Australian Institute for Teaching and School Leadership Limited

Reading Instruction Evidence Guide

August 2020

Contents

Background

This section describes the development of criteria to enable Initial Teacher Education (ITE) providers to identify well-designed studies, trustworthy research, and meaningful evidence to inform decisions relating to the teaching of reading instruction. The purpose of these criteria is to assist ITE providers to be able to scope and define levels or gradations of evidence-based practice and illustrate the strength, validity, and reliability of different types of evidence to support critical consumption of evidence.

It is necessary for ITE providers to collect, collate, and maintain a set of evidence-informed resources relating to reading instruction. It is unlikely that one study or resource will be wholly fit for purpose. Rather, a collection of evidence will be needed to provide coverage over the range of criteria. This is consistent with other approaches in reading research in education arguing that strong evidence is derived from "numerous well-designed qualitative and/or quantitative studies, with high convergence of findings" (Singer et al., 2012 p. 18).

Collating a set of evidence at many levels

There is a long history of attempting to create hierarchies of evidence relating to their type or the methods used to prioritise one piece of evidence over another (Guyatt et al., 1995). While this is predominantly related to interventions in health science and epidemiology, there have been attempts in education (Centre for Education Statistics and Evaluation, 2020; Singer et al., 2012; What Works Clearinghouse, 2020). There are notable criticisms of this approach, which include the fact that hierarchies may miss some of the complexity present in intervention science and prioritise poor-quality research of one type/level over high-quality research of another type/level (Stegenga, 2014).

These criticisms are relevant in reading instruction as it is an applied teaching practice. It is important to establish that there are sound quantitative studies that support the validity and reliability of an intervention as having a substantive positive effect. However, the fact that some practices or interventions are impactful is insufficient information – readers of this research need to know why an intervention works, how to implement it, and how to measure and track progress and impact. A broad range of evidence is required at both the macro- and micro-levels (including qualitative research and expert opinion) in order to unpack how educators can carry out reading instruction effectively.

It is more difficult to quantify the quality of qualitative research and the way that it builds evidence of reading instruction at scale. Ideally, qualitative research used to justify an ITE provider's approach to reading instruction would focus on the implementation of approaches found to have a significant effect in quantitative studies. The role of qualitative research in this context is, therefore, to unpack why something works, or whether it continues to work in a given context.

Some existing frameworks exclude qualitative research, including the What Works Clearinghouse (2020) framework. This approach, however, is too narrow for an ITE provider focusing on the implementation of effective reading instruction.

Qualitative research should, therefore, be included for the purpose of contextualising the implementation of specific strategies, programs, or interventions. For example, to describe or explain

how an educator deploys their expertise (e.g., about their students and the community, content, and pedagogical content knowledge) to make decisions about their practice and interactions.

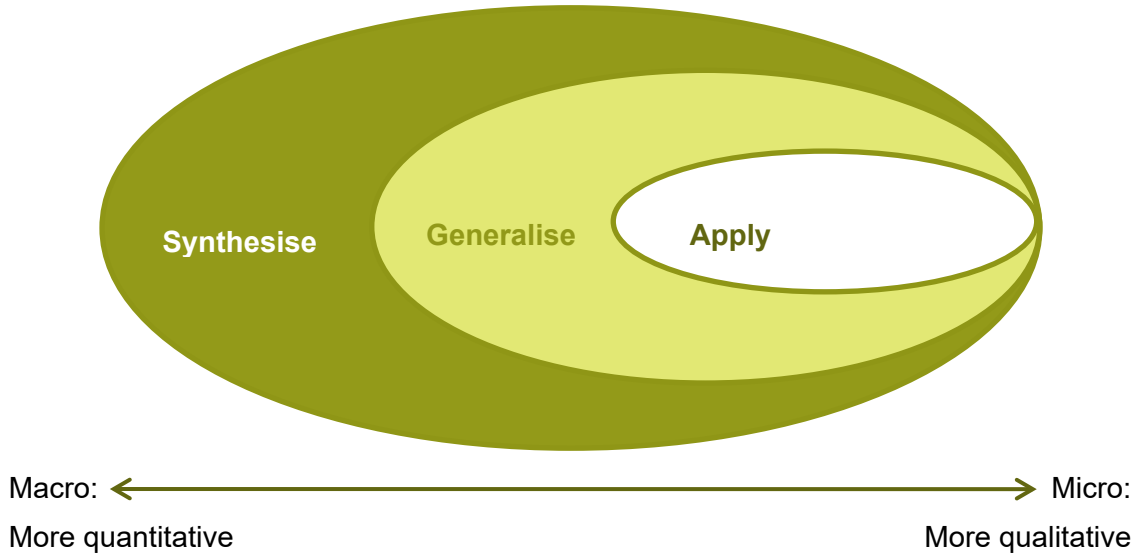

Figure 1

Figure 1 provides an overview of the types of research that would be represented in an evidence set. Each ellipse characterises a different category of research.

- 'Synthesise' includes research that collates and synthesises results from a range of studies – particularly meta-analysis and systematic reviews. Findings in these types of studies, while comprehensive, can lack specific details to help the reader transfer what they have learnt to the classroom for implementation. This category is primarily focused on generating a compelling body of evidence that a particular approach or intervention works across many good-quality studies.
- 'Generalise' includes empirical research, using quantitative methods that can be generalised with some certainty to a given population. For example, the effectiveness of reading groups in Year 1, 2, and 3 classrooms. Results from this type of research provides specific evidence about a single intervention protocol on a set of outcomes, but may lack detail about how the intervention was implemented, what it took to implement and maintain it, and the experience of the participants. This category is primarily focused on demonstrating that a particular approach or intervention is likely to work in general within a given population, using sampling methods and other statistical concepts like reliability to justify this claim to a degree of certainty.
- 'Apply' consists of qualitative research and expert opinion. This kind of evidence can be situated in the classroom and explanatory in nature, providing specific information about how knowledge and pedagogy can be used to make the teaching and learning of reading more effective. For example, using prompts and visuals such as a slinky, elastic band or hand gesture can remind children to segment their words into phonemes when decoding and encoding words while reading. This category is primarily focused on demonstrating how a particular approach or intervention works, including in specific contexts or applications.

It should be noted that while the categories above imply that qualitative and quantitative studies are distinct, it is entirely reasonable that a single piece of evidence could belong in both the 'apply' and 'generalise' categories. This is particularly true for mixed-methods research that may incorporate

elements of each to both establish that an intervention can be generalised to some populations, and to describe how and why it works or under what circumstances.

While it is possible to begin the collection of evidence at any point in the ellipses, for the reliability and validity of results, it is logical to start with quality research at the macro levels ('Synthesise'). Evidence at this level can be robust, having quantifiable findings that can be supported with research from the 'Generalise' and the 'Apply' level.

Applying the quality criteria

To support the collation of a set of evidence across the quality criteria, a template is provided that will allow ITE providers to see at a glance how well the evidence covers the criteria. The template provides a brief and simple guide to key criteria that should be considered in the evaluation of the quality of a piece of evidence. Space for notes is provided to allow for qualifications of the way in which the evidence meets the criteria, or to identify an alternative basis for meeting the criteria for a particular piece of evidence. Meeting all, or most of the template criteria is an indication of the quality of the evidence. The quality criteria are grouped under four headings: fit-for-purpose, practical considerations, evidence quality layer 1, and evidence quality layer 2.

