Australian Education Research Organisation

Which skills are important for future literacy and numeracy learning?

How the Australian Early Development Census data reveal the building blocks for future reading, writing and numeracy performance

September 2023

The Australian Education Research Organisation (AERO) is Australia's national education evidence body, working to achieve excellence and equity in educational outcomes for all children and young people.

Acknowledgement

AERO acknowledges that this publication was made possible by the joint funding it receives from Commonwealth, state and territory governments.

Acknowledgement of Country

AERO acknowledges the traditional custodians of the lands, waterways, skies, islands and sea country across Australia. We pay our deepest respects to First Nations cultures and Elders past and present. We endeavour to continually value and learn from First Nations knowledges and educational practices.

Authors

Christine Jackson, Dr Wai Yin Wan, Dr Eunro Lee, Tess Marslen, Dr Lucy Lu, Dr Lisa Williams, Dr Ashleigh Collier, Dr Kelly Johnston and Dr Melanie Thomas.

Contributing authors

Bridget Healey and Dr Melody McCormick.

We would also like to thank Associate Professor Jon Quach for peer-reviewing the report.

Research ethics

Advice on the research ethics requirements of this research was sought from an external NHMRC-registered research ethics review provider. This resulted in the research being assessed as negligible and therefore exempt from research ethics review as per the National Statement: 2.17; 5.1.8; and 5.1.22–5.1.23.

How to cite

Jackson, C., Wan, W.-Y., Lee, E., Marslen, T., Lu, L., Williams, L., Collier, A., Johnston, K., & Thomas, M. (2023). *Which skills are important for future literacy and numeracy learning? How the Australian Early Development Census data reveal the building blocks for future reading, writing and numeracy performance*. Australian Education Research Organisation. https://www.edresearch.edu.au/resources/literacy-numeracy-skills-future-learning

Disclaimer

This report uses data from the Australian Early Development Census (AEDC). The AEDC is funded by the Australian Government Department of Education. The findings and views reported are those of the author and should not be attributed to the Department or the Australian Government.

Publication details

ISBN 978-1-923066-08-3 (Online)
Cover image: AERO

Contents

Figures

Tables

Executive summary

Overview

The research covered in this report was conducted in 2022 and sought to identify the specific areas of literacy and numeracy in which children[1] need the most support. This document reports on analysis of a broad set of children's skills upon entry into school and how these relate to later literacy and numeracy outcomes. The report details findings from one element of the 'Literacy and numeracy project' conducted by the Australian Education Research Organisation (AERO). The project will generate practical, evidence-based resources to support teaching and learning in schools.

The key aim of the 'early childhood' element of the project was to understand how children's learning and development progresses in the early years of school, using available national data. The early years of a child's life, from birth to age 8, are a crucial period for learning and development. High-quality early childhood education and care (ECEC), and continuity from ECEC into school, provide children with a strong foundation for lifelong learning. Understanding how skills related to language and cognition develop during this period of a child's life can help policymakers and education practitioners optimise the support offered to children.

The study contains analysis of linked data on 2,459 children who commenced full-time school in Australia in 2009, drawn from the Australian Early Development Census (AEDC) and National Assessment Program – Literacy and Numeracy (NAPLAN). The AEDC is a triennial national census, in which teachers report on children's development during their first year of full-time school against 5 key developmental domains: Physical Health and Wellbeing, Social Competence, Emotional Maturity, Language and Cognitive Skills (school-based), and Communication Skills and General Knowledge.[2] NAPLAN is a national annual assessment of literacy and numeracy skills, undertaken in Years 3, 5, 7 and 9.

A mixed-methods approach to analysis was implemented across 4 stages. Structural equation modelling was used to predict Year 3 and Year 5 outcomes using AEDC domain and subdomain scores, and Rasch modelling was used to understand the relative difficulty of the AEDC items. Multilevel modelling was used in Stage 3 to predict Year 3 outcomes using individual AEDC items. The final stage used qualitative techniques to align the AEDC item progression to the Australian Curriculum and learning progressions.

The report builds on previous analysis demonstrating the relationship between AEDC and the reading and numeracy skills in Years 3, 5, 7 and 9, when used in conjunction with NAPLAN data (Brinkman et al., 2013) going one step further by including NAPLAN Writing as part of the analysis. The AEDC is used to track Australian children's development, strengthen communication across ECEC and school settings, and to identify opportunities for working collaboratively. This report demonstrates the alignment of the AEDC with the Australian Curriculum and the National Literacy and Numeracy Learning Progressions and the potential usefulness of such alignment for teaching and learning. This means that when AEDC reveals that a child is not yet able to demonstrate a skill, the teacher may locate that skill within the Australian Curriculum and National Literacy and Numeracy Learning Progressions, to ensure it is explicitly taught.

1 We acknowledge that in schools learners are referred to as 'students' but for the purposes of this report we will refer to learners in both ECEC and schools as 'children'.

2 This report acknowledges that the AEDC full domain names are Physical Health and Wellbeing, Social Competence, Emotional Maturity, Language and Cognitive Skills (school-based),Communication Skills and General Knowledge. Where possible we have used the full names, however there are some instances in the report where the shortened version 'Language and Cognitive Skills' and 'Communication' were used for editorial, graphical and syntactical purposes.

Key findings

The key findings of the analysis presented in this report are:

- All 5 domains assessed by the AEDC are related to Years 3 and 5 NAPLAN performance. The skills children are already able to demonstrate at the beginning of school significantly predict their performance 3 years later. Children's performance in NAPLAN at Year 5 can be traced back not only to their performance in Year 3, but to 5 years earlier at the start of school.
- The Language and Cognitive Skills domain of the AEDC has the strongest positive association with children's academic performance in Year 3, with additional and sustained impact into Year 5. The Language and Cognitive Skills domain consists of 4 subdomains: Basic Literacy, Advanced Literacy, Basic Numeracy, and Interest in Literacy, Numeracy and Memory. Performance in Basic Literacy and Basic Numeracy significantly explained Year 3 Reading outcomes with small to medium effect sizes.
- A set of 8 skills from the Language and Cognitive Skills domain was shown to have a strong correlation with NAPLAN performance in Year 3 across Numeracy, Reading and Writing. The analysis showed that children entering primary school who had acquired these skills were more likely to experience future literacy and numeracy success throughout the primary years. These 8 skills are: Recognise numbers 1 to 10; Read complex words; Count to 20; Remember things easily; Attach sounds to letters; Use one-to-one correspondence; Write his/her own name in English; Aware of writing directions in English.
- The early skills that predict later high performance are not all advanced skills. Of the 26 skills within the AEDC Language and Cognitive Skills domain, the top 5 found to have the strongest correlation with each NAPLAN test domain were not always the most difficult skills. This highlights the need for achieving basic skills in order to progress to more advanced literacy and numeracy skills. (See Figure 1).
- The progression of early literacy, numeracy and cognitive skills, as derived from the analysis of AEDC Language and Cognitive Skills domain items, aligns with the Australian Curriculum and the National Literacy and Numeracy Learning Progressions (NLNLP). This means that, apart from minor exceptions, the pattern of the acquisition of the foundational skills identified by the AEDC follows the same pattern identified by the Australian Curriculum and the NLNLP.
- The skill indicators in the AEDC Language and Cognitive Skills domain do not distinguish between developmental levels as well as the indicators belonging to the other domains. This finding is important as further analysis highlights that, for children at the high end of the developmental scale, the AEDC is a 'blunt instrument' and does not provide a granular indicator of more advanced literacy and numeracy skills.

Figure 1: Strongest indicators of future Numeracy, Reading and Writing performance

Numeracy performance

- Recognise numbers 1 to 10
- Read complex words
- Count to 20
- Remember things easily
- Attach sounds to letters

Reading performance

- Read complex words
- Recognise numbers 1 to 10
- Remember things easily
- Attach sounds to letters
- Count to 20

Writing performance

- Read complex words
- Recognise numbers 1 to 10
- Use one-to-one correspondence
- Write his/her own name in English
- Aware of writing directions in English

Implications

The findings of this research have implications for teachers and policymakers, which include the following:

- Given that early skills development, particularly in Language and Cognitive Skills and in Communication Skills and General Knowledge, has been found to indicate a significant influence on a child's schooling achievement, there is potential for a greater focus on targeting these skills in the early years of schooling.
- It has long been highlighted that children's learning and development in the years prior to school is critical; in particular, the development of the 'whole child' is important to build language, literacy, and social-emotional skills. With this in mind, supporting children's development in the range of domains included in the areas of Social Competence, Emotional Maturity and Physical Health and Wellbeing is also important. These skill sets contribute to children's abilities in language and cognition and communication skills that directly impact later literacy and numeracy achievement.

1. Introduction

This report is a research priority as part of the Australian Education Research Organisation's (AERO) 'Literacy and numeracy project' and will present key insights from the analysis of the Australian Early Development Census (AEDC) and the National Assessment Program Literacy and Numeracy (NAPLAN) nested in the Longitudinal Study of Australian Children (LSAC) dataset. This project is part of AERO's vision for Australia to achieve excellence and equity in education outcomes for all children and young people through effective use of evidence. The Literacy and numeracy project aimed to identify the specific areas of literacy and numeracy where children and young people need the most support and to generate practical, evidence-based resources to enhance the teaching and learning of those skills. This report describes the research related to early childhood development. Reports and resources from across the entire project are available from AERO's website.

This section of the report introduces the project with a discussion of its context, aims and research design, as well as the outputs of the project.

1.1. Context

The early years of a child's life from birth to age 8 are a critical period for learning and development. Experiences and interactions in early childhood have a great influence on a child's cognitive, physical, social and emotional development (Saunders et al., 2022), with evidence showing that children who have a strong start in their learning and wellbeing will have better outcomes as they grow older (Kral et al., 2021; Organisation for Economic Co-operation and Development [OECD], 2020). High-quality early childhood education and care (ECEC), and continuity from ECEC and into school, provides children with a strong foundation for lifelong learning.

Development within the first 5 years of life has greater impact and occurs more rapidly than at any other time across the lifespan (OECD, 2018). A key acknowledgement of this in the Australian context is found in the Alice Springs (Mparntwe) Education Declaration (Education Council, 2019), which highlights the significance of access to high-quality education and care in the early years in supporting positive outcomes for children that extend across their lifespans. Internationally, research reinforces that ECEC is of particular value to children experiencing disadvantage as it improves their opportunities and outcomes (Australian Children's Education & Care Quality Authority [ACECQA], 2020; OECD, 2018).

The AEDC provides a periodic snapshot of how Australia's children are faring as they start their first year of full-time school. The AEDC occurs nationally between May and July (Terms 2–3) and represents a commitment to understanding how children are developing from both vulnerability and strengths-based perspectives, to ensure that all children start primary school in a position of equal opportunity.

In a similar way, the NAPLAN assessments represent a regular snapshot of educational outcomes across Australia. NAPLAN measures children's learning progress in literacy and numeracy and is managed by the Australian Curriculum, Assessment and Reporting Authority (ACARA), in collaboration with representatives from all states and territories and non-government school sectors. NAPLAN consists of literacy and numeracy tests (reading, writing, language conventions and numeracy) taken in Years 3, 5, 7 and 9 each year. The intention of the tests is to understand whether children are achieving the ‘fundamental literacy and numeracy skills that every child needs to become successful learners in school and beyond’ (ACARA, 2022). The first NAPLAN tests were administered in 2008. The tests have continued every year, apart from 2020, when they were cancelled due to the COVID-19 pandemic.

