still emphasises the auditory skills of segmenting words into sounds and blending the sounds together. The visual display of the letter is almost like a subliminal teaching of phonics without explicitly saying, "This alphabetic symbol 'a' makes the sound 'ah'." Others explain that there isn't a bright line between phonemic awareness and phonics, and they can be taught in tandem.

The authors of the latest 2024 meta-analysis had hoped to give teachers more guidance on how much classroom time to invest on phonemic awareness. But unfortunately, the classroom studies they found didn't keep track of the minutes. The researchers were left with only 16 high-quality studies, all of which were interventions with struggling students. These were small-group or individual tutoring sessions on top of whatever phonemic awareness lessons children may also have been receiving in their regular classrooms, which was not documented. So, it's impossible to say from this meta-analysis exactly how much sound training students need.

The lead author of the 2024 meta-analysis, Florina Erbeli, an education psychologist at Texas A&M, said that the 10.2 hours number in her paper isn't a 'magic number'. It's just an average of the results of the 16 studies that met her criteria for being included in the meta-analysis. The right amount of phonemic awareness might be more or less, depending on the child.

Erbeli said the bigger point for teachers to understand is that there are diminishing returns to auditory only instruction and that students learn much more when auditory skills are combined with visible letters.

I corresponded with Heggerty, the market leader in phoneme awareness lessons, which says its programs are in 70% of US school districts. The company acknowledged that the Science of Reading has evolved and that's why it

Brady's concern is the misunderstanding that teachers need to teach all the phonemes before moving on to phonics. It's not a precursor or a prerequisite to reading and writing.

revised its phonemic awareness program in 2022 to incorporate letters and introduced a new program in 2023 to pair it with phonics. The company says it is working with outside researchers to keep improving the instructional materials it sells to schools. Because many schools cannot afford to buy a new instructional program, Heggerty says it also explains how teachers can modify older auditory lessons.

The company still recommends that teachers spend 8 to 12 minutes a day on phonemic awareness through the end of first grade. This recommendation contrasts with the advice of many reading researchers who say the average student doesn't need this much. Many researchers say that phonemic awareness continues to develop automatically as the child's reading skills improve without advanced auditory training.

NWEA literacy consultant, Peltier, whom I quoted earlier, suggests that phonemic awareness can be tapered off by the fall of first grade. More phonemic awareness isn't necessarily harmful, but there's only so much instructional time in the day. She thinks that precious minutes currently devoted to oral phonemic awareness could be better spent on phonics, building vocabulary and content knowledge through reading books aloud, classroom discussions and writing.

Another developer of a phonemic awareness program aimed at older, struggling readers is David Kilpatrick, Professor Emeritus at the State University of New York at Cortland. He told me that five minutes a day might be enough for the average student in a classroom, but some struggling students need a lot more. Kilpatrick disagrees with the conclusions of the meta-analyses because they lump different types of students together. He says severely dyslexic students need more auditory training. He explained that extra time is needed for advanced auditory work that helps these students build long-term memories, and the meta-analyses didn't measure that outcome.

Another reading expert, Susan Brady, Professor Emerita at the University of Rhode Island, concurs that some of the more advanced manipulations can help some students. Moving a sound in and out of a word can heighten awareness of a consonant cluster, such as taking the 'l' out of the word 'plant' to get 'pant', and then inserting it back in again. But she says this kind of sound subtraction should only be done with visible letters. Doing all the sound manipulations in your head is too taxing for young children.

Brady's concern is the misunderstanding that teachers need to teach all the phonemes before moving on to phonics. It's not a precursor or a prerequisite to reading and writing. Instead, sound training should be taught at the same time as new groups of letters are introduced. "The letters reinforce the phoneme awareness, and the phoneme awareness reinforces the letters," said Brady, speaking at a 2022 teacher training session. She said that researchers and teacher trainers need to help educators shift to integrating letters into their early reading instruction. "It's going to take a while to penetrate the belief system that's out there," she said.

I once thought that the reading wars were about whether to teach phonics. But there are fierce debates even among those who support a phonics-heavy Science of Reading. I've come to understand that the research hasn't yet answered all our questions about the best way to teach all the steps. Schools might be over-teaching phonemic awareness. And children with dyslexia might need more than other children. More importantly, the Science of Reading is the same as any other scientific inquiry. Every new answer may also raise new questions as we get closer to the truth.