Not all of the criteria will be applied to all individual pieces of evidence. Therefore, it is advisable to apply the criteria in a staged approach in the following order:

- *Fit-for-purpose* is a broad filter to ensure overall good coverage of the big six strands of reading instruction, the presence of a good mix of evidence types (from macro to micro), as well as identifying that each piece of evidence is appropriate for its intended purpose.
- *Practical considerations* are about ensuring the evidence is applicable in settings relevant to ITEs.
- *Quality Layer 1 is* to check the quality of the evidence is generally high, before proceeding. Qualitative and quantitative evidence should meet most of these criteria.
- *Quality Layer 2* is primarily appropriate for quantitative studies or the quantitative elements of a mixed study (e.g., original research including evaluations or interventions) and broadly considers the criteria of empirical validity and reliability.

The path through the quality criteria is illustrated in Figure 2.

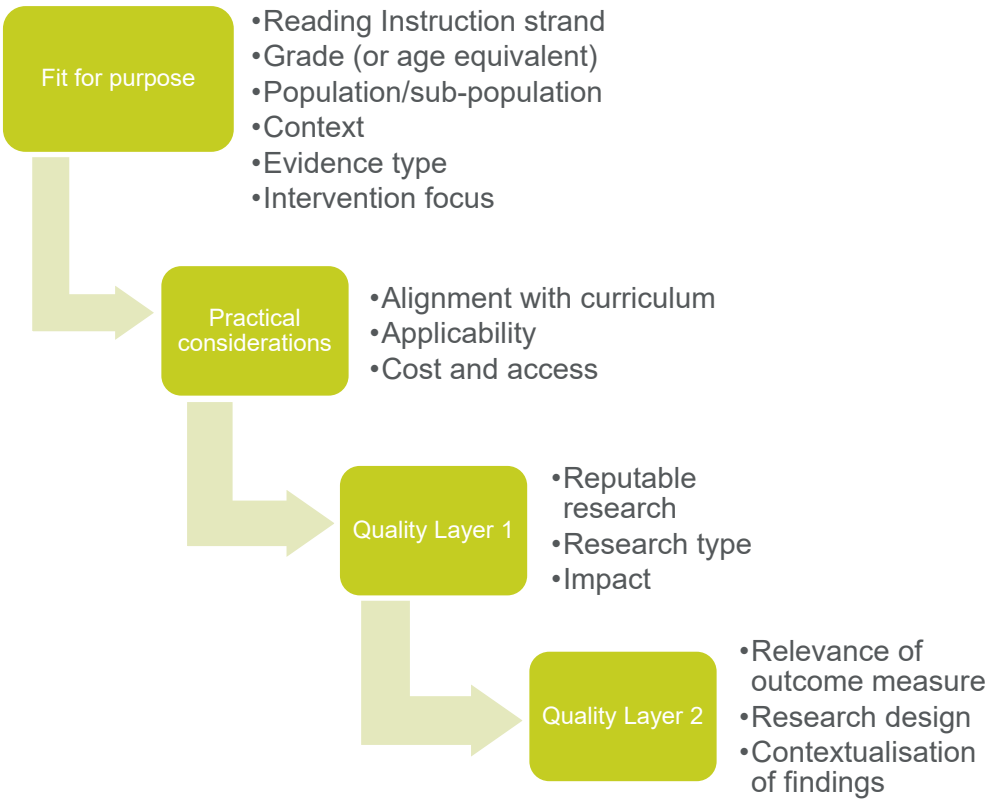

Figure 2.

Supporting resources

Evaluating the quality criteria should be done with reference to the following resources:

- **Glossary –** The glossary within this document provides definitions of the terms used, along with commonly encountered terms in the reading literature.
- **Template –** The template is a tool for collating evidence and rating it against the quality criteria.

Quality Criteria

Fit-for-purpose

Fit-for-purpose is considered under six criteria.

- Reading instruction strand
- Grade/age
- Population
- Context
- Evidence Type
- Intervention focus

To establish if evidence is fit-for-purpose, the evidence being collated should be initially reviewed to ensure it is relevant to one or more of the reading instruction strands, within the scope of F-6 classrooms, and with a relevant population and context. In addition, the type of evidence and the focus of the intervention provides information about the breadth of the set of evidence being collated.

Reading instruction strand

The evidence guide is structured around six key elements of early reading instruction, called strands in these documents. These are:

- phonological awareness (includes phonemic awareness)
- phonics
- fluency
- vocabulary
- comprehension
- oral language.

The first five strands were identified by the National Reading Panel Report (National Reading Panel, 2000). This was a seminal and hugely influential meta-analysis of the research evidence related to early reading and reading instruction that was undertaken in the USA. It was primarily concerned with reading in English. The report identified five pillars of reading: phonemic awareness, phonics, fluency, vocabulary, and comprehension. These pillars have defined early reading constructs in many educational jurisdictions and much of the subsequent research literature addressing reading in English.

The National Reading Panel Report identified phonemic awareness as one pillar. The evidence guide identifies this strand as phonological awareness because this is the overarching construct which includes phonemic awareness. Phonemic awareness is the ability to hear phonemes (the smallest units of sound in English) in words. Phonological awareness includes the ability to hear larger chunks of sound within a word such as onset and rime or syllables, which are skills students typically develop before phonemic awareness.

The role of oral language and early literacy experiences as an additional pillar in the development of early reading proficiency has been convincingly argued by Konza in a systematic review of the literature (Konza, 2014). Konza named the original five pillars, plus her additional pillar, as the ‘Big Six.’

The inclusion of oral language, and use of phonological awareness rather than phonemic awareness, complete the six key strands of reading instruction used in these documents. The terms phonological

awareness; phonics; fluency; vocabulary; comprehension; and oral language, are defined in the Glossary that forms part of the Reading Instruction Evidence Guide materials.

The strands are inter-related and have different roles at different times in the development of children's early reading skills. Some, such as phonological awareness and phonics are constrained skills, which are relatively quickly learned and largely mastered by the time the child starts reading independently for meaning. Others, such as oral language, vocabulary and comprehension, require deep conceptual development and are unconstrained, which means they can continue to develop for the rest of the child's life (Turner et al., 2018).

These six strands are reflected in the Australian Curriculum: English.

- Oral language, vocabulary and comprehension are key elements across the three strands of language, literature and literacy that address listening, reading, viewing, speaking, writing, and creating from Foundation to Year 10.
- Comprehension is elaborated in both the literature strand (e.g., 'responding to literature', 'examining literature') and the literacy strand (e.g., 'interpreting, analysing, evaluating').
- Phonological awareness, phonics, and vocabulary are elaborated in the early level content descriptions for language (e.g., 'phonics and word knowledge' and 'expressing and developing ideas'). Fluency is also elaborated in the Early Years' literacy strand (e.g., Year 2 interpreting, evaluating, analysing).

The national learning progressions also reflect these six strands with oral language, vocabulary and comprehension reflected across many elements of the progressions, and phonological awareness, phonics and fluency addressed as sub-elements (Australian Curriculum, Assessment and Reporting Authority, 2020).