This study provides key insights into the crossover between ECEC and learning at school, with a focus on literacy and numeracy. Being literate or numerate involves being able to explore, understand and apply language, reading and writing or mathematical skills respectively in everyday life (MacDonald, 2018; Serpell et al., 2004; Yelland et al., 2014).

The Early Years Learning Framework (EYLF, V2.0), which is the national early childhood curriculum framework for children from birth to 5 years, provides definitions of numeracy and literacy that are underpinned by effective communication skills and capabilities. Literacy in the early years is contextualised and defined as:

> … the capacity, confidence and disposition to use language in all its forms through written, oral, visual and auditory. Literacy incorporates a range of modes of communication including music, movement, dance, storytelling, visual arts, media and drama, as well as talking, listening, viewing, reading and writing. Active listening and a strong foundation of oral language is key to ongoing and lasting literacy learning. Children enjoy and learn from different texts. Contemporary texts include electronic and print based media. In an increasingly technological world, the ability to critically analyse texts is a key component of literacy. Children benefit from opportunities to explore their world using technologies and to develop confidence in using digital media (Australian Government Department of Education [DoE], 2022, p. 57).

Meanwhile, numeracy in the early years is understood in context and defined as:

> … the capacity, confidence and disposition to use mathematics in daily life. All children bring new mathematical understandings through engaging with problem solving. It is essential that the mathematical ideas with which young children interact are relevant and meaningful in the context of their current lives. Educators require a rich mathematical vocabulary to accurately describe and explain children’s mathematical ideas and to support numeracy development. To build their numeracy, children explore powerful mathematic ideas in their world including spatial sense, geometric and algebraic reasoning, structure and pattern, number sense, data and probability reasoning and measuring, along with drawing connections and argumentation structure and pattern, number, measurement, data argumentation [DoE, 2022, p. 57].

Numeracy in the school context is defined as the knowledge, skills, behaviours and dispositions that children need to purposefully use mathematics in a wide range of situations, as well as understanding the role of mathematics in the world (ACARA, 2017).

Research by He and colleagues (2021) has stated that 'the critical window for the development of early literacy and numeracy skills is before children arrive at school; hence major differences in early academic, learning and behavioural regulation skills can emerge at preschool or school entry' (He et al., 2021, p. 3). This notion of foundational numeracy and literacy skills being developed in the early childhood period before starting school has been further supported elsewhere in the literature (Elek et al., 2022; Harris et al., 2018; Zubrick et al., 2015).

Children's ability levels as they enter school, then, are key factors in setting their academic trajectories and may have implications across the life course (Brinkman et al., 2013). Furthermore, engagement in formal preschool programs has been found to be associated with higher performance in middle-childhood NAPLAN results (Department of Education and Early Childhood Development [DEECD], 2013). Recognising these implications for children's learning presents a unique opportunity to provide support early (He et al., 2021). It also sheds light on the mechanisms of learning through understanding the ways in which the attainment of earlier skills sets a child up for learning more advanced skills. This provides a clearer evidence-based roadmap for improving children's learning outcomes.

The years prior to starting formal schooling have been found to be critical in the development of emergent literacy skills that precede learning to read and write (Elek et al., 2022). AEDC data (2009) indicate that children who attend a preschool program have a lower likelihood of being in the developmentally vulnerable range for 4 out of 5 of the AEDC domains (Language and Cognitive Skills; Communication and General Knowledge; Physical Health and Wellbeing; and Social Competence) (Goldfeld et al., 2016). AEDC data from 2015 identified that 22% of 5-year-old children are developmentally vulnerable, reinforcing the importance of opportunities that foster learning and development in the early childhood years (Taylor et al., 2022).

1.2. Research aims

This study aimed to understand how children's learning and development progresses throughout the early years of school. Specifically, the research examined the language and cognitive skills children have acquired on entry to school, and the relationship of those skills to later literacy and numeracy achievement.

The focus on literacy and numeracy in the context of the early years aligns with and supports AERO's Strategic Plan and Research Agenda, approved by Education Ministers in September 2021. The Strategic Plan, Research Agenda and annual Work Plan are all intended to reflect and enable AERO's vision and objectives; to achieve excellence and equity in educational outcomes for all children and young people through effective use of evidence.

1.3. Research questions

The research questions that guided this study are:

1. Does a child's ability to successfully complete Basic and Advanced Literacy and Basic Numeracy tasks provide an indication of future performance in NAPLAN?
2. How do the responses to the items from the Language and Cognitive Skills domain within the AEDC demonstrate a sequenced progression of early literacy and numeracy skill acquisition on a progressive scale?
3. Does the progression of Language and Cognitive Skills, derived from AEDC data, align to the Australian Curriculum (for English and Mathematics) and the National Literacy and Numeracy Learning Progressions (NLNLP)?

1.4. Project design and outputs

The study described in this report uses mixed methods to analyse pre-existing datasets linked by a common ID that incorporates both AEDC and NAPLAN results, as well as demographic information. Using the linked AEDC and NAPLAN data, the analysis identified the foundational aspects of literacy and numeracy included in the AEDC that are the strongest predictors of future academic performance in reading, writing and numeracy as assessed by NAPLAN. The analysis was also used to establish a platform of basic and advanced literacy and numeracy progression, before linking this to indicators of success in NAPLAN literacy and numeracy performance.

Understanding how children progress from the level of competency they held in their first year of school to their later performance on NAPLAN literacy and numeracy measures can help practitioners identify and target key skills, to support children in reaching their potential. This study provides an opportunity to consider the utility of the AEDC data, not just as a means for system and community *feedback*, but as a means to *feedforward*, as we take into account the key skill indicators of future literacy and numeracy success beyond the school foundation year.

This study also situates the analysis of the Language and Cognitive Skills AEDC domain in the context of the Australian Curriculum and National Literacy and Numeracy Learning Progressions (NLNLP). This alignment provides an opportunity to consider the relevance of this dataset in the context of the classroom. It also outlines the skills that children starting school find easier, compared with those that are identified here as being more challenging.

This document is a final report on the project detailing the methodology and findings. Where necessary, further information is provided in appendices or via links to associated documents. The research findings will be translated in a variety of ways so that the learnings are made meaningful and valuable for teachers and leaders. All outputs are available from the AERO website.

2. Background

2.1. Literacy and numeracy

Early steps in language, communication and literacy include the skills, knowledge and attitudes that a child develops in relation to reading and writing throughout early childhood (Lennox & Westerveld, 2014). These skills develop from birth, and include speaking and listening (exposure to oral language development), understanding that print carries meaning, basic alphabet knowledge, and early phonological awareness (Elek et al., 2022; Nicholas & Rouse, 2021). In a similar way, early numeracy skills involve the basic foundations required for later proficiency in numeracy. This includes building an understanding of mathematical concepts, such as number, pattern, measurement, spatial sense and geometry, as well as data and probability (DoE, 2022; MacDonald, 2018). Early numeracy development also includes building capacity with processes for exploring mathematical concepts and developing dispositions for mathematical thinking and learning (DoE, 2022; MacDonald, 2018, 2015).

Formal schooling nurtures children's literacy and numeracy skills and provides the bridge between children's ability to apply skills learnt in the home and in ECEC settings to those to which they have had no exposure until primary school. Particularly, formal schooling provides an opportunity for all children to move from biologically primary literacy and numeracy skills (for example, vocabulary development and visual-spatial processing), which can be acquired through experiences and responsive interactions, and biologically secondary skills (for example, metalinguistic awareness and symbolic numerical processing), which must be explicitly taught (Huo et al., 2021; Sweller et al., 2019). The Australian Curriculum from the foundation year through the primary years, along with the NLNLP, maps the development of biologically secondary skills to enable literacy and numeracy development.

When children enter primary schooling, there are formal structures and assessments in place to understand if they are acquiring reading, writing and numeracy skills. NAPLAN is a large-scale assessment of literacy and numeracy, the purpose of which is to 'provide governments, education authorities, schools and the community with nationally comparable data about how young Australians are meeting educational outcomes in the key areas of literacy and numeracy' (ACARA, 2017, p.3). In 2017, national key performance measures for schooling were developed (ACARA, 2017), including 'NAPLAN participation, the proportion of students achieving at or above the national minimum standard for NAPLAN reading, writing and numeracy and the mean scale scores' (ACARA, 2017, p. 4). The linking of the AEDC data to NAPLAN provides an opportunity to look at the key skills that children can achieve and whether they are precursors to positive outcomes in reading, writing and numeracy based on the NAPLAN tests.

2.2. Skill development in early childhood

Development from infancy through to later childhood hinges on many skills being nurtured, supported and developed, enabling later, more complex, skills to be built on their foundation. Supporting children to be ready for a positive transition to primary school includes enabling these skills to be learnt. Quality ECEC services provide an opportunity for skill building across both literacy and numeracy for all children, including enabling identification and support for those who are performing below the level expected by normative data (Tayler et al., 2015).

Children's early reading skills have been shown to be dependent on the development of their oral language and phonological awareness; that is, their ability to recognise and manipulate the spoken parts of sentences and words (Nicholas & Rouse, 2021). Emergent literacy refers to the skills, knowledge and attitudes that lead to learning to read and write, and the environment provided to children that supports this learning (Elek et al., 2022). Emergent literacy skills can be developed and extended through immersion in pre-literacy activities (Nicholas & Rouse, 2021). This could include sharing oral language through stories, looking at and listening to books; singing and rhymes; letter-sound knowledge; alphabetic knowledge; word recognition (Campbell, 2020; Chaitow et al., 2022); and early writing skills, such as mark making, writing names, letters and words (Chaitow et al., 2022) both in a child's home learning environment and ECEC programs such as long day care and preschool (Elek et al., 2022).

Pre-literacy and early literacy skills and knowledge can be provided through both child-led and intentional teaching through play-based learning experiences in ECEC (Campbell, 2020). Play-based learning also provides a rich environment that supports brain development and learning, and allows children to explore and make meaning of their world (DoE, 2022). Existing research has indicated that learning experiences that promote children's oral development and pre-literacy skills include 'dramatic play; exposure to music and songs; dialogic shared reading; the collective use of audio-visual technologies to display video, songs or other electronic texts; and exposure to non-verbal communication, such as visual aids and gestures' (Nicholas & Rouse, 2021, 5).

Within the Australian Curriculum, literacy development occurs when children acquire the knowledge, skills and dispositions to confidently use and interpret language for communicating and learning (ACARA, n.d.). Literacy includes 'students listening to, reading, viewing, speaking, writing and creating oral, print, visual and digital texts, and using and modifying language for different purposes in a range of contexts' (ACARA, n.d.). Becoming literate also involves the development of learning dispositions, such as self-sufficiency; working well with others; including ideas and opinions from and about diverse cultures; revising work; and questioning meanings and assumptions in texts (ACARA, n.d.).

Some research has suggested that 'children who start school with poor language, cognitive and emergent literacy skills' (Elek et al., 2022, p. 89) may be unlikely to catch up to their peers in literacy achievement as they get older (Goss et al., 2016; Lonigan & Shanahan, 2008; O'Connor et al., 2019). Zubrick et al. (2015) investigated 'patterns and predictors of children's oral language and literacy abilities at 4, 6, 8 and 10 years' (2015, p. 1) using data from the Longitudinal Study of Australian Children (LSAC). The literacy skills were measured using the Academic Rating Scale: Language and Literacy Subscale (ARS). The study found that 69% of the children maintained the same level of literacy skills and ability with which they started school (Zubrick et al., 2015). This suggests a strong effect of early literacy on literacy at age 10, but also identifies a large (31%) subgroup of children who were observed to have changed literacy levels between the two time-points.