Jill Barshay [@jillbarshay on X] writes the weekly 'Proof Points' column about education research and data, covering a range of topics from early childhood to higher education. She taught algebra to ninth graders for the 2013–14 school year. Previously, Barshay was the New York Bureau Chief for Marketplace, a national business show on public radio stations. In 2019, she received the American Educational Research Association's award for excellence in media reporting on education research. A graduate of Brown University, Barshay holds master's degrees from the London School of Economics and Columbia University's Graduate School of Journalism.

What words do children encounter when they read for pleasure?

Maria Korochkina

Kathy Rastle

The ability to read opens up worlds. Reading enables children to progress into post-primary education and provides the basis for lifelong learning and prosperity into adulthood. Importantly, the journey to becoming a skilled reader requires not only high-quality classroom instruction but also many years of practice through independent book reading.

We wanted to learn more about the vocabulary that children encounter when they read for pleasure. To do this, we analysed the words in 1200 books popular with British children aged 7–16. The original research article is open access and free to download, and we summarise the key insights from this work below.

Books contain a vast number of words

Reading seems so fast and automatic that sometimes people think that to be able to read we just memorise printed words. This idea has led to strategies that try to teach children to memorise the shapes of words. However, this type of rote learning is effortful and takes a long time. For instance, children in China need to memorise 2500 characters during primary school, and this takes around 9 hours per week for 6 years!

A working knowledge of 2000–3000 characters is enough to understand most modern texts in standard Chinese. Not so in English – the 1200 children's books that we analysed contain over 100,000 different words! There is no way that a child could memorise so many printed words. That's why *phonics is so powerful*. Without understanding the connections between letters and sounds, children won't be able to break down the wide variety of words that they will encounter during independent reading.

Books contain many words that children may not know

We found that around 40% of words used in children's books do not appear on BBC television programs aimed at children of the same age. Similarly, one fifth of words used in books for young people aged 13–16 are not encountered on BBC channels targeting adult audiences. The most common of these words include rare and sophisticated vocabulary – often of foreign origin – related to science (e.g. 'meridian', 'homunculus'), arts (e.g. 'aria'), history (e.g. 'marquis', 'inquisitor'), politics (e.g. 'communists', 'suffragists', 'abolitionist', 'legislature') and religion (e.g. 'quaker', 'missionary'). Typically, if a word is not in our spoken vocabulary, we can use context to infer meaning. However, if a book includes too many words that a child does not know, reading it will be a struggle.

The large numbers of unfamiliar words mean that *books present a unique opportunity for enhancing children's vocabulary*. However, the other side of the coin is that, f*or many children, reading is likely to pose a challenge from the earliest years of independent reading.*

What words do children encounter when they read for pleasure?

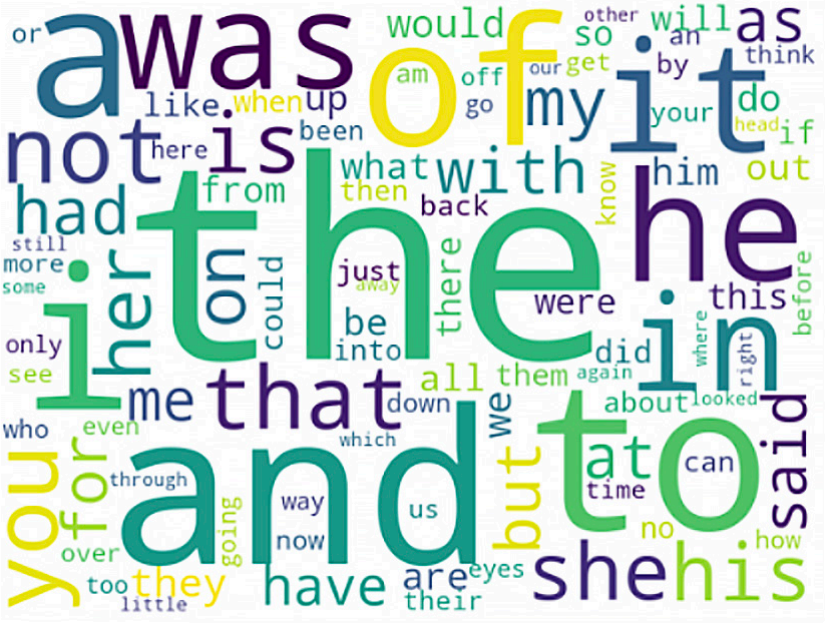


Figure 1: 100 most common words in the 1200 children's books we analysed.