Year Level (or age equivalent)

Relevance of the school year levels from F to 6 (or age-equivalent) reading is rated by indicating, on a presence or absence basis, whether students in Foundation/Year 1, Years 2 and 3, or Years 4, 5 and 6 are explicitly covered in the piece of evidence. Alternatively, evidence can be rated on an age-equivalence. This may be useful for international research that does not use the same mapping of student age to school year level (e.g., different school starting age).

Population or Sub-population

Population is categorised as:

- mainstream
- Aboriginal and Torres Strait Islander/Indigenous peoples
- disability
- learning difficulties
- English as an additional language.

These criteria relate to whether the main focus of the evidence is on mainstream students or includes detailed coverage of the diversity of learners in Australian classrooms. Rating is done by indicating whether the evidence substantially relates to specific sub-populations, including:

- Aboriginal and Torres Strait Islanders in Australian studies
- Indigenous students in other contexts
- students with a learning delay or difficulty
- students with a disability
- children learning English as an additional language.

This is not about prioritising evidence relating to one particular sub-population, but ensuring consideration is given to the breadth of evidence available. A comprehensive set of evidence should include some studies specific to these sub-populations.

Context

Context is categorised as follows:

- Australian or international
- metropolitan, regional, rural, or remote
- socioeconomic status is low, medium, or high
- English not spoken at home.

Context identifies where the evidence was primarily collected, key factors of SES status and whether or not students come from homes where English is the primary language. These factors are known to affect learning outcomes and interpretations of evidence.

Evidence type

Evidence type is categorised as:

- systematic review – a collation of secondary sources that meet the definition of a systematic review. e.g., (Higgins et al., 2019)
- literature review
- meta-analysis - a collection of the results of similar studies that are then analysed quantitatively to give an average effect size
- original research article / case study
- professional opinion.

If the evidence piece is not one of these types, then the template cell for this piece of evidence will be blank, indicating it is another type of evidence, e.g., a popular media piece written by a non-education expert. This type of evidence can be noted in the adjacent column.

Intervention focus

An intervention covers any purposeful, systematic attempt to improve reading instruction. While the teacher may be the key instigator, the student is not necessarily passive. An intervention may focus on increasing student agency and having students participate in shaping the intervention. The substantive focus of an intervention is categorised as:

- curriculum
- pedagogy
- philosophy, approach, or policy
- classroom program or practice.

It is recognised that these elements are inter-related as all teachers work within the broad context of prescribed curricula, philosophical and pedagogical approaches, and recommended interventions that often become highly specific in a particular school. The purpose here is to identify the substantial focus of a piece of evidence, which may cover more than one element in detail.

These criteria are not about promoting one intervention focus above another but relate to the level of coverage of different types of intervention evidence. Improving early reading requires a multi-faceted approach and consideration of different kinds of interventions. It is not reasonable to only focus on school policies relating to reading instruction, for example.

In general, **a curriculum** would include a description of the breadth of key knowledge skills and abilities, or a description of growth through key knowledge skills and abilities (e.g., could include a

learning progression, or a developmental sequence). It is used as the primary document to support the planning and development of classroom activities to support reading instruction.

A **pedagogy** is the interaction in the classroom that educators deploy to support learning and development, and may include particular strategies (the use of feedback or questioning, the modelling of language) or more general approaches (e.g., play-based learning, inquiry-based learning, project-based learning).

A **philosophy, approach, or policy** is an over-arching approach to reading instruction that may broadly align with different curricula and a range of pedagogies, or may be integral to the motivation or implementation of a particular curriculum or pedagogy. For example, a broadly aligned approach to early reading interventions may relate to the time devoted to reading instruction and support for reading (e.g., appointing a reading lead, or a reading community of practice) across a range of pedagogies. A more embedded approach may concern the implementation of standards (e.g., hours or particulars of specific kinds of professional development) or other policies that are clearly aligned to a specific curriculum and pedagogy.

A **classroom program/practice** refers to a teaching program or packaged reading intervention that will include aspects of the previous three aspects to intervention. It is typically a detailed set of instructions about classroom implementation of a particular intervention. This may or may not be a branded intervention (e.g., Reading Recovery) or a commercial program (e.g., Jolly Phonics)

Practical considerations

These considerations are categorised as:

- alignment with the curriculum
- applicability to classroom implementation
- cost and access.

Alignment

The content of the evidence should be aligned with the Australian Curriculum. Alignment should not be rated strictly, but relates more to ensuring that the findings are relevant to policies, practices, or approaches that could be reasonably enacted in an Australian classroom delivering the Australian Curriculum.

This evidence guide is concerned with the quality of evidence about the six key strands of phonological awareness, phonics, fluency, vocabulary, comprehension, and oral language. It was identified earlier that these strands form a substantive part of the Australian Curriculum: English, where the curriculum addresses reading, with the more constrained skills aligning with those described in the Early Years. Most evidence that is focussed on early reading instruction in English is likely to be relevant to the English curriculum, with some evidence also applying to the General Capabilities.

Alignment is generally unlikely to be at the level of matching specific content descriptions from the curriculum to a research study. The findings of most pieces of evidence are likely to be relevant to a number of content descriptions across several years, and may align with content descriptions across two or three strands. Some evidence may be aligned to teaching strategies to improve reading instruction that are broadly applicable across pedagogical approaches advocated by the curriculum.

Applicability

This is considered in relation to the following:

- relates directly to classroom practice

- informs professional understanding of pedagogy/content
- informs policy.

This criterion considers the extent to which the evidence is immediately applicable in a classroom context, or more broadly informative.

Cost and access

This is considered in terms of the following:

- freely available
- requires registration
- incurs costs
- restricted access.

Practical issues include consideration of costs, for example, the requirement to purchase specific curriculum materials or ongoing registration costs.

Is there a requirement of qualification beyond that of Teacher Registration (the most common example would be an assessment that is used in the piece of evidence, that requires a specialist qualification, e.g., as a speech pathologist or occupational therapist)? Availability of materials needs to be considered. Some evidence may use curriculum materials, or other supporting materials or resources (e.g., a learning community website) that are simply not available (e.g., not documented, no longer available, not sold or available in Australia). Lastly, are the materials freely available, e.g., open-source or freely available for use?

Evidence quality

This set of criteria is broken into two stages. See the flow chart above (*Figure 2*). Quality Layer 1 is the first step. It is a top-level quality check. All evidence should meet at least one of the criteria for reputable research and for impact in order to be considered of sufficient quality. Quality Layer 2 is the second step, which relates to a finer-grained review of the quality of empirical research.

The quality of research evidence is an area of much debate (across disciplines, between ontological and epistemological perspectives). This section aims to provide a snapshot of the quality of the collected evidence without getting bogged down in a particular perspective. There is, however, a weight given to evidence that is representative of Australian students and likely to be able to be replicated because of careful controlling of extraneous factors. This is consistent with the approaches taken in other large-scale research and evidence repositories that attempt to rate or quantify the quality of the entries (What Works Clearinghouse, 2020).