Early development of numeracy is similarly nuanced. Research reinforces that children are born with innate mathematical capabilities, showing spatial awareness and number sense from infancy (Chen et al., 2017; Feigenson et al., 2002; Xu & Spelke, 2000). When this potential for mathematical learning is recognised and supported in the early years, children have better long-term outcomes with numeracy (Chen et al., 2017), however this extends beyond just building knowledge of mathematical concepts and processes. The early childhood years are a time when children build a conceptualisation of themselves as a learner. Research reinforces that when young children develop positive dispositions and attitudes towards mathematics, they show higher levels of achievement (Clerkin & Gilligan, 2018; Perry et al., 2016). Young children learn mathematics more effectively when it is embedded in routines and play-based learning, and this begins in infancy (Bjorklund, 2012; Lee, 2012). Such an approach is at the heart of ECEC pedagogical philosophy and practice and supports numeracy development.

Early literacy and numeracy skills develop holistically and are interrelated with other domains of learning and development. Early strength in mathematics supports the development of reading and writing skills and general achievement levels at school (Duncan et al., 2007). Additionally, regulation of emotion and attention are an important foundation for mathematical thinking and influence mathematical development and later achievement (Williams et al., 2016). These synergies are an important consideration when seeking to understand the importance of early mathematical ability and the most effective ways to support mathematical thinking and learning.

Many of the skills discussed here, particularly those in language, communication and literacy, and numeracy, are incorporated for measurement into the AEDC (the AEDC instrument is described in detail in Section 3.1.1). The AEDC is a population-based measure of children's development that is administered every 3 years to children who are in their first year of full-time school (Australian Government Department of Education, Skills and Employment, 2022). It provides feedback on the development of communities of children within 5 key domains, with a specific focus on tracking areas of developmental vulnerability at the community level. Results are typically reported in terms of the proportion of children that are on track or developmentally vulnerable in one or more domains. In this way, the AEDC has been designed to be used to strengthen communication across ECEC and school settings, and to identify opportunities for working collaboratively to improve children's outcomes.

Recent results contained in the 2021 AEDC report indicated that, for the first time since the AEDC was launched, the percentage of children nationally who were on track in all of the 5 domains decreased (from 55.4% in 2018 to 54.8% in 2021; Department of Education, Skills and Employment, 2021). Specifically, developmental vulnerability in the AEDC increased across 4 out of 5 domains, and inequalities in outcomes according to children's family backgrounds continued to persist. When considering the decrease reported between 2018 and 2021, it is important to note that 2019 to 2021 was a time of significant disruption for ECEC and school attendance for children in Australia due to natural disasters and the COVID-19 pandemic. ECEC services and schools were interrupted with closures and other safety and social distancing measures (ACECQA, 2020). There is an expectation that building a comprehensive understanding of the impact of the pandemic on children's learning may take some time, and also that children from disadvantaged backgrounds will be found to have been the most affected (Sacks et al., 2020).

It is important therefore to understand what flow-on effects might be seen as a child progresses through school, based on their level of developmental vunerability in their starting year. Critically, early assessment of children entering school should be further supported with targeted ongoing formative assessment. Formative assessment is essential to identify prior learning and progress, and to monitor any gaps in learning that need to be addressed. Formative assessment combined with other evidence-based practices such as explicit instruction and mastery learning are the most consistently effective teaching practices to support all children (including vulnerable children) to achieve benchmark literacy and numeracy.

2.3. Linking early childhood to primary school outcomes

2.3.1. Can we anticipate primary school outcomes using knowledge of skills attained in early childhood?

Previous research has demonstrated that early language, communication and literacy skills of speaking and listening, understanding that print carries meaning, basic alphabet knowledge, and early phonological awareness, are all strong predictors of how quickly children will develop, and how well they will read in Years 1, 2 or 3 (Neumann, 2016). Similarly, number knowledge, especially relating to symbolic numbers, has been found to be a very strong predictor of later mathematics success as well as success in other academic domains (Heckman, 2011; Raghubar & Barnes, 2017). Additionally, research has found a positive association between preschool attendance and Year 3 NAPLAN test scores, particularly in the domains of reading, spelling and numeracy (DEECD, 2013).

Research conducted in the United States adapted the Early Development Instrument (the AEDC in Australia and the EDI in Canada and the US) and collected data from 3,000 children in Orange County, California, following them longitudinally from kindergarten to year 3 (Duncan et al., 2020). The study assessed children's progress using the EDI in kindergarten and state assessments in literacy and numeracy in year 3. It found that children's ratings at kindergarten were strongly linked to outcomes and achievements for the children in year 3. The researchers concluded that the study 'demonstrated that the kindergarten assessment of the EDI strongly predicted' (Duncan et al., 2020, p. 294) year 3 proficiency in literacy and numeracy, with the EDI providing a strong indication of future achievement.

In the Australian context, the AEDC domains have been shown to predict later health, wellbeing and academic success (Department of Education and Training, 2019, p. 6). Previous research has indicated that the developmental levels assessed through AEDC are predictive of later reading and numeracy skills in Years 3, 5, 7 and 9, when used in conjunction with NAPLAN data (Brinkman et al., 2013). The strongest predictions appear to be evident between developmental vulnerability on the AEDC Language and Cognitive Skills domain and the NAPLAN Reading and Numeracy tests (Brinkman et al., 2015). Interestingly, a more recent Australian study using LSAC data from 2,018 children highlighted the potential in the early years of primary school for improving children's learning trajectories, with 40% of children in the study who were initially identified as being academically vulnerable being classed as academically resilient by Year 3 (Williams et al., 2022).

AERO's scoping report, *Early childhood data in Australia*, highlights the value of linked datasets such as the Multi-Agency Data Integration Project (MADIP). The scoping report argues that linked datasets could be better leveraged to identify where the system is working well, and where more support is required, adding nuance to decision-making in policy and practice (AERO, 2022). Linked AEDC and NAPLAN data provide a powerful data asset, allowing researchers to analyse children's learning over time, and providing important contextual information about the child's background characteristics when they start school and as they move through schooling.

The research that is reported here builds on the work by Brinkman et al. (2013), by looking at the Language and Cognitive Skills domain of the AEDC in the context of a linked dataset containing all the AEDC and NAPLAN domains. In doing this, the study takes a granular look at the items within the Language and Cognitive Skills AEDC domain specifically, to understand how particular skills might predict later NAPLAN achievement.

2.3.2. Can knowledge of skills gained in early childhood inform strategic support of a child's progressive skill building?

In a recent report on children's mathematical development, Geary (2022) highlighted that, while there has been extensive research relating to the development of young children's mathematics-related skills and knowledge, there was 'relatively little research that links these early developing competencies with mathematical achievement in school' (p. 3). Geary's study sought to address this gap in research by identifying the key early competencies that predict mathematics achievement during preschool and that are foundational for formal learning at school. The report indicated that children who start school significantly behind their peers in fundamental numerical knowledge are at risk of long-term difficulties in mathematics. Given this foundational knowledge emerges during preschool years, such difficulties can also be determined by age 3.5 to 4 years.

In Australia, detailed descriptions of journeys through learning in key content areas are contained in the NLNLP, which describe the observable indicators of increasing sophistication in children's use of Standard Australian English language and key numeracy concepts. The progressions present a conceptual view of literacy and numeracy learning, providing an opportunity to develop and support targeted teaching and learning programs for children (ACARA, 2020). The NLNLP stemmed from the National Partnerships Agreement on Literacy and Numeracy, with the intention of bringing together the resources of the Australian Government, state and territory governments and education sectors to create infrastructures to deliver 'sustained improvement in literacy and numeracy outcomes for all students, especially those who are falling behind' (Australian National Audit Office, 2012).

In 2016, ACARA in partnership with the NSW Department of Education started a national collaborative project to develop the progressions from 2016 to 2017 (ACARA, 2020). The NLNLP were trialled in 2017 and validated against NAPLAN student performance data (where feasible) in late 2016 and early 2017. Version 2 was approved in October 2017 and made available to Australian teachers. Version 3 of the NLNLP was completed in 2020 under the discovery phase of the Learning Progressions and Online Formative Assessment national initiative, a joint project of ACARA, Education Services Australia Limited (ESA) and the Australian Institute for Teaching and School Leadership Limited (AITSL), under the direction of the Education Council. The intent of the NLNLP is to:

> describe the skills, understandings and capabilities that students typically acquire as their proficiency increases in a particular aspect of the curriculum over time. They describe the learning pathway(s) along which students typically progress in particular aspects of the curriculum regardless of age or year level, and are designed to help teachers ascertain the stage of learning reached, identify any gaps in skills and knowledge, and plan for the next step to progress learning [ACARA, 2020].

The NLNLP sit within the General Capabilities, and supplement and underpin the broader framework of the Australian Curriculum. They do not replace the Australian Curriculum (ACARA, 2022).

In examining the skills targeted by the AEDC – particularly in the Language and Cognitive Skills domain – the current study aims to expand the existing evidence base pertaining to the Australian Curriculum (English and Mathematics) and the NLNLP. This study explores what early literacy and numeracy skill progression looks like, as derived from analysis of the AEDC and whether this sequence aligns with the Australian Curriculum (English and Mathematics) and the NLNLP. This is important for several reasons. First, it will help support understanding of where children are situated in their understanding of literacy and numeracy as part of the curriculum, and whether there are opportunities for these children to develop advanced literacy and numeracy skills that may be beyond their year level. Second, it allows governments, communities and schools to gain a better understanding of what AEDC results mean in relation to the Australian Curriculum. They can then see where additional support might be required, particularly for children who are unable to demonstrate these key literacy and numeracy skills, to ensure the best opportunity for future literacy and numeracy engagement and success.

2.4. Summary

Early language, communication, literacy and numeracy ability, knowledge and skills form an essential foundation for successful mastery of literacy and numeracy. This, coupled with conceptual development, is necessary for long-term retention and transference of skills throughout school and beyond. The current study has been designed to provide new insight into the skills that impact future literacy and numeracy success, and to link these to the Australian Curriculum and NLNLP. In doing so, it presents an opportunity for teachers to focus on foundational skills within the curriculum and ensure the continual and targeted development of children's literacy and numeracy skills. Teachers can then make full use of every identified opportunity to build both academically vulnerable and resilient children into confident learners, as they progress through the first few years of schooling.

3. Methodology

The study used a mixed-methods approach to explore the research questions outlined in Section 1.3. Below, we have provided detailed descriptions of the data sources that were used (Section 3.1) and the dataset (Section 3.2). Following this, we describe the statistical methods that were used (Section 3.3) and the research design by which these were applied (Section 3.4) in the context of our data and assessment tools to address the research questions.

3.1. Data sources

3.1.1. Australian Early Development Census

The AEDC is a population-based measure of children's development, taken in their first year of full-time school (Australian Government Department of Education, Skills and Employment, 2022). It is administered between May and July every 3 years by all schools in Australia, with a response rate of over 95% of eligible children nationally across all collection years. Results are released in the year following testing. The intent of the AEDC data is to provide insights at a community level into the developmental vulnerabilities of children, allowing communities to observe changes over time, as different approaches to supporting children and families are adopted. Commonly, AEDC results are reported as the number and percentage of children who are either developmentally vulnerable, at risk, or on track.