Few words are used repeatedly in books

We found that the most common 100 words make up around 54% of the 1200 books that we analysed. Most of these words belong to a class of words called function words: these are words like 'do', 'and', 'not', 'but', or 'is', which are used to express relationships between other words. Every second word encountered in children's books is a function word, and children will quickly learn to recognise these words by sight. This form of sight-word recognition increases the speed of reading, but these words carry little meaning, and being able to read them quickly will not be enough to understand what a text is about.

To illustrate, consider a sentence from one of the books we analysed where all but the function words have been removed: "Then … a … her … and she … her … and … her." Can you guess what this is about? Now consider the original sentence: "Then a mischievous thought flashed across her eyes, and she pursed her lips together and pushed her tongue forward." This example shows why *being able to read the top 100 words effortlessly is not sufficient to read for meaning.*

New words are encountered in every book

It turns out that the vast majority of words in children's books are only encountered a few times and in a small subset of books. One consequence of this is that books written for children of the same age tend to vary greatly in the words they use. This is particularly so for books aimed at younger primary school children. The words in these are less similar to one another compared to the words in books written for older children. Likewise, books for older children include many words that are not encountered in books for younger children. For instance, more than one third of words in books for children aged 10–12 are never used in books for younger children, and more than one third of words in books for young people aged 13–16 do not occur in books targeting older primary school children. This means that reading is likely to continue to pose a challenge as children grow older.

With book vocabulary being so intense, it is crucial to develop reading skills and motivation early on. And because different books use different words, it is important that children read widely.

Most new words have complex structure

We have said that the vocabulary in books is more sophisticated than the vocabulary on television and tends to get richer as children age. One way that we see this richness is through morphological complexity. Morphologically complex words are words that consist of several elements that are themselves meaningful: for example, the word 'mistrustfulness' consists of four elements, 'mis-', '-trust-', '-ful', and '-ness'. To understand the meanings of these words, a child needs to know what each individual component means and how it contributes to the overall meaning of the complex word. This body of knowledge is referred to as morphological knowledge.

The 1200 children's books that we analysed use thousands of morphologically complex words. Examples include words like 'inexpensively', 'unwinnable', 'unlawfulness', 'speechlessness' or 'outlandishly' – these words appear in books for children aged 10–12, but not in books for younger children. As skilled readers, we can easily understand these words even if we haven't encountered them before because each of these words is created by combining elements we already know (e.g. 'speech' + '-less' + '-ness'). However, these words will be very challenging to those who have not yet learned that 'in-' and 'un-' mean 'not', or that '-ness' denotes a noun and '-ly' an adverb. For these reasons, strong morphological knowledge is key to being able to read well.

Conclusions

It is widely accepted that reading ability is a strong predictor of how much children choose to read. Our work suggests that a failure to acquire good phonic and morphological knowledge early in reading acquisition is likely to have a negative snowball effect on a child's reading habits. On the other hand, our analyses show that the books popular with British children today offer a wonderful opportunity to build vocabulary, particularly if children read widely.

This article originally appeared on the Rastle Lab blog.

Maria Korochkina (https://mariakna.github.io/) is a Postdoctoral Research Fellow in the Department of Psychology at Royal Holloway, University of London. Maria's research to date has focused on how we learn and consolidate new words, and on how word learning is influenced by the context in which new words are encountered.

Kathy Rastle [@Kathy_Rastle on X] is Professor of Psychology at Royal Holloway, University of London. She studies the neurocognitive mechanisms that underpin how people read and learn to read. She is very interested in using insights from research on reading to improve outcomes for children.

Response to the ACT inquiry into literacy and numeracy teaching

Recently, the Australian Capital Terrority (ACT) Government called for public submissions to inform their independent inquiry into the teaching of literacy and numeracy in public schools. Below are the key points excerpted from MultiLit's submission. The entire submission can be accessed via the link at the end of this excerpt.

Key points in this submission

As an organisation that has been working closely with Australian schools and school systems for more than 25 years to improve student outcomes, MultiLit welcomes the opportunity to respond to the Consultation Paper put forward by ACT's Literacy and Numeracy Education Expert Panel.