Layer 1: Reputable research

This is considered in the following terms:

- authors are experts
- contains citations
- publisher is reputable
- acknowledges limitations.

This section is rated on a presence or absence basis on each of four elements. The first relates to the expertise of the authors: the authors should have a track record of research or providing expert advice in reading instruction. This may be indicated by a higher degree or publication history in the discipline. The second relates to the publication itself; it should contain citations to support claims made. The citations should be collected into a reference list, and the evidence provided should be accessible to

the reader (that is, they cite publicly available evidence themselves). The third relates to the reputation of the publisher of the evidence. The publisher should be reputable in the sense that they have a track record of publishing education research or findings, and not be a predatory publisher. Finally, the authors acknowledge the limitations of the evidence (e.g., debates in the literature, limitations of the methods, gaps or future work required).

Layer 1: Research type

The research type is an indicator of the type of evidence presented:

- quantitative
- qualitative
- mixed.

As discussed above, this is not an implied hierarchy. It is recognised that a mix of evidence types is likely needed to provide adequate coverage of research into effective reading instruction.

Layer 1: Impact

This is considered in terms of:

- recent (last 25 years)
- cited by others at least 5 times in the past 10 years.

The evidence should have been published in the previous 25 years (e.g., relative to 2020, in or after 1995), or be a significant part of a review article or meta-analysis in the past 10 years. Similarly, the evidence should form part of a coherent body of literature. This is evidenced by the piece of work itself being cited in recent peer-reviewed literature, or by the relevant intervention or approach being cited in recent peer-reviewed literature. For the evidence to have had a high impact, it should have been cited at least 5 times in the past 10 years.

Layer 2: Relevancy of Outcome measures

Quality evidence for early reading should be shown to be relevant to reading and valid through the use of an outcome measure for the results or impact described in the evidence. This should be theorised against a framework or learning progression, so the results make sense in the broad context of early reading development and are aligned with similar findings and other relevant evidence.

To be **relevant** to early reading the measure must be related to one of the big six strands of reading[1] .

To be a **valid** measure, the results or impact must be linked to, and explained through, the broad context of early reading development with evidence-based justifications of the framework or learning progression that is used. Validation may include the publication of a validation study (whether as a separate publication or included in the piece of evidence being reviewed). It should be noted that validity in this case refers to the theoretical justification of the coverage of the construct (construct

[1] It should be noted that the measure used may be relevant without being a direct assessment of student learning. A study may, instead, focus on the acquisition of teacher pedagogical content knowledge (directly related to one or more of the big six strands) or a pedagogical skill within the area of reading instruction (e.g., focusing on improving the quality of teacher feedback to support oral language acquisition). This kind of evidence should not be given less priority, so long as the implied connection between improved outcomes (e.g., pedagogical content knowledge) and student outcomes is well theorised and supported in the literature.

validity[2]), being made up of relevant content deemed appropriate and important in the assessment context (content validity), and exhibiting longer term associations with outcomes, like NAPLAN (criterion validity).

Relevancy of outcomes measured is rated as 0, 1 or 2 for each of the six early reading strands. A rating of 2 is given for validated measures across the scope of a strand and a rating of 1 for validated measures of a sub-strand within a strand. A rating of 0 is given if the evidence is not relevant to one of the six strands, or is not validated.

For example, a measure of an economic outcome (employment, lifetime earnings) would not be considered a relevant outcome (and be rated 0). A measure of an overlapping cognitive outcome (e.g., comprehension knowledge (Evans et al., 2002; Phelps et al., 2005)) or a measure of a sub-strand of one of the big six strands (e.g., receptive vocabulary using the Peabody Picture Vocabulary Test) would be considered relevant and rated 1. An outcome measure that comprehensively covers at least one strand of early reading would be considered highly relevant and rated 2. An example of a literacy assessment covering several strands is the literacy components of the school entry assessments used by some states such as the English Online Interview (Victoria), Best Start (Literacy) (NSW) and Early Start (Literacy) (Queensland), and the Progressive Achievement Tests in Reading.

Layer 2: Research design

This is rated as 0, 1 or 2 for:

- reliability
- repeated measures
- control for selection effects.

Aspects of the design of the research itself influence the quality of the measure used. For example, highly relevant measures used in a low-quality study will not yield useful results. In this section, the criteria relate to the applied use of the measure. This is different from the above layer that focuses on the choice of measure and whether it is relevant to reading instruction in its own right. Here, the question is asked: given the research design, can the chosen measure exhibit reliability and therefore be used validly?

Three components of research design are considered: reliability; repeated measures; and control for selection effects.

Reliability[3] is a measure of the degree to which the findings in a piece of evidence would be recreated were the study to be replicated in another (equivalent) sample. Reliability therefore depends strongly on the design of the study: how big is the sample size, how long is the measure being used (test length), and how well matched are the underlying abilities of the students being observed to the contents of the assessment (test targeting)? At the simplest level, a piece of evidence should report a

[2] This is typically demonstrated by convergent validity (this measure should correlate with other measures that have been shown to measure the outcome of interest) and/or divergent validity (this measure should not correlate strongly with other, unrelated measures).

[3] Measures of reliability are affected by multiple sources, including the sample size (the number of students assessed), and the number of items in a measure (test length). It can be shown that a low-reliability measure applied in a large enough sample can yield meaningful results (Adams, 2005). However, this is not the typical pattern, and therefore considerations of validity should focus both on the reliability of the measure (e.g., estimated in other studies and samples) and the sample of children who are measured (e.g., the design of this study).

measure of the reliability of the outcome measure. The most common measure is Cronbach's Alpha, and standard rules of thumb suggest that a minimum of 0.7 should be expected (Nunnally, 1978).

Reliability is rated as 0, 1 or 2. For a rating of 1, a piece of evidence would have a Cronbach's Alpha score of 0.7. Evidence at a higher standard and rated at 2 would be evidence that discusses the targeting of the measure to the sample. This is typical in studies that deploy item response theory.

Repeated measures indicate a robust design. Evidence that demonstrates intra-subject growth tends to have higher precision. That is, a study that simply takes one snapshot of reading ability and compares the mean of an intervention and a control group at the end of a study relies on strong sampling to avoid confounding the initial abilities of students with their final status. Repeated measures are rated as 0, 1 or 2. Evidence that relies on a single snapshot should be rated 0, evidence that shows growth[4] over one point in time (e.g., baseline and end-line) should be rated 1, and evidence demonstrating outcomes over more than two repeated observations should be rated 2.

Control for selection effects is concerned with the way a piece of evidence controls for the selection of subjects into treatment and how it affects the degree to which systematic bias may impact the findings. This is also rated as 0, 1 or 2. Evidence that is simply observational, with no attempt to control for bias should be rated 0; evidence that uses approaches to control for selection such as including covariates e.g., SES, gender, or other contextual features, or uses methods such as propensity score matching should be rated 1 (Duncan & Gibson-Davis, 2006); and randomised control trials or other approaches that legitimately claim to force exogenous allocation to groups (instrumental variables) should be rated 2.

Layer 2: Contextualisation of findings

This is categorised as:

- learning progressions
- interpretable metric
- relative effect size.