Administration of the AEDC for each child involves their classroom teacher answering about 100 questions on what the child is able to do, using their knowledge of the child's learning and development. The questions, also referred to as items, align to 5 key domains of development:

- Physical Health and Wellbeing
- Social Competence
- Emotional Maturity
- Language and Cognitive Skills (school-based)
- Communication Skills and General Knowledge.

As an example, one item from the Language and Cognitive Skills AEDC domain is 'Able to read complex words' (Australian Government Department of Education, Skills and Employment, 2022). The teacher judgement options for each item vary: some items are binary (yes = 1, no = 2), some are on a 3-point scale (poor/very poor = 1, average = 2, very good/good = 3) and some have an option of 'don't know'. Teachers are supported in carrying out their assessment with a 'Guide to completing the Australian version of the Early Development Instrument' (Department of Education, Skills and Employment [DESE], 2021a), which provides examples of how the instructions for the completion of each question are applied. Questions are further supplemented with examples to support teachers in determining whether the child is able to demonstrate the skill. For the question 'Is able to read complex words', children who are judged 'Yes' would be able to 'read a few words of 2 or more syllables (a few of them is enough). The child may read complex words with the use of prompts'; alternatively a 'No' response is 'The child reads no complex words, even with prompts' (p. 17).

For each of the 5 domains, children receive a raw domain score between zero and 10. This is calculated based on teacher responses to the relevant domain questions, with zero indicating children who are the most developmentally vulnerable. More detailed descriptions of the 5 AEDC domains can be found on the AEDC website.

Rather than the raw domain scores, the current study used item responses to all items in all 5 domains of the AEDC, with a particular focus on children's achievement in the Language and Cognitive Skills domain. This study also used the subdomain structure of the Language and Cognitive Skills domain of the AEDC, representing the analytical insights through the lens of 4 subdomains: Basic Literacy, Advanced Literacy, Basic Numeracy, and Interest in Literacy, Numeracy and Memory.

Validation studies have shown that the AEDC is a valid and reliable predictor of children's literacy and numeracy outcomes throughout their primary school years and later in life (Brinkman et al., in press). Further evaluation of teacher judgement of children's skills, according to the AEDC items and supporting documentation, is beyond the scope of this research.

3.1.2. Longitudinal Study of Australian Children (LSAC)

The Longitudinal Study of Australian Children (LSAC) is an Australian sample-based study that began in 2003 (Sanson et al., 2002; Soloff et al., 2005). The intent of the LSAC is to identify the factors that determine children's pathways through life to positive or negative outcomes. It also identifies the factors that influence change in these pathways, especially at key transition points, such as entry into early childhood services or school settings (Sanson et al., 2002). More detail on the LSAC study is available at the Growing up in Australia website.

Two initial cohorts of 5,000 children were randomly selected from 10% of Australian postcodes, with the number of participating children proportional to the population size within each state. The first cohort recruited children in their first year of life (Birth [B] cohort), while the second cohort recruited children aged 4 to 5 years (Kindergarten [K] cohort). The first wave of data collection began in 2004. After initial assessment, follow-up assessment with both cohorts has occurred biannually since 2004. In the most recent collection (wave 9), data were collected from 3,806 respondents, representing a retention rate of 37.7% from wave 1.

The LSAC uses parents as the main respondent at each follow-up point, but, where appropriate, it also includes responses from the child or young person, their carers, and teachers. It uses a range of methods, such as in-person interviews, phone interviews and paper-based questions. The data collected are wide-ranging, covering many aspects of child development.

In the current study, LSAC wave 4 data were used for children in the B cohort who had also completed assessment with the AEDC (Section 3.1.1) in their first year of full-time school, in 2009.

3.1.3. National Assessment Program – Literacy and Numeracy (NAPLAN)

NAPLAN is a census assessment that is administered to Years 3, 5, 7 and 9 cohorts annually across Australia as part of the Australian National Assessment Program. The assessments are designed to measure children's literacy and numeracy skills in a nationally consistent way (ACARA, 2016), such that NAPLAN results can provide an opportunity for schools and systems to understand how educational approaches are working and identify opportunities for improvement. The assessment is also an important document for parents/carers to understand how children are progressing in literacy and numeracy in their community and against national standards (ACARA, 2016). For more details about NAPLAN, see the NAPLAN website.

The NAPLAN assessment is conducted in the first half of each school year, and involves all of Years 3, 5, 7 and 9. NAPLAN is made up of tests in 4 key domains:

- Reading
- Writing
- Language conventions (spelling, grammar and punctuation)
- Numeracy.

In this study, NAPLAN scaled scores[3] from Year 3 and Year 5 were used, where they were available for children who also had been assessed on both the AEDC (Section 3.1.1) and LSAC (Section 3.1.2) assessments.

3.1.4. My School

My School is a repository of data about schools. It includes enrolment numbers, school sectors, an Index of Community Socio-Educational Advantage (ICSEA) pertaining to each school, and geolocations of schools. ICSEA is a measure of a school's advantage calculated by ACARA, which provides an indication of the socio-educational backgrounds of a student population relative to those of other schools. ICSEA is calculated using information about parents' occupation, parents' education, geographical location, and proportion of Aboriginal and Torres Strait Islander students (ACARA, 2014).

School-level information was used in the current study to understand the relevance of such environmental factors and to statistically control for these factors where necessary. Before 2016, My School referred to 4 categories of geolocations, denoting varying degrees of proximity from a state or territory's capital city: metropolitan, provincial, remote, and very remote (ACARA, 2015).[4] AERO's data analysis did not use My School's geolocation categories, instead referring to the terms used in the LSAC data dictionary, which were defined by the Australian Statistical Geography Standard 2011: major urban, other urban, bounded local, and rural (Australian Bureau of Statistics (ABS), 2010).

3 These are scores on the NAPLAN measurement scales, as reported to parents and to schools.
4 Since 2016, 5 categories of geographic location are used to describe school locations: major cities, inner regional, outer regional, remote, and very remote (ACARA, 2015).

3.2. Data

3.2.1. Sample description

The data in this study came from a large sample (*N* = 2,459) of Australian children. On average, at entry to the earliest assessment point (in 2009[5]), children were 5.2 years old (*SD* = 0.4), and almost half (47.9%) of the children were female. About 2.7% were Aboriginal and Torres Strait Islander children.[6] Once at school, more than half of the children (61.9%) attended a school from the government sector, while 21.8% attended a Catholic school and 9.9% attended an independent school. As aforementioned (Section 3.4.1), to describe the geolocation of the children in the sample, we used labels corresponding to the Australian Statistical Geography Standard 2011: major urban, other urban, bounded local, and rural (ABS, 2010). More than half (62.5%) of the children went to school in a major urban area, 21.3% had other urban area schools, 5% were bounded local and 10.8% were rural. English was noted as a second language (ESL) for 4.6% of the children.[7] About 4.2% of children were identified as requiring adjustments from teachers and educators due to a physiological health condition, a disability, a mental health condition, or specific support needs.[8]

All data collection was carried out using standard research practices, including informed consent from the Longitudinal Studies Data Strategy Branch to use the *Growing Up in Australia: Longitudinal Study of Children in Australia (LSAC) — General Release 9C1 Australian Early Development Census (AEDC) linked data file* (application #835153).

3.2.2. Linked dataset description

Securing consent from parents for the process of conducting data linkage and full data de-identification was undertaken by data owners before the study began.

At the LSAC interviews, 80% of the parents of the wave 1 LSAC B cohort (4,110 study children) returned consent forms for the LSAC-AEDC data linkage. The eligible LSAC sample for data linkage, however, reduced to 2,765 children whose parents provided valid consent, and who participated in the study at the time of linkage and started the first year of full-time primary school in 2009. The LSAC-AEDC data integration and linkage process successfully linked 2,459 children from the eligible cohort.

5 The study began with on-entry children before the launch of the Australian Curriculum in 2010; however, children in this study completed the Year 3 NAPLAN in 2012 and Year 5 NAPLAN in 2014.

6 The term 'Aboriginal and Torres Strait Islander' is used throughout this document to correspond with the derived binary Aboriginal and Torres Strait Islander variable in the 2009 AEDC data dictionary. The question in the 2009 wave of data collection eliciting this information asked families to select the option that best described them from: (1) Aboriginal but not Torres Strait Islander origin, (2) Torres Strait Islander but not Aboriginal origin, (3) Both Aboriginal and Torres Strait Islander origin, (4) Neither Aboriginal nor Torres Strait Islander origin, or (5) Not stated or unknown. Responses to this question informed the derived binary variable of Aboriginal and Torres Strait Islander.

7 The term 'English as a second language (ESL)' is used throughout this document to reflect the language used in the 2009 AEDC collection. The question used in the AEDC data collection was 'Is the child considered ESL (English as a Second Language)?'.

8 The 2009 wave of AEDC data collection used the label of 'special needs' to refer to a large array of conditions that families identified as potentially impacting on children. 'Special needs' included many different health conditions as well as developmental differences. This report uses the term 'requiring adjustment' instead of 'special needs' as it is a preferable term that broadly includes a wide range of circumstances leading to specific support needs.

When the NAPLAN data was linked to the LSAC-AEDC dataset, some further attrition occurred. At Year 3, full data was available for 2,306 children in Reading, and for 2,307 children in Numeracy. At Year 5, the number of complete records reduced to 2,230 for Reading and 2,214 for Numeracy.

The final step in data linkage added school-level variables from My School data, which reduced the available sample size to *N* = 1,927. Where applicable, multiple imputation of missing data was conducted to secure more rigorous analysis results with the whole data (*N* = 2,459).

3.2.3. Data preparation

AERO analysis shows that, where NAPLAN results were missing for LSAC-AEDC records, they were not missing completely at random.[9] Out of the 2,451 LSAC-AEDC records, 47 children who had missing Year 3 NAPLAN scores on average had a lower domain score for all five AEDC domains compared with their peers with non-missing Year 3 NAPLAN scores. Multiple imputation (MI) is a practical and principled statistical method to reduce bias from missing data (Little & Rubin, 1987; Schafer, 1999). For this analysis, multiple imputation using chained equations (MICE) was conducted and 5 imputed datasets[10] were generated. MICE operates under the assumption that, after controlling all of the observed variables in the imputation model, any remaining missingness is completely random (Graham, 2009). In the MICE procedure a series of regression models are run, whereby each variable with missing data is modelled conditionally upon the other variables in the data. This means that each variable can be modelled according to its distribution; for example, the categorical school sector variable can be modelled using multinomial regression and continuous NAPLAN score variables can be modelled using linear regression.

The MICE models used in this analysis produced 5 sets of plausible values for the missing data in these variables: NAPLAN scores across the domains, the state where the children resided when undertaking the NAPLAN tests, school ICSEA and school sector variables. Other independent variables with complete data, as well as the AEDC domain measurement scores from the Rasch analysis,[11] were also incorporated in the imputation models, including gender; whether children were Aboriginal and Torres Strait Islander peoples; whether they spoke English as a second language (ESL); and whether children were requiring adjustments to accommodate learning with health conditions, disabilities, or specific support needs. The imputed values for the missing data and the observed data were combined to form the dataset (referred to below as the 'imputed dataset'). The imputed dataset was then used in all analyses to manage potential issues with bias on the statistical model parameter estimates stemming from missing data. Results generated from the imputed datasets were also compared with the results generated using original data.