In developing this response paper, we have chosen to address the questions that relate most to the content and format of literacy instruction, intervention and assessment. All our responses align well with the Education Directorate's 'Key Teaching Strategies', listed on pages 13–14 of the Consultation Paper. The key points can be summarised below:

1. The Response to Intervention (RtI) framework was developed to provide all students in schools with the opportunity to access additional services if they do not respond to instruction received at a whole-class level.
2. High-quality, evidence-based instruction can easily and effectively be embedded into an RtI framework. Timely implementation of this learning content may enable teachers to respond equitably to the range of skills and backgrounds of students in their classrooms before learning gaps become entrenched and the associated negative impacts on wellbeing take effect.
3. Assessment (including ones that allow for screening and progress monitoring) should inform teachers' decisions around what instructional tier is most suitable for each student.
4. Mandating the provision of evidence-based instruction in ACT government schools represents another step towards more equitable learning outcomes for students, given that it will result in less variability in the quality of content received by students.
5. Prescribing evidence-based instruction and intervention will alleviate teacher workload by constraining the selection of materials to those that are most likely to be effective. Prescribing assessments and an assessment protocol that aligns with an RtI framework will be similarly cost- and time-effective if the measures are valid and reliable, and if they directly inform decisions around students' learning.

The entire submission can be accessed from the ACT Government's inquiry 'submissions' webpage.

Nomanis Notes
Issue 17 | April 2024

Do ‘brain breaks’ help students learn?

Jennifer Buckingham and Maddy Goto

Statement of the problem

It is essential for attention to be maintained for learning to happen effectively. In a classroom situation, there are several internal and external factors that can lead to inattention and a loss of focus. Orienting attention to a specific input or task (concentrating) requires conscious effort. This draws on executive functions that are still developing in children and they therefore can have difficulty attending to lessons for long periods of time.

Proposed solution

Punctuating learning with ‘brain breaks’, typically either a physical or mindfulness/meditation activity for 1 to 5 minutes, is a popular tactic among teachers to reset and refocus students’ attention to the learning task. This allows them to briefly shift focus to a less cognitively demanding activity.

The theoretical rationale – how does it work?

It is hypothesised that a brief shift in focus will allow the brain to reach a state of low cognitive load that will let the information being held in working memory begin its transfer to long term memory, before returning to a learning activity. For young children who are unused to sitting still and paying attention, or for children with attention and/or hyperactivity disorders, brain breaks are seen as a way to release energy and then re-engage with learning.

What does the research say? What is the evidence for its efficacy?

Several studies have examined the effects of active breaks on academic achievement and cognitive functions involving primary school children of a range of ages.

A study by Mavildi et al. (2019) with Australian students in Years 3 and 4 found that active breaks resulted in significant improvements in engagement and significant effects for mathematics performance (d = 0.4, p = 0.045). Mazzoli et al. (2019) studied the relationship between time spent sitting, stepping and sit-to-stand movement with cognitive functions and brain activity in younger Australian students. They concluded that students who spent longer sitting were more easily distracted, but the results for cognition measures were inconclusive. Müller et al. (2021) studied active breaks for Year 4 and 5 students and reported a significant positive effect on attention but not reading comprehension. The reverse was reported for mindfulness breaks, with small positive effects on reading comprehension but no effect on attention.

A systematic review by Watson et al. (2017) included four studies of academic outcomes and found only one significant effect for maths. A meta-analysis by de Greeff et al. (2018) found that active breaks had a positive small to moderate effect on attention (d = 0.43) and mixed but weak results for reading (d = 0.17) and maths (d = -0.18). Likewise, Daly-Smith et al. (2018) described active breaks as resulting in no change in cognitive outcomes and weak effects on academic performance. Masini et al.’s (2020) systematic review described the results of studies of active breaks on cognitive functions as inconclusive and determined that active breaks have “limited or no impact on academic achievement”.

Conclusion

Overall, evidence for the effect of active classroom breaks on cognitive and executive functions such as attention/active engagement is moderately positive, but this does not necessarily translate into learning. There is mixed but weak evidence of the effect of active breaks and mindfulness breaks on academic achievement. This may be due to the quality of the studies, or differences in the type, frequency and duration of the breaks; however, based on the current research, there is insufficient evidence to support the benefits of ‘brain breaks’ for learning.