High quality research makes a point of putting context around the magnitude and meaning of the findings. This includes, for example, providing a discussion of what skills and abilities have likely changed as a result of a period of growth or an intervention.

- **Learning progressions –** Visualising or describing learning along a developmental progression, such that the results can be described in terms of the knowledge, skills, and abilities that have been modified.
- **Interpretable metric –** Quantifying impact in an interpretable way (e.g., months of growth) and/or ability relative to some expectation (e.g., national norms, or outcome-for-age expectation)
- **Relative effect sizes –** Using effect sizes and comparing to other studies.

[4] There is a debate about how reliable simple gain scores are and how to appropriately model growth, that is beyond the scope of this document (Collins, 1996; Williams & Zimmerman, 1996; Zimmerman & Williams, 1982).

Glossary

This is a glossary of key terms and definitions stemming from the Guide and research scan. The glossary aims to use plain language and includes synonyms and examples where applicable.

Concepts related to evidence

Term	Definition	Sources
Construct validity	The degree to which an assessment or test measures what it claims to be measuring.	Cronbach & Meehl 1955
Content validity	Evidence about the degree to which elements of an assessment or test are relevant to and representative of the targeted construct for a particular assessment purpose.	Almanasreh, Moles & Che 2019, p. 214
Control group	In reference to a research design, the control group is the group that does not receive an intervention and is used to benchmark the relative effect of an intervention.	Duncan & Gibson-Davis 2006
Control variable	In reference to a statistical analysis, a control variable is a variable that is not the independent or dependent variable(s), and is held constant to improve precision or reduce the influence of extraneous factors.	Duncan & Gibson-Davis 2006
Convergent validity	A subtype of construct validity. The degree to which two measures of constructs that theoretically should be related, are in fact related.	Trochim 2020
Cronbach's Alpha	A measure of the reliability of an assessment or test that describes the internal consistency of a measure. A measure of the inter-correlations of the items that make up an assessment or test, such that 0 represents no internal consistency and 1 represents perfect internal consistency.	Cronbach 1951
Effect size / Relative effect size	Effect size is calculated as the difference in performance between the average scores of a group in a trial or experimental condition and those in a comparison condition, divided by the standard deviation of the comparison group (or more typically, divided by the pooled standard deviation	Cohen 1988; Rowe 2005, p. 85

Term	Definition	Sources
	of both groups). An effect size of ≤ 0.2 is regarded as 'weak'; 0.5 is considered as 'moderate'; and 0.8 or larger as 'strong'. Note that there are other types of effect sizes used to describe the magnitude of other statistical tests including correlations and regression coefficients. A common measure of effect size is Cohen's d: a way of quantifying the size of the mean-difference between two groups.	
Effectiveness	Relates to the efficacy of an intervention or program, but usually to the extent that the intervention or program works in a real-world setting.	Singal, Higgins & Waljee 2014
Efficacy	Relates to the extent to which an intervention or program achieves its stated or intended effect.	Singal, Higgins & Waljee 2014
Evidence-based	Concept or strategy that is derived from or informed by objective evidence, and involves the application of rigorous, objective methods to obtain valid answers to clearly specified research questions. "It includes research that: (1) employs systematic, empirical methods that draw on observation and/or experiment designed to minimise threats to validity; (2) relies on sound measurement; (3) involves rigorous data analyses and statistical modelling of data that are commensurate with the stated research questions; and (4) is subject to expert scientific review."	EdGlossary 2016; Rowe 2005, p. 85
Interventions	A collection of approaches, perspectives, curricula, pedagogies, or other strategies that are deployed to change a given outcome.	ATED 2013, What Works Clearinghouse Glossary n.d.
Literature review	A collection, critique, and synthesis of the scholarly work underpinning a topic.	ATED 2013
Meta-analyses	"A statistical method used for summarising findings from many studies that have investigated a similar problem. It provides a numerical way of assessing and comparing the magnitudes of 'average' results – typically expressed as effect sizes."	Rowe 2005, p. 86

Term	Definition	Sources
Mixed method research	An approach to research that includes both qualitative and quantitative approaches to address research questions.	Leech & Onwuegbuzie 2008
Original research	A primary source, usually whereby the researchers who undertook the study have published results or findings.	ATED 2013
Professional opinion (or expert opinion)	A belief or judgement made by a qualified practitioner within a given area or field. The belief or judgement may be made up of personal experiences or knowledge accumulated from other sources.	Ponce et al. 2017
Qualitative research	Research providing detailed narrative descriptions and explanations of phenomena investigated, with lesser emphasis given to numerical quantifications. Methods used to collect qualitative data include ethnographic practices such as observing and interviewing.	ATED thesaurus
Quantitative research	Empirical investigation of observable phenomena derived from the positivist perspective. Usually associated with the development of mathematical and statistical models to represent complex systems.	Box 1976
Randomised control trials	A research design that directly allocates participants to control and intervention conditions probabilistically as a way of controlling for selection bias, particularly bias due to unobserved (uncontrolled) extraneous factors.	Styles & Torgerson 2018
Reliability	The degree to which a measurement (e.g., from an assessment or test) is accurate and reproducible. Also, the degree to which a measurement reduces uncertainty in the estimation of some trait.	Adams 2005
Systematic review	A type of literature review that uses “a transparent and systematic process to define a research question, search for studies, assess their quality and synthesise findings qualitatively or quantitatively.”	Armstrong et al. 2011, p. 147

Term	Definition	Sources
Validity	The extent to which a test, inventory, rating scale, questionnaire, etc., is an effective index of what it is used or intended to measure.	ATED 2013

Concepts related to pedagogy

Term	Definition	Sources
Curriculum	Refers to the *Foundation to Year 12 Australian Curriculum*, or alternative curriculum frameworks that have been assessed by the Australian Curriculum, Assessment and Reporting Authority (ACARA) as meeting the requirements of the Australian Curriculum, or any curriculum authorised by jurisdictional authorities, and the *Early Years Learning Framework for Australia*.	AITSL 2019, p. 11
Explicit instruction	"Explicit teaching practices involve teachers clearly explaining to students why they are learning something, how it connects to what they already know, what they are expected to do, how to do it and what it looks like when they have succeeded. Students are given opportunities and time to check their understanding, ask questions and receive clear, effective feedback." "A systematic method of teaching with emphasis on proceeding in small steps, checking for students' understanding and achieving active and successful participation by all students".	NSW DEC 2020b, p. 11; Rosenshine 1987, p. 34
Learning progressions	Learning progressions describe the skills, understandings, and capabilities that students typically acquire as their proficiency increases in a particular aspect of the curriculum over time.	Online Formative Assessment Initiative 2020
National learning progressions	National learning progressions sit within the broader framework of the Australian Curriculum and can help inform the refinement of the Australian Curriculum.	Online Formative Assessment Initiative 2020

Term	Definition	Sources
Pedagogy	The art and science of teaching. Interaction in the classroom that may include particular strategies (the use of feedback or questioning, the modelling of language) or more general approaches to the way teachers deliver the curriculum, e.g. play-based learning, inquiry-based learning, project-based learning.	ATED 2013; Marzano 2007
Scaffolding	Purposeful, temporary flexible support provided to students to help them achieve beyond their independent level.	Hammond, 2001