9 LSAC weights were not used in this analysis.

10 The number of imputed datasets to create has been recommended up to 40 to increase power, particularly for data with high missing rates, while the feasibility for large samples and complex data is a factor in the planning of imputation (Azur et al., 2011; Graham et al. 2007). Because of the relatively low missing rate of 1.9% the conventional number of 5 datasets were created through the MICE technique in the analysis. Supplementary analysis was conducted with 40 imputed datasets and the results were similar and no substantial power fall was suggested for small effect sizes of the candidate explanatory variables in the model.

11 See Section 3.3.2 for the description of the Rasch modelling.

A total of 368 children had missing data on their schools. In addition, most schools (*N* = 1,445; 63%) were only represented by one child in the dataset, while the rest had 2 to 14 children per school. Nevertheless, the proportion of total variability that was found to pertain to between-school differences (intraclass correlations) was about 10%, indicating some interdependence of the children's data nested within schools. Therefore, when required, multilevel analysis was conducted in parallel with single-level analysis so that school-level variability could be adequately incorporated.

3.3. Statistical methodologies

This section briefly explains the types of statistical methodologies used and how they are best conceptually understood. The section on design (Section 3.4) then connects these methodologies with the data and research questions being explored in the current study.

3.3.1. Structural equation modelling (SEM)

Structural equation modelling (SEM) is a set of statistical techniques used to explore variables that are not directly measurable, and to examine explanatory pathways between variables.

A variable that is not directly measurable is referred to as a *latent variable*, an example of which could be ability in literacy or ability in numeracy. To get an index of how well developed such an ability might be, test items relevant to the ability, referred to as *indicators*, are administered. In simple terms, SEM uses the shared variance of these *indicators*, or the extent to which performance on the items varies together, to define their measurement of the *latent variable* of ability.

Explanatory pathways between variables can be estimated using either latent variables or *observed variables* – variables that are directly measurable, like age or years of schooling. Pathways between variables can be expressed as either *direct* paths or *indirect* paths. *Direct* paths pertain between variables that are related without the involvement of any other variable; for example, Year 3 literacy might be directly related to Year 5 literacy, which, conceptually, would mean that a large part of children's literacy ability in Year 5 could be anticipated by understanding what their literacy ability was in Year 3. *Indirect* paths refer to pathways between 2 variables that travel through a third (or additional) variable, called a *mediator*. In terms of the example of literacy, it could be found that Year 3 literacy ability was related to engagement in literacy learning in Year 4, which in turn was related to Year 5 literacy ability. This would be an example of an *indirect* pathway between Year 3 and Year 5 literacy, through the *mediator* of Year 4 engagement. Accordingly, in SEM, multiple outcome variables can be estimated in a single model while also being examined as potential mediators.

All paths between variables are numerically estimated in SEM, together with an index of whether they show a substantial link; referred to as *significance*. In simple terms, the further the pathway coefficient is away from zero, the stronger the effect that is being estimated. Positive path coefficients indicate that higher values on the predictor variable are associated with higher values on the outcome variable, while negative path coefficients indicate that lower values on the predictor variable are associated with higher values on the outcome variable. Importantly, each path is estimated while controlling for the effects of the other paths included in the relevant section of the model; when a path is significant, then, the relationship between the variables is described as being independent of the effects of the other variables modelled.

Results from SEM analyses are depicted graphically. *Latent variables* are drawn with ovals, *observed variables* are drawn with rectangles, and arrows between variables show *pathways*, with numerical estimates of these written alongside. Pathways that were not found to be *significant* are either drawn as dotted lines or are left out of the picture for visual simplicity. *Fit indices* for models are also reported, which broadly provide an indication of how well the model explained the data. While a number of such indices have been developed (Byrne, 2013), those that are used here are:

- Chi-square (χ^2) and degrees of freedom (*df*) with *p-value*. A non-significant χ^2 ($p > 0.05$) indicates good fit, but as this is undermined by large sample sizes the list of indices below is used as well.
- Root Mean Square Error of Approximation (*RMSEA*). Values close to zero indicate good fit, while values over 0.08 indicate poorer fit.
- Comparative Fit Index (*CFI*). Values approaching 1 or over 0.95 indicate good fit.
- Tucker Lewis Index (*TLI*). Values approaching 1 or over 0.95 indicate good fit.
- Standardised Root Mean Square Residual (*SRMR*). Values close to zero indicate good fit, while values over 0.08 indicate poorer fit.

This study makes use of SEM to explore relationships between AEDC domains and later NAPLAN outcomes, while also incorporating demographic variables pertaining to children and schools.

AERO conducted single-level as well as multilevel SEM analyses, which were able to take into account the nested structure of the data (see description of data in Section 3.2). Results presented in Section 4.1 are those from the single-level SEM analysis, as they are very similar to those from the multilevel SEM analysis. Multilevel SEM analysis results are available on request.

3.3.2. Rasch modelling

Rasch modelling is a psychometric method that can be used to examine whether a set of items cooperate with each other sufficiently to measure a single underlying construct, and to estimate the relative difficulties of these items. At the same time, the person ability estimates within the Rasch model reflect the relative competency levels of a cohort of test-takers. The item difficulties and person ability estimates are placed on the same measurement scale (Wright & Masters, 1982).

In Rasch modelling, it is assumed that a single dimension of measurement underlies performance by test-takers on test items. This single dimension is called the *Rasch dimension*, and can be modelled according to the underlying ability being tested by the test items (sometimes referred to as *indicators*), or according to the underlying ability demonstrated by the test-takers (sometimes referred to as *persons*). So, for example, performance on a test comprising questions about addition would be assumed to measure a single underlying ability of addition – the *Rasch dimension* – with *indicators* testing this ability at differing levels of difficulty and *persons* displaying this ability at different levels of capability. A measurement tool that is well suited to describing the population it assesses should show a distribution of *item difficulties* that covers the range of the *person abilities*. That is, enough test items with a corresponding range of difficulty levels are required to assess test-takers' abilities from the lowest to the highest. The alignment between the item difficulty ranges and test-takers' ability ranges suggests whether the measurement tool is adequately targeting the cohort of test-takers.

When there is evidence that items function well to define a single latent ability, Rasch analysis not only can place items along the measurement scale in terms of their relative difficulty, but the order of items also shows the order in which the skills involved in the ability are acquired. So, in terms of the addition example, item difficulties in a Rasch model would show not only that it is easier to calculate 1+1 than it is to calculate 123+45; but also, that a person typically would need to know how to calculate 1+1 before being capable of calculating 123+45.

The final key aspect of item difficulty estimates that is of relevance here is that they are presented on a measurement scale, which shows not only the order of the items' difficulties but the gaps between them. The ability to space items out in a scientific way, and show differential distance between them, is an aspect of Rasch modelling that differentiates this measurement approach from other modelling techniques and from raw scores. The measurement scale allows us to quantify a child's learning leap on the underlying proficiency scale, which indicates their progression from a lower skill level to the adjacent higher skill level. This type of evidence is often used to understand the increase in the complexity of the knowledge or skill when moving from mastering one skill to the next.

A series of additional analyses are carried out in the context of a Rasch analysis to ensure the Rasch model adequately describes the data. First, the unidimensionality of the domain is tested by examining item fit and performing a *Principal Components Analysis (PCA)* on the standardised item residuals (differences between observed and predicted item scores divided by the variance of the item scores) from the Rasch model. If the Rasch model explains the variance in the data adequately, these residuals should reflect nothing more than random variation. The PCA extracts factors from the covariation of the item residuals, which, if substantial, would show that the unidimensionality assumption does not hold. Specifically, 3 criteria are used to evaluate unidimensionality (Fisher, 2007; Linacre, 2009):

- At least 50% of the *total variance in the response data* should be explained by the Rasch dimension (item and person)
- The *variance explained by the first contrast* from the PCA should be less than 15%; and
- The *eigenvalue of the first contrast* from the PCA should be less than 2.

Two further reliability and separation statistics are used to evaluate the AEDC instrument:

- The *person reliability coefficient* is a statistic showing the extent to which it would be expected that the person and item locations would be reproduced under consistent conditions. Values range between zero and 1, with a person reliability of 0.8 representing the minimum preferable value (Fisher, 1992); and
- The *person separation index* gives an estimate of the spread of individuals along the continuum of ability and reflects the number of distinct groups in which the sample or items can be divided (Bond & Fox, 2015). It indicates whether the test can distinguish between high and low performers. This index has a value starting from zero, with values of 2 and higher reflecting a reasonable level of separation (Fisher, 1992).

Additional analyses for inspecting item fit includes examination of Item Characteristics Curves (*ICCs*) and item fit statistics. Collectively, this body of evidence informs the assessment of whether the Rasch model describes the data well and whether the indicators distinguish low-achieving and high-achieving children well.

Where there is evidence against the data being able to be explained by a single Rasch dimension, an additional modelling method can be used. *Multidimensional item response theory (MIRT)* models broadly speaking extend the logic of the Rasch model to the multidimensional case. However, some further distinctions are introduced in that the between-item *MIRT* model assumes that each item belongs to only one subscale (for example, subdomains in AEDC), and that there are no items in common across subscales.

This study sequentially used the procedures described above, starting from the testing of unidimensionality to the application of *MIRT* when warranted, to investigate the 5 domains of the AEDC. Conceptually, this allowed for the differing latent abilities of children to be accurately measured and for the relative difficulties of the test items to be identified. In addition, it allowed for the extent to which the AEDC domains can be described by Rasch dimensions to be demonstrated.

3.3.3. Multilevel modelling

The aim of multilevel modelling is to explore relationships between variables. To do this, multilevel modelling makes use of natural nested structures in datasets. For example, individuals live within households, that are placed within communities, that feed into schools, that sit inside of states. Given that children's developmental and learning outcomes within a particular level (for example, school) are usually more similar, multilevel modelling can be used to control for this similarity such that the estimated associations between the factors and outcomes are not biased by the effects of these different levels. At each of these levels, systematic factors are at work that can influence children's learning. Placing variables at different levels according to where they hierarchically exist in a structure like this allows for effects to be explored on particular levels and across levels while controlling for the effects pertaining to other levels.

Before performing multilevel modelling analysis, it is important to look at the strength of pairwise relationships between the variables of interest. This is done through *bivariate correlations*. Variables that are highly correlated with other variables but weakly correlated with the outcome variable would be considered for exclusion from the analysis to avoid any bias on the regression coefficients. Resulting correlation coefficients range between -1 and 1, with weaker relationships indicated by a correlation coefficient close to $r = 0$. Positive correlations indicate that high scores on one variable in the pair are associated with higher scores on the other. Negative correlations indicate low scores on one variable in the pair correspond to higher scores on the other.

The multilevel modelling analysis is run after the correlation analysis. While many different aspects of the model are investigated to ensure appropriate model fit, this report focuses on the interpretation of the regression coefficients; specifically, *unstandardised regression coefficients* and their *effect sizes*. These can be understood as follows:

- *Unstandardised regression coefficients (b)* show the number of units by which the outcome variable changes when scores on the predictor variable increase by one unit. So, for example, in the models used here, unstandardised regression coefficients will show the number of scale score points changed on the NAPLAN scale when the AEDC indicator moves up to the next score category. This result is true, on average, after applying all of the other levels of statistical control that have been included in the model – a mathematical equivalent of saying 'all else being equal'.
- *Effect sizes* are a way of understanding the strength or significance of the regression coefficients. In the current context, the effect size for an AEDC indicator is defined as its regression coefficient relative to the standard deviation of the NAPLAN outcome for the 2012 Year 3 cohort. In the context of this study, where there are multiple factors influencing outcomes, we have considered a conventionally small effect size of 0.2 (Cohen, 1988) or higher to indicate a noteworthy effect.