Key references

Daly-Smith, A. J., Zwolinsky, S., McKenna, J., Tomporowski, P. D., Defeyter, M. A., & Manley, A. (2018). Systematic review of acute physically active learning and classroom movement breaks on children’s physical activity, cognition, academic performance and classroom behaviour: understanding critical design features. B*MJ Open Sport & Exercise Medicine, 4*(1). https://doi.org/10.1136/bmjsem-2018-000341

de Greeff, J. W., Bosker, R. J., Oosterlaan, J., Visscher, C., & Hartman, E. (2018). Effects of physical activity on executive functions, attention and academic performance in preadolescent children: a meta-analysis. *Journal of Science and Medicine in Sport, 21*(5), 501–507. https://doi.org/10.1016/j.jsams.2017.09.595

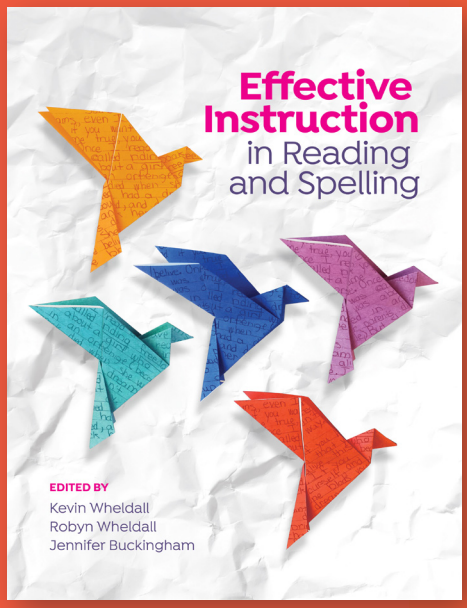
Masini, A., Marini, S., Gori, D., Leoni, E., Rochira, A., & Dallolio, L. (2020). Evaluation of school-based interventions of active breaks in primary schools: A systematic review and meta-analysis. *Journal of Science and Medicine in Sport, 23*(4), 377–384. https://doi.org/10.1016/j.jsams.2019.10.008

Watson, A., Timperio, A., Brown, H., Best, K., & Hesketh, K. D. (2017). Effect of classroom-based physical activity interventions on academic and physical activity outcomes: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity, 14*(1). https://doi.org/10.1186/s12966-017-0569-9

New from MRU Press

MRU PRESS

These new books from MultiLit's academic imprint, MRU Press, are highly recommended for educators wanting to engage with the latest research and inform their practice.



Effective Instruction in Reading and Spelling

Edited by Kevin Wheldall, Robyn Wheldall and Jennifer Buckingham

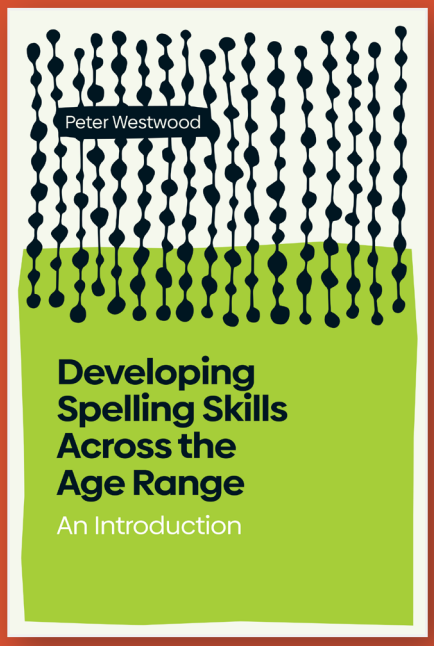
This textbook is an accessible, up-to-date guide to evidence-informed practices in teaching reading and spelling, grounded firmly in the Science of Reading and its application in classrooms.
It is ideal for use in initial teacher education (ITE) degrees and other higher education courses for primary school teachers. It is also a practical yet scholarly reference book for any teacher of reading.

The book covers theories of reading, the scientific evidence base on how children learn to read, the Five Big Ideas of reading, reading-related skills, intervention and assessment, with chapters written by respected Australian and international experts.

Developing Spelling Skills Across the Age Range: An introduction

By Peter Westwood

For too long, the explicit teaching of spelling was neglected. In this clear and concise text, author and educator Peter Westwood steps through the skills required to be an accurate speller, and how teachers can impart these skills to students of all ages – from those in the preschool years right through to adults, with an emphasis on explicit teaching strategies. The book also includes useful print and online resources, making it a practical addition to the bookshelf of any teacher looking to improve their students' spelling.



Visit bookshop.multilit.com to purchase these and other academic books from MRU Press.