Concepts related to reading instruction

Term	Definition	Sources
'Big Six'	See Reading instruction strands	
Alliteration	A recurrence of the same phoneme at the beginning of words in close succession (for example, 'ripe, red raspberry').	ACARA English Glossary
Alphabet knowledge	The ability to identify and name the letters of the English alphabet, to recall the corresponding letter shapes uppercase and lowercase and to represent letters graphically with correct formation.	Picker 2012
Alphabetic principle	"The concept that English uses graphemes (letters and letter combinations) to represent phonemes. Grasping the alphabetic principle depends on both phonemic awareness and familiarity with letters."	Moats 2020
Analogy phonics	An approach that teaches children to use parts of written words they already know to identify new words.	NRP 2000, p. 2-89
Analytic phonics	An approach that avoids having children pronounce sounds in isolation to figure out words. Rather children are taught to analyse letter sound relations once the word is identified. Phonics programs that tend to start with children's known language and introduce shared reading. An explicit focus on words	Ewing p. 11; NRP 2000, p. 2-89

Term	Definition	Sources
	from these sources follows, including teaching children letter-sound correspondences and analysis of words into their component parts. The emphasis is on the larger sub-parts of words (i.e. onsets and rimes, spelling patterns) and phonemes.	
Decoding (Reading)	"A process of efficient word recognition in which readers use knowledge of the relationship between letters and sounds to work out how to say and read written words."	ACARA English Glossary
Embedded phonics	Approach in which "children are taught letter-sound relationships during the reading of connected text. Since children encounter different letter-sound relationships as they read, this approach will not be a preconceived sequence, but can be thorough and explicit."	Ewing 2018, p. 11
Genre	How texts are grouped depending on their social purpose.	ACARA English Glossary
Grammar	A description of a language as a system. In describing a language, attention is paid to both structure (form) and meaning (function) at the level of a word, a sentence, and a text.	ACARA English Glossary
Grapheme	A letter or group of letters that spell a phoneme in a word; can be one, two, three or four letters in English.	ACARA English Glossary
Literacy	"The knowledge and skills students need to access, understand, analyse and evaluate information, make meaning, express thoughts and emotions, present ideas and opinions, interact with others and participate in activities at school and in their lives beyond school." "The ability to identify, understand, interpret, create, communicate and compute, using printed and written (and visual) materials associated with varying contexts. Literacy involves a continuum of learning to enable an individual to achieve his or her goals, to develop his or her knowledge and potential and to participate fully in the wider society."	ACARA n.d.; UNESCO 2004, p. 13

Term	Definition	Sources
Morpheme	The smallest meaningful or grammatical unit in a language. It may be a word or part of a word; it may be a single sound /s/, one syllable, e.g., -ful (suffix), or multiple syllables, e.g. inter- (prefix).	ACARA English Glossary
Morphology	The study of meaningful units in a language and how the units are combined in word formation.	ATED 2013
Onset	The consonant(s) of a word or syllable that precede the vowel, for example, the onset in *hit* is /h/. Not all words have an onset, for example, in the word *it* there is no onset as the whole word is the rime.	Center 2005, p. 267
Onset-rime phonics	The phonological units of a spoken syllable. A syllable can normally be divided into two parts: the onset which consists of the initial consonant or consonant blend and the rime which consists of the vowel and any final consonants, for example: bark b (onset), ark (rime); inside (no onset), in (rime), s (onset), ide (rime).	NESA Glossary n.d.; NRP 2000; NSW DEC 2020a
Oral language	A reading instruction strand "made up of at least five key components: phonological skills, pragmatics, syntax, morphological skills and vocabulary (also referred to as semantics)."	Hill 2010, Konza 2014; Moats 2020
Orthographic mapping	Orthographic mapping involves the formation of letter-sound connections to bond the spellings, pronunciations, and meanings of specific words in memory. It explains how children learn to read words by sight, to spell words from memory, and to acquire vocabulary words from print.	Ehri 2014, p. 5
Orthography	"The visual representation of language in writing as conditioned by the phonological, syntactic, morphological, and semantic features of the language."	Gregory, 2008; Joshi & McCardle 2018
Phoneme blending	The ability to blend several sequential phonemes together to form a word, for example /m/-/a/-/t/. In continuous blending the student does not stop between the sounds, whereas in discontinuous blending there is a pause between each of the uttered sounds.	Anderson et al. 1985; Hempenstall, 1997; Konza 2006, p. 39; Weisberg & Savard 1993

Term	Definition	Sources
Phoneme grapheme correspondence	Relationship between speech sound (phoneme) and written symbol (grapheme).	ATED 2013
Phoneme isolation	Articulating a specified sound in a word in isolation from the word itself, for example, hearing and articulating the initial sound in fish, /f/, and the final sound in mat, /t/.	Ericson & Juliebo 1998
Phoneme segmentation	The ability to detect, separate, and articulate the individual phonemes which make up a syllable or word, for example, plan: /p/-/l/-/a/-/n/.	Burns, Griffin & Snow 1999; Center 2005; Konza 2006
Phonemes	The smallest units in oral language which represent the individual sounds in a word, for example, the word *sat* is composed of the three phonemes /s/-/a/-/t/.	Center, 2005
Phonemic awareness	The ability to focus on and manipulate phonemes in words, for example, knowing that the word *cat* is composed of three phonemes, and that if the middle phoneme is replaced by /u/ the word becomes *cut*.	Gillon 2004; Moats & Tolman 2019
Phonetics	Study and classification of speech sounds, including their production, transmission, and perception. Articulatory phonetics refers to the way sounds are physically produced in the human vocal tract.	ERIC thesaurus n.d.
Phonics	A reading instruction strand that develops specific knowledge of the letter-sound relationships used in reading and writing It involves the combined knowledge of phonological awareness (phonemes) and letters (graphemes).	Rohl, 2000
Phonics through spelling	Programs that teach children to segment and write the phonemes in words.	Rowe 2005, p. 88
Phonological awareness	A reading instruction strand that develops the “realisation that a continuous stream of speech can be broken up into separate words, that words can also be broken up into one or more syllables and that syllables are made up of a sequence of separate single sounds called phonemes”. Phonological awareness is an umbrella term that	Konza, 2006, p. 36; Moats & Tolman 2019

Term	Definition	Sources
	includes conscious awareness of words, syllables, onset and rime, phonemes, word boundaries, stress patterns, syllables and phonemes.	
Phonology	Study of the ways in which speech sounds form systems and patterns in language.	ERIC thesaurus
Reading comprehension	A reading instruction strand that develops understanding of texts "through an active process of making, constructing, or deciphering the meaning of language input through listening, reading, viewing, touching (as in braille or tactile signing) and through combinations of these modes. It involves elements of decoding, working out meaning, evaluating and imagining. The process draws upon the learner's existing knowledge and understanding, text–processing strategies and capabilities; for example, making inferences or applying knowledge of text types and social and cultural resources."	ACARA English Glossary; Victorian DET 2020
Reading fluency	A reading instruction strand that aims to develop "an ability to produce signed, spoken or written language with appropriate phrasing, rhythm and pace. It involves the smooth flow of language, lack of hesitation or undue pausing and characterises the largely accurate use and automatisation of the target language." "Fluent readers are able to devote their finite cognitive resources to the more important task in reading – that is, comprehension."	ACARA English Glossary; NRP 2000 p. 32 & p. 510; Rasinski & Samuels 2011, p. 95
Reading instruction strands	Six strands of reading instruction: • Fluency • Oral language • Phonics • Phonological awareness • Reading comprehension • Vocabulary	Konza 2014
Register	A variety of language shaped by its use in a particular social context.	Gibbons, 2009
Rime	The part of a word or syllable after the onset which includes a vowel and following consonants, for example, the rime in both *hit* and *sit* is *it* (Center 2005, p. 267).	Center 2005, p. 267