3.4. Design

This section describes the design of the current study. The data analysis for this project was conducted in 4 stages. The aim of Stage 1 was to explore how AEDC domains at school entry together with background variables can predict NAPLAN results for literacy and numeracy in later years (Research Question 1). Stage 2 used Rasch modelling to understand the sequencing of skills measured in the AEDC on a developmental continuum (Research Question 2). In Stage 3, we extended on Stage 1 analyses by investigating the way in which the individual indicators of the AEDC Language and Cognitive Skills domain can predict NAPLAN results for literacy and numeracy while controlling for other demographic and background variables (Research Question 1). Finally, Stage 4 built on Stage 2 and Stage 3 to explore whether the sequencing of skills found in this study aligned with the Australian Curriculum year levels and the National Literacy and Numeracy Learning Progressions (NLNLP) sub-elements (Research Question 3). Each of these stages of analysis is detailed below.

3.4.1. Stage 1: Predicting Years 3 and 5 literacy and numeracy using AEDC domains

The aim of Stage 1 of the analysis was to understand the predictive pathways from the AEDC domains (Physical Health and Wellbeing, Emotional Maturity, Social Competence, Language and Cognitive Skills, and Communication Skills and General Knowledge), representing skills attained in early childhood, to Reading and Numeracy results in Years 3 and 5, as assessed by NAPLAN. In doing this, a range of demographic groups were considered, including gender; Aboriginal or Torres Strait Islander children; children who spoke English as a second language (ESL); and children requiring adjustments due to health conditions, disability, or specific support needs.[12] School-level information was also included: the school Index of Community Socio-Educational Advantage (ICSEA), school type (government as compared to Catholic or independent), and school geolocation (major urban as compared to other urban, bounded local, or rural).

12 Described in the dataset as 'special needs'.

Structural equation modelling (SEM; described in Section 3.3.1) was used throughout Stage 1 of the analysis. All modelling in Stage 1 used the multiple sets of imputed data, described in Section 3.2.3. Models were selected to address Research Question 1 based on theories and previous evidence, and guided by statistical methods and indices. Model competition processes guided the choice of the final model that best fit the data parsimoniously, with the least number of variables and paths to explain the largest amount of variance in the outcome variable. The 3 different SEM models are described below.

3.4.1.1. Models predicting Year 3 NAPLAN outcomes

The first set of models was constructed to predict NAPLAN Year 3 results for Reading and Numeracy respectively. The initial analysis used the raw domain scores from the AEDC data as well as all of the demographic and school variables listed above as the predictors. After Rasch modelling was conducted for each of the 5 domains (see Section 3.4.2), estimates of the child's achievement in the 5 AEDC domains using Rasch modelling replaced these raw scores as predictors (Physical Health and Wellbeing, Emotional Maturity, Social Competence, Language and Cognitive Skills, and Communication Skills and General Knowledge) in the final models. This is because estimates from the Rasch analysis were scientific and more rigorous estimates of the child's achievement in each domain, as the measurement analysis carefully considered the order and gap of the difficulty level of the AEDC indicators and the measurement error associated with the estimate. From this point onwards, AEDC domain score refers to the latent person estimate on that domain from the Rasch modelling.

In the design of the SEMs, demographic and school variables were used to predict achievement in the AEDC domains as well as the NAPLAN outcome directly. Achievement in some AEDC domains were used to predict one another, and achievement in all AEDC domains were used to predict the NAPLAN outcome directly. Therefore, both direct and indirect (via individual AEDC domain) pathways were explored from demographic and AEDC variables through to predicting the NAPLAN outcome. The results for these models are presented in Section 4.1.1.

When compared with the model using raw data with missing values, notable differences in these SEM models using imputed data included the increased explained variance in the outcome from 26% to 30% in Reading. Furthermore, most coefficients of equity variables became larger. These results confirmed that the data were not missing completely at random and children with disadvantaged backgrounds were less likely to participate in NAPLAN tests or to be tracked through the LSAC period.

3.4.1.2. Models predicting Year 5 NAPLAN outcomes

The second set of SEM models expanded from what has been described above to also include Year 5 NAPLAN results in Reading and Numeracy respectively. For these models, the variables being predicted are the Year 5 NAPLAN outcomes. Demographic and school variables were used to predict the achievement in AEDC domains and Year 3 and Year 5 NAPLAN results directly. Some AEDC domains predicted one another as mediators, as well as both the Year 3 and Year 5 NAPLAN results. And finally, the Year 3 NAPLAN result was used to predict the Year 5 NAPLAN result. In this way, the direct predictive power for all variables on the Year 5 outcomes was tested, together with indirect paths moving through AEDC domains and Year 3 NAPLAN results. The results for these models are presented in Section 4.1.2.

As above (Section 3.4.1.1), most equity group variables showed larger coefficients from SEM models using the imputed datasets as compared with the models that had missing data. Again, missingness and imputation suggested that critical attrition occurred involving children with more disadvantages.

3.4.1.3. Models using 4 subdomains of Language and Cognitive Skills to directly predict Year 5 outcomes

The third set of SEM models built on the second set of models by taking the Language and Cognitive Skills domain of the AEDC and identifying further latent dimensions of 4 subdomains: Basic Literacy, Advanced Literacy, Basic Numeracy, and Interest in Literacy, Numeracy and Memory (Warren et al, 2018). The person ability estimates from the multidimensional item response theory (MIRT) model of these 4 subdomains (see further explanations in 3.4.2.1) were included from this point of the analysis forward in the models in place of the Language and Cognitive Skills domain score. All of the demographic and school variables and other AEDC domain scores were used to predict Year 3 NAPLAN results, which in turn was used to predict Year 5 NAPLAN results, together with the 4 subdomain scores of Language and Cognitive Skills. Therefore, indirect paths to Year 5 NAPLAN results were modelled for all variables through Year 3 NAPLAN results, while direct paths were modelled from Year 3 to Year 5 NAPLAN results and from the 4 subdomain scores of Language and Cognitive Skills to Year 5 NAPLAN results. The results for the third set of models are presented in Section 4.1.3.

As with the models described above, the imputed data model with subdomains of Language and Cognitive Skills showed larger effect sizes when explaining Year 3 and Year 5 outcomes in comparison with models run with missing data.

3.4.2. Stage 2: Understanding the relative difficulty of the AEDC items

The aim of Stage 2 of the analysis was to address Research Question 2 by building understanding of the measurement properties of the instrument for each of the 5 AEDC domains: Physical Health and Wellbeing, Emotional Maturity, Social Competence, Language and Cognitive Skills, and Communication Skills and General Knowledge. This was done using Rasch modelling (techniques are described in Section 3.3.2) and began with an assessment of whether each AEDC domain described a single dimension. Focus then fell, as in Stage 1, on the Language and Cognitive Skills domain, with an investigation of the item and person distributions pertaining to the domain, and with a detailed description of the indicators in order of difficulty.

3.4.2.1. Assessing the unidimensionality of each of the 5 AEDC domains using Rasch modelling

Unidimensional Rasch models (see Section 3.3.2) were fitted to the item responses in each of the 5 AEDC domains: Physical Health and Wellbeing, Emotional Maturity, Social Competence, Language and Cognitive Skills, and Communication Skills and General Knowledge. Principal component analysis (PCA) was performed on the standardised residuals for each Rasch model to check the unidimensionality of each AEDC domain. The metrics described in Section 3.3.2 were used to form a judgement about the likely unidimensionality of each AEDC domain.

As there were some indications of multidimensionality in the AEDC domains, multidimensional item response theory (MIRT) models were used from hereon to estimate the subdomain logit scores. Before applying the MIRT model, exploratory factor analysis (EFA) was conducted on each of the 5 AEDC domains to explore the subdomain structure within each domain. The results of these are not reported here, but they largely confirmed the subdomain structure proposed by the AEDC developer. Hence, a between-item MIRT model was fitted to the item response data from the Language and Cognitive Skills domain to estimate the subdomain scores.

3.4.2.2. Person and item distributions for the Language and Cognitive Skills domain

The exploration of the importance of the Language and Cognitive Skills domain of the AEDC that began in Stage 1 of the analysis was continued as part of this analysis. This was done by investigating the targeting of the indicators in this domain on the study cohort through graphical representation of the person and item distributions from the Rasch modelling, and the degree to which they were aligned.

3.4.2.3. Item difficulty in the Language and Cognitive Skills domain

Item difficulty estimates (in terms of logits) were generated for each item belonging to the AEDC Language and Cognitive Skills domain. As well as showing the relative difficulty of each item to the others, this procedure provided evidence of the sequence by which skills are developed in this domain (see Section 3.3.2 for further explanation of this). Note that the sequencing of the skills in the other domains of the AEDC was also analysed and the results can be shared with readers who are interested in understanding the developmental spectrum of these domains.

One key aspect of item difficulty estimates that has been focused on here is the spacing of these scores on the measurement scale, quantifying the children's learning leaps on the underlying proficiency scale. This type of evidence has been used to understand the increase in the complexity of the knowledge/skill moving from mastering one indicator skill to another.

3.4.3. Stage 3: Predicting Year 3 literacy and numeracy using individual AEDC indicators

The aim of Stage 3 analyses was to connect the understanding gained about items belonging to the AEDC Language and Cognitive Skills domain in Stage 2 with the SEM analyses predicting NAPLAN outcomes in Stage 1. This was done using multilevel modelling (see Section 3.3.3) and provided a second set of results to address Research Question 1.

The first step of this analysis was to use bivariate correlations to look at the strength of relationships between each Language and Cognitive Skills item, or indicator, and each of the Year 3 NAPLAN outcomes: Reading, Writing and Numeracy. This bivariate correlation analysis was performed using the non-imputed data, as this was an indicative analysis to understand the strength of relationship associated with each indicator and each NAPLAN outcome.

The second step of the Stage 3 analysis applied multilevel modelling (Section 3.3.3) to investigate the impact of individual AEDC indicators on Year 3 NAPLAN outcomes while controlling for variability linked to schools. Note that the multilevel model for each NAPLAN domain was fitted to the 5 imputed datasets where about 6% of the Year 3 NAPLAN data and 10% of Year 5 NAPLAN data were imputed. The relationships of interest were those pertaining between AEDC Language and Cognitive Skills domain items and the Year 3 NAPLAN results, viewed after other patterns in the dataset had been accounted for through statistical control.

The initial full multilevel models were built including all AEDC indicators that had correlations of greater than 0.2 with Year 3 NAPLAN outcomes. The model was then reduced so that focus could rest on the AEDC indicators that were most significant for predicting NAPLAN results. The first step in reducing the model was to remove indicators that were weakly correlated with Year 3 NAPLAN outcomes ($r < .2$). Second, to avoid known problems with either spurious correlation or multicollinearity, any non-significant indicators that were highly correlated with the significant indicators were also removed from the model. Finally, a likelihood ratio test was performed to determine whether the final reduced model provided a better fit to the data compared with the preceding multilevel models.

A brief summary of the results is outlined in Section 4.3. However, further information on the regression coefficients for the AEDC indicators and demographic variables in the 3 multilevel models for Numeracy, Reading and Writing, are provided in Appendix B.

3.4.4. Stage 4: Aligning the AEDC item progression to Australian Curriculum and learning progressions

The aim of Stage 4 was to address Research Question 3 by connecting evidence about the sequence of skill progression as derived from the AEDC item analysis to current Australian Curriculum and learning progressions. To this end, Stage 4 linked strongly to evidence from Stage 2 Rasch modelling and evidence from Stage 3 regarding the predictive power of specific AEDC items with respect to NAPLAN outcomes.

The analysis was conducted by first using the Stage 2 item difficulty distribution for Basic Literacy and Numeracy and Advanced Literacy to align each item to the Australian Curriculum English and Mathematics. Second, the Stage 2 item difficulty distribution was used to align the items to the National Literacy and Numeracy Learning Progressions (NLNLP) in:

- Reading and Viewing (Phonological Awareness, Phonic knowledge and word recognition, Understanding texts)
- Writing (Creating texts, Spelling, Handwriting and keyboarding); and
- Numeracy (Number sense and algebra [Number and place value, Counting processes], Measurement and geometry [Understanding geometric properties, Measuring time]).

The alignment was conducted by AERO researchers and verified by literacy and numeracy experts. Some of the sub-elements from the NLNLP were necessarily excluded due to the limitations of the AEDC Language and Cognitive Skills domain items.

The key documents supporting this alignment were the Australian Curriculum General Capabilities (ACARA 2022) and the Online Formative Assessment Initiative's *Measurement Scales and Signposts Report* (OFAI 2022).

4. Results

The 4 stages of the analysis were conducted as described in Section 3.4. The results for each of the 4 stages are presented below.

4.1. Stage 1: Structural equation modelling predicting Years 3 and 5 literacy and numeracy using AEDC domains

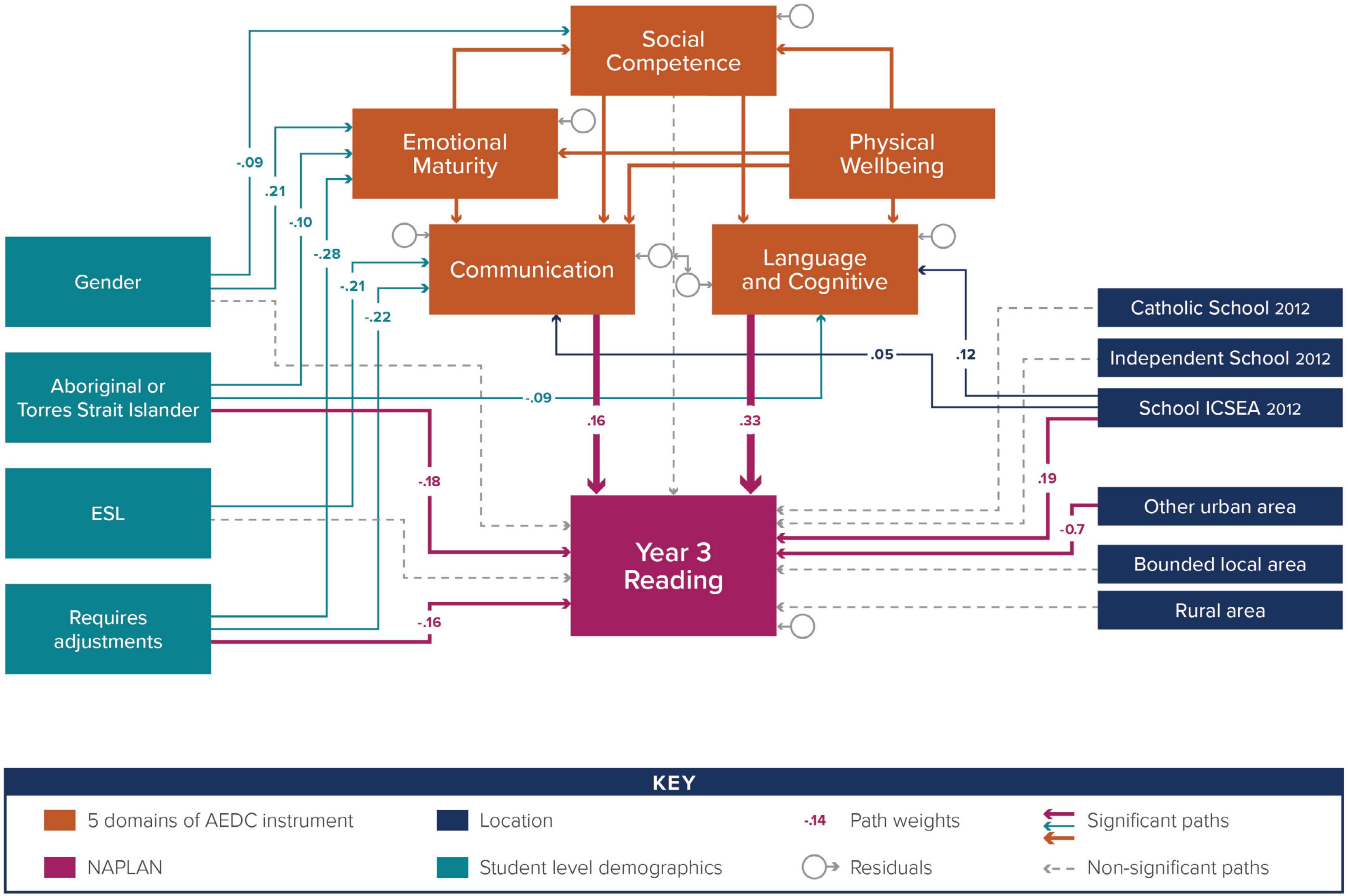
4.1.1. Models predicting Year 3 NAPLAN outcomes

The first set of SEM models predicted Year 3 NAPLAN results using a combination of AEDC domains and demographic and school variables. Results for the models pertaining to Reading and Numeracy were broadly similar, so a detailed description is provided for only one of these: Reading. As shown in Figure 2, the AEDC Language and Cognitive Skills domain score directly predicted Year 3 Reading ($\beta = 0.33$). It did so more strongly than the significant demographic groups of Aboriginal and Torres Strait Islander children; children requiring adjustments; and community socio-educational advantage (ICSEA). This means that children who were identified on the AEDC as having stronger Language and Cognitive Skills achieved better results on their Year 3 NAPLAN test. The results also showed that children from higher ISCEA schools had higher achievement levels on Year 3 Reading in the NAPLAN.

The Communication Skills and General Knowledge domain score was the second strong AEDC predictor of Year 3 Reading ($\beta = 0.16$), and, together with Language and Cognitive Skills, served as a mediator for the other 3 AEDC domains' indirect impact on this outcome. Specifically, Figure 2 shows that Physical Health and Wellbeing has significant pathways to all of the other AEDC domains, meaning that stronger Physical Health and Wellbeing predicted better scores on the other domains. Similarly, higher levels of Emotional Maturity predicted better Social Competence and better Communication. Stronger Social Competence in turn predicted stronger Communication and stronger Language and Cognitive Skills, which both had significant positive predictive relationships with Year 3 NAPLAN scores.

All of the key demographic characteristics – gender; Aboriginal and Torres Strait Islander background; ESL; and requiring adjustments due to health conditions, disability, or specific support needs – together with community socio-educational advantage (ICSEA), had indirect effects on Year 3 Reading via the AEDC domains (see Figure 2). The results for gender indicated that girls had higher levels of Emotional Maturity but lower Social Competence than boys. Aboriginal and Torres Strait Islander children had lower developmental levels for Emotional Maturity and Language and Cognitive Skills. Children who had English as a second language had reduced Communication scores, and children requiring adjustments to accommodate learning with health conditions, disability, or specific support needs had lower scores for Emotional Maturity and Communication. This group also had lower scores for the Year 3 NAPLAN Reading test. Higher ICSEA scores (higher community socio-educational advantage) also predicted higher scores for Language and Cognitive Skills and for Communication Skills and General Knowledge.

Figure 2: Imputed data model – AEDC domain score explaining Year 3 Reading scores when demographic variables were controlled: standardised single-level SEM coefficients



Note. The coefficients are pooled results. Model fit (from first imputed dataset): χ^2 (35, *N* = 2,457) = 70.25, ρ < 0.001; RMSEA = .02; CFI = .98; TLI = .97; SRMR = .01. Green paths show significant prediction of Year 3 NAPLAN results; orange paths depict AEDC mediations; teal and dark blue paths indicate demographic and school variables' explanation of AEDC factors; grey dotted paths indicate non-significant pathways in the model. Gender: Girls were compared to boys. [a]: School sectors were compared with the government sector. [b]: Neighbourhood geolocations were compared with major urban cities.

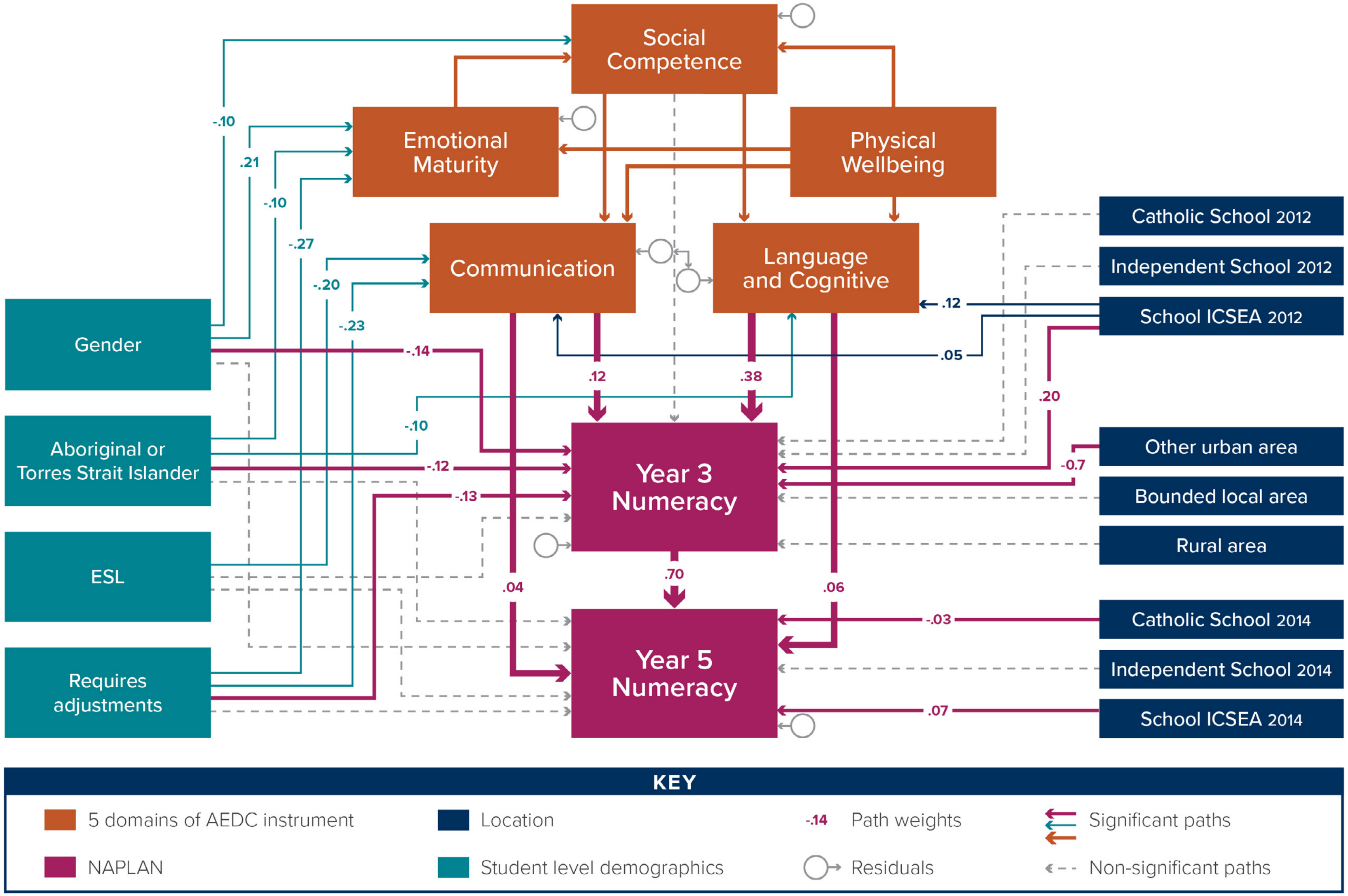
Similar results were found when testing the same SEM model to predict Year 3 Numeracy results. The main difference was that the direct gender impact on the Year 3 outcome was significant ($\beta = -0.14$), suggesting lower scores for girls. Interestingly, the indirect effects of gender indicated two things. First, girls tended to have higher scores for AEDC Emotional Maturity, which via higher Communication in turn predicted higher Year 3 Numeracy outcomes. Second, girls had lower AEDC Social Competence scores, which was associated with lower Year 3 Numeracy outcomes. In total of these various effects, girls were more likely to have lower Year 3 Numeracy results than boys because of the stronger direct gender effect, overriding the weaker positive indirect effect from higher Emotional Maturity and even weaker negative indirect effect of lower Social Competence as indexed by the AEDC measures.