Term	Definition	Sources
Semantic information	Information related to meanings used when reading. Semantic information includes a reader's own prior knowledge and the meanings embedded in a text.	ACARA English Glossary
Sight word	A word that is stored in memory and read automatically as a unit.	Ehri, 2005
Standard Australian English	The variety of spoken and written English language in Australia used in more formal settings.	ACARA English Glossary
Syllable	A unit of a speech which contains a vowel or vowel-like speech sound with or without an accompanying onset or rime.	Harris & Hodges 1995
Syntax	The ways in which sentences are formed from words, group/phrases and clauses.	ACARA English Glossary
Synthetic phonics	Approach that places emphasis on the process of synthesising or blending individual sounds together when teaching sounds, and teaches children to convert graphemes into phonemes (e.g., to pronounce each letter in 'stop', /s/-/t/-/o/-/p/) and then blend the phonemes into a recognisable word.	Rowe 2005, p. 88
Vocabulary	A reading instruction strand. The words known and used by a particular person or text.	Victorian DET 2020
Whole-language approach	Whole-language, as a movement, has at its core that learning is 'holistic'. That is, a whole-language approach views listening, speaking, reading and writing as integrated, not separate entities. It is meaning-centred and recognises that students learn the subsystems of language as they engage in it.	Turbill in Rowe 2005, p. 90
Whole-word approaches	Whole-word approaches to the teaching of reading (also known as 'look-say' methods) make no attempt to encourage children to analyse words into letter-sound relationships until a corpus of 'sight words' has been learnt.	Rowe 2005, p. 90

Term	Definition	Sources
Word awareness	The understanding that speech can be broken up into individual words and individual words can be spoken about independently of the object they represent (Rohl, 2000; Konza, 2006).	Rohl 2000; Konza 2006

Additional glossaries

ACARA n.d., F-10 Curriculum English Glossary, https://www.australiancurriculum.edu.au/f-10-curriculum/english/Glossary

Australian Thesaurus of Education Descriptors (ATED) 4th ed. 2013, Australian Council for Educational Research, https://acer.org/ated

Department of Education and Training (2020). Literacy glossary. Literacy Teaching Toolkit. https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/english/literacy/Pages/litglossary.aspx

EdGlossary 2016, Great Schools Partnership. https://www.edglossary.org

ERIC thesaurus n.d., Institute of Education Sciences, https://eric.ed.gov/?ti=all

Macquarie Dictionary 2020, https://www.macquariedictionary.com.au

NSW Education Standards Authority n.d., Glossary, https://educationstandards.nsw.edu.au/wps/portal/nesa/k-10/learning-areas/english-year-10/english-k-10/glossary

Western Australian Curriculum 2020, *Glossary*, https://k10outline.scsa.wa.edu.au/home/p-10-curriculum/curriculum-browser/english/overview/glossary

What Works Clearinghouse Glossary n.d., https://ies.ed.gov/ncee/wwc/Glossary

References

Adams, R. J. (2005). Reliability as a measurement design effect. *Measurement, Evaluation, and Statistical Analysis*, *31*(2), 162–172. https://doi.org/10.1016/j.stueduc.2005.05.008

Almanasreh, E, Moles, R, & Chen, TF 2019, 'Evaluation of methods used for estimating content validity', *Research in Social and Administrative Pharmacy*, vol. 15, no. 2, pp. 214–221, https://doi.org/S1551741118302687.

Anderson, R et al. 1985, *Becoming a Nation of Readers*, National Academy of Education (U.S.). Commission on Reading, Champaign, Ill., http://files.eric.ed.gov/fulltext/ED253865.pdf

Armstrong, R et al. 2011, 'Scoping the scope of a Cochrane review', *Journal of Public Health*, vol. 33, no. 1, pp. 147–150, https://doi.org/10.1093/pubmed/fdr015.

Australian Curriculum, Assessment and Reporting Authority (ACARA). (n.d.). Australian Curriculum. https://www.australiancurriculum.edu.au

Australian Curriculum, Assessment and Reporting Authority. (2020). *National Literacy Learning Progression Version 3.0 March 2020*. Australian Curriculum, Assessment and Reporting Authority. https://www.lpofai.edu.au/media/kw2lslug/national-literacy-learning-progression-v3-for-publication.pdf

Australian Institute for Teaching and School Leadership 2019, Accreditation of initial teacher education programs in Australia, AITSL, Melbourne. https://www.aitsl.edu.au/docs/default-source/national-policy-framework/accreditation-of-initial-teacher-education-programs-in-australia.pdf

Box, GEP 1976, 'Science and statistics', *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 791–799, https://doi.org/10.1080/01621459.1976.10480949.

Burns, MS et al. (eds.) 1999, Starting out right: a guide to promoting children's reading success, National Academy Press, Washington, DC, https://files.eric.ed.gov/fulltext/ED439781.pdf

Centre for Education Statistics and Evaluation. (2020). *How we use evidence*. https://www.gtil.cese.nsw.gov.au/how-we-use-evidence

Center, Y 2005, *Beginning reading: A balanced approach to reading instruction in the first three years at school*, Allen & Unwin, Sydney

Cohen, J 1998, *Statistical power analysis for the behavioral sciences*, 2nd edition, L. Erlbaum Associates, Hillsdale, N.J.

Collins, L. M. (1996). Is Reliability Obsolete? A Commentary on "Are Simple Gain Scores Obsolete?" *Applied Psychological Measurement*, *20*(3), 289–292. https://doi.org/10.1177/014662169602000308

Cronbach, L J 1951, 'Coefficient alpha and the internal structure of tests', *Psychometrika*, vol. 16, no. 3, pp. 297–334, https://doi.org/10.1007/BF02310555

Cronbach, L J & Meehl, P E 1955, Construct validity in psychological tests. *Psychological Bulletin,* vol. 52, no. 4, pp. 281–302. https://doi.org/10.1037/h0040957

Duncan, GJ & Gibson-Davis, CM 2006, 'Connecting child care quality to child outcomes: Drawing policy lessons from nonexperimental data', Eval Rev, vol. 30, no. 5, pp. 611–630, https://doi.org/10.1177/0193841X06291530

Ehri, LC 2014, 'Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning', *Scientific Studies of Reading*, vol. 18, no. 1, pp. 5–21, viewed 22 June 2020, http://doi.org/10.1080/10888438.2013.819356.

Ericson, L & Juliebö, MF 1998, *The phonological awareness handbook for kindergarten and primary teachers*, International Reading Association, Newark, DE.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*(2), 246–262.

Ewing, R 2018, *Exploding some of the myths about learning to read: a review of research on the role of phonics*, New South Wales Teachers' Federation, Sydney, https://www.alea.edu.au/documents/item/1869

Gibbons, P 2009, *English Learners Academic Literacy and Thinking*. Portsmouth, NH: Heinemann.

Gillon, GT 2005, 'Phonological awareness: Evidence to influence assessment and intervention practices', *Language, Speech, and Hearing Services in Schools*, vol. 36, no. 4, pp. 281–284, https://doi.org/10.1044/0161-1461%282005/028%29.

Gregory, E 2008, *Learning to Read in a New Language*, Sage, London

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., Cook, R. J., Bass, E., Gerstein, H., Haynes, B., Holbrook, A., Jaeschke, R., Laupacls, A., Moyer, V., & Wilson, M. (1995). Users' Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations. *JAMA*, *274*(22), 1800–1804. https://doi.org/10.1001/jama.1995.03530220066035

Harris, TL & Hodges, RE (eds.) 1995, *The literacy dictionary: the vocabulary of reading and writing*, International Reading Association, Newark, DE

Hempenstall, K 1997, 'The whole language-phonics controversy: A historical perspective', *Educational Psychology,* vol. 17, pp. 399-418

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Hill, S 2010, 'Oral language play and learning', *Practically Primary,* vol. 15, no. 2, https://www.alea.edu.au/documents/item/1038

Joshi, RM & McCardle, P 2018, 'Models of reading in different orthographies: An Introduction', *Journal of Learning Disabilities*, vol. 51, no. 5, pp. 419-421, https://doi.org/10.1177/0022219417718196

Konza, D 2006, *Teaching children with reading difficulties*, Thomson Social Science Press, South Melbourne

Konza, D. (2014). 'Teaching Reading: Why the "Fab Five" should be the "Big Six"' *Australian Journal of Teacher Education*, *39*(12). http://dx.doi.org/10.14221/ajte.2014v39n12.10

Leech, NL & Onwuegbuzie, AJ 2008, 'A typology of mixed methods research designs', *Quality and Quantity*, vol. 43, no. 2, pp. 265-275, http://doi.org/10.1007/s11135-007-9105-3

Marzano, RJ 2007, *The art and science of teaching: a comprehensive framework for effective instruction*, Association for Supervision and Curriculum Development, Alexandria, Va

Moats, LC 2020, *Speech to print: language essentials for teachers*, Third edition, Paul H. Brookes Publishing, Baltimore, Maryland

National Reading Panel (NRP) 2000, Report of the National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development

NSW Department of Education and Communities (DEC) 2020a, Teaching strategies: Reading, https://www.education.nsw.gov.au/teaching-and-learning/student-assessment/smart-teaching-strategies/literacy/reading

NSW Department of Education and Communities 2020b, *What works best 2020 Update*, Centre for Education Statistics and Evaluation. Centre for Education Statistics and Evaluation, Sydney, https://www.cese.nsw.gov.au/publications-filter/what-works-best-2020-update

Nunnally, J. C. (1978). *Psychometric theory* (2d ed.). McGraw-Hill.

Online Formative Assessment Initiative 2020, Frequently Asked Questions, https://www.lpofai.edu.au/faqs

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The General (g), Broad, and Narrow CHC Stratum Characteristics of the WJ III and WISC-III Tests: A Confirmatory Cross-Battery Investigation. *School Psychology Quarterly*, *20*(1), 66.

Picker, K 2012, *Teaching phonological awareness in Western Australian Pre-primary government school classrooms* (Master's Thesis), The University of Western Australia, Crawley, WA

Ponce, OJ et al. 2017, 'What does expert opinion in guidelines mean? a meta-epidemiological study', *Evidence Based Medicine*, vol. 22, no. 5, pp. 164–169, http://ebm.bmj.com/lookup/doi/10.1136/ebmed-2017-110798

Rasinski, TV & Samuels, SJ 2011, 'Reading fluency: what it is and what it is not', In SJ Samuels & AE Farstrup (eds.), *What research has to say about reading instruction* (pp. 94 –114). International Reading Association, Newark, DE

Rohl, M 2000, 'Learning about words, sounds and letters', In C Barratt-Pugh & M Rohl (eds.), *Literacy learning in the early years* (pp. 57-80), Open University Press, Buckingham, UK

Rosenshine, B 1987, 'Explicit teaching and teacher training', *Journal of Teacher Education*, vol. 38, no. 3, pp. 34-36

Rowe, K 2005, *Teaching Reading: Report and Recommendations*, National Inquiry into the Teaching of Literacy (Australia), Department of Education, Science and Training, https://research.acer.edu.au/tll_misc/5

Singal, AG, Higgins, PDR, & Waljee, AK 2014, 'A primer on effectiveness and efficacy trials', *Clinical and Translational Gastroenterology*, vol. 5, no. 1, p. e45, https://doi.org/10.1038/ctg.2013.13

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering.* National Academies Press.

Stegenga, J. (2014). Down with the Hierarchies. *Topoi*, *33*(2), 313–322. https://doi.org/10.1007/s11245-013-9189-4

Styles, B & Torgerson, C 2018, 'Randomised controlled trials (RCTs) in education research: methodological debates, questions, challenges', *Educational Research*, vol. 60, no. 3, pp. 255–264, https://doi.org/10.1080/00131881.2018.1500194

Trochim, WMK 2020, *Research Methods Knowledge Base*. Conjoint.ly, https://conjointly.com/kb

Turner, R., Adams, R. J., Schwantner, U., Cloney, D., Scoular, C., Anderson, P., Daraganov, A., Jackson, J., Knowles, S., O'Connor, G., Munro-Smith, P., Zoumboulis, S., & Rogers, P. (2018). *Development of Reporting Scales for Reading and Mathematics: A report describing the process for building the UIS Reporting Scales*. Australian Council for Educational Research. https://research.acer.edu.au/monitoring_learning/33/

UNESCO 2004, The plurality of literacy and the implications of its policies and programs, Education Position Paper, http://unesdoc.unesco.org images/0013/001362/136246e.pdf

Victorian Department of Education and Training (DET) 2020, 'Comprehension', *Literacy Teaching Toolkit*, https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/english/literacy/readingviewing/Pages/litfocuscomprehension.aspx

Weisberg, P & Savard, CF 1993, 'Teaching preschoolers to read: Don't stop between the sounds when segmenting words', *Education & Treatment of Children*, vol. 16, no. 1, pp. 1-18, https://www.jstor.com/stable/42899291

What Works Clearinghouse. (2020). *Standards Handbook, Version 4.1*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/handbooks

Williams, R. H., & Zimmerman, D. W. (1996). Are Simple Gain Scores Obsolete? *Applied Psychological Measurement*, *20*(1), 59–69. https://doi.org/10.1177/014662169602000106

Zimmerman, D. W., & Williams, R. H. (1982). Gain Scores in Research Can Be Highly Reliable. *Journal of Educational Measurement*, *19*(2), 149–154. JSTOR.

aitsl.edu.au

Telephone: +61 3 9944 1200
Email: info@aitsl.edu.au

AITSL is funded by the Australian Government