4.1.2. Models predicting Year 5 NAPLAN outcomes

The second set of SEM models predicted Year 5 NAPLAN results using Year 3 NAPLAN results with a combination of AEDC domain scores and demographic variables. The focus in this section is on results for the imputed data model predicting Numeracy, shown in Figure 3. As with the SEM models described above (Section 4.1.1), the AEDC domain of Language and Cognitive Skills and Communication Skills and General Knowledge showed the strongest impacts on NAPLAN Numeracy outcomes, both at the Year 3 and Year 5 levels. Specifically, this group of SEM models showed that while both Language and Cognitive Skills, and Communication Skills and General Knowledge, had an indirect effect on Year 5 Numeracy through the very strong predictive effect of Year 3 Numeracy, they each had an independent direct effect on Year 5 Numeracy as well. This means that an additional part of the downstream impact of skills attained in early childhood (captured by the AEDC) comes through at Year 5, separate from the effects already seen in Year 3.

The other variable to show this type of long-acting incremental effect is the school ICSEA variable. As can be seen in Figure 3, ICSEA indirectly impacted Year 5 Numeracy through its positive prediction of Year 3 Numeracy. An additional effect not explained through Year 3 results was also found through the direct ICSEA pathway to Year 5 Numeracy. This means that community socio-educational advantage impacts children's outcomes at Year 5 in some ways that are captured at Year 3 and some ways that are independent of Year 3 outcomes.

Figure 3: Imputed data model – AEDC domain scores explaining Year 3 and Year 5 Numeracy scores when demographic variables were controlled: standardised single-level SEM coefficients



Note. The coefficients are pooled results. Model fit (from first imputed dataset): χ^2 (59, N = 2,457) = 146.83, $p < 0.001$; RMSEA = .03; CFI = .97; TLI = .95; SRMR = .01.

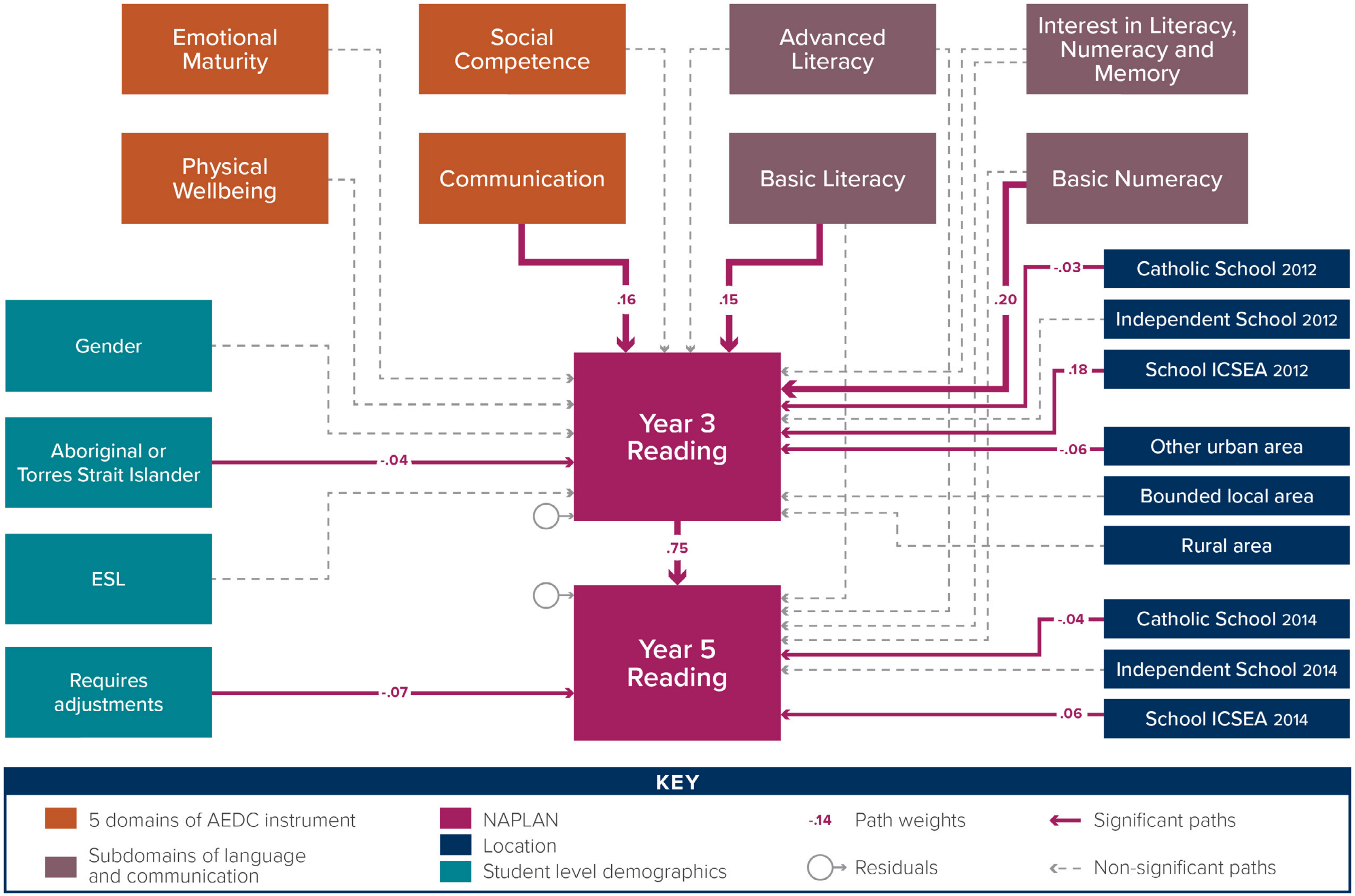
Similar results were found in the above model when it was used to predict NAPLAN Reading results. The only substantial point of difference was that gender was not found to significantly predict NAPLAN Reading results.

4.1.3. Models using 4 subdomains of Language and Cognitive Skills to directly predict Year 5 outcomes

The third set of SEM models used the 4 subdomain scores instead of the AEDC Language and Cognitive Skills domain score as a whole, examining their effects on Year 3 and Year 5 NAPLAN outcomes. They also included the other AEDC domain scores and demographic variables as predictors of Year 3 NAPLAN results. The NAPLAN results for Reading are the focus of the model presented below in this part of the analysis. Of the 4 subdomains, achievement in Basic Literacy and Basic Numeracy significantly explained Year 3 Reading outcomes with small to medium effect sizes (β = 0.15, 0.20 respectively). The achievement in the other 2 subdomains did not significantly explain the scores. However, although the subdomain scores did not have significant direct effects on Year 5 NAPLAN outcomes, their indirect effects on Year 5 scores were significant through Year 3 outcomes.

A SEM model predicting NAPLAN Numeracy results at Year 3 and Year 5 showed similar results. As expected, Basic Literacy showed a smaller coefficient (β = .10) in explaining Year 3 Numeracy scores, whereas Basic Numeracy had a larger effect (β = .25) than in the Reading model. While Communication Skills and General Knowledge had a smaller impact (β = .09) on Numeracy than on Reading, Social Competence showed a small but significant association (β = .06). Female children were likely to have significantly lower Numeracy scores in Year 3 (β = -.14) than their male peers. Finally, unlike what was seen in the outcomes for the Reading models, ESL children were estimated to have higher Numeracy scores (β = .11).

Figure 4: Imputed data model with subdomain scores of Language and Cognitive Skills and other AEDC domain scores explaining Year 3 and Year 5 Reading scores when demographic variables were controlled: standardised single SEM coefficients



Note. The coefficients and model statistics are from the results with the first imputed dataset as pooled results included no standardised coefficients. The other 4 imputed data models showed similar statistics. Model fit: χ^2 (16, N = 2,457) = 49.43, $p < 0.001$; RMSEA = .03; CFI = .97; TLI = .91; SRMR = .004. Year 3 Reading R^2 = .32. Year 5 Reading R^2 = .64.

4.2. Stage 2: Rasch modelling for understanding the relative difficulty of the AEDC items

4.2.1. Assessing the unidimensionality of the 5 AEDC domains using Rasch modelling

Unidimensional Rasch models (see Section 3.3.2) were fitted to the item responses, or scores for each item, in each of the 5 AEDC domains. The appropriateness of this was assessed using principal component analysis (PCA) performed on the standardised residuals.

Table 1 reports the results for each domain. The Rasch dimension explained more than 50% of the variance in the response data for 4 out of the 5 domains; 47.1% of the variance in the data for the Social Competence domain was explained. Only 2 of the 5 domains met the criteria of the eigenvalue of the first contrast being less than 2; Social Competence, Emotional Maturity, and Language and Cognitive Skills all had higher eigenvalues. All 5 domains met expectations with the variance explained by the first PCA contrast falling well below 15%. There are some indications of a secondary dimension for some of the AEDC domains, particularly Social Competence, Emotional Maturity, and Language and Cognitive Skills domains.

Table 1: Summary statistics and PCA results from the unidimensional Rasch models

Summary statistics	Physical Health and Wellbeing	Social Competence	Emotional Maturity	Language and Cognitive Skills	Communication Skills and General Knowledge
Number of indicators	12	24	26	26	8
Percentage of variance in the data explained by Rasch dimension	50.6%	47.1%	51.2%	50.4%	52.0%
Eigenvalue for 1st contrast	1.97	3.50	4.65	2.20	1.63
Variance explained by the 1st contrast	8.1%	7.7%	8.7%	4.2%	9.8%
Person reliability	0.71	0.88	0.87	0.69	0.82
Person separation index (PSI)	1.56	2.71	2.63	1.50	2.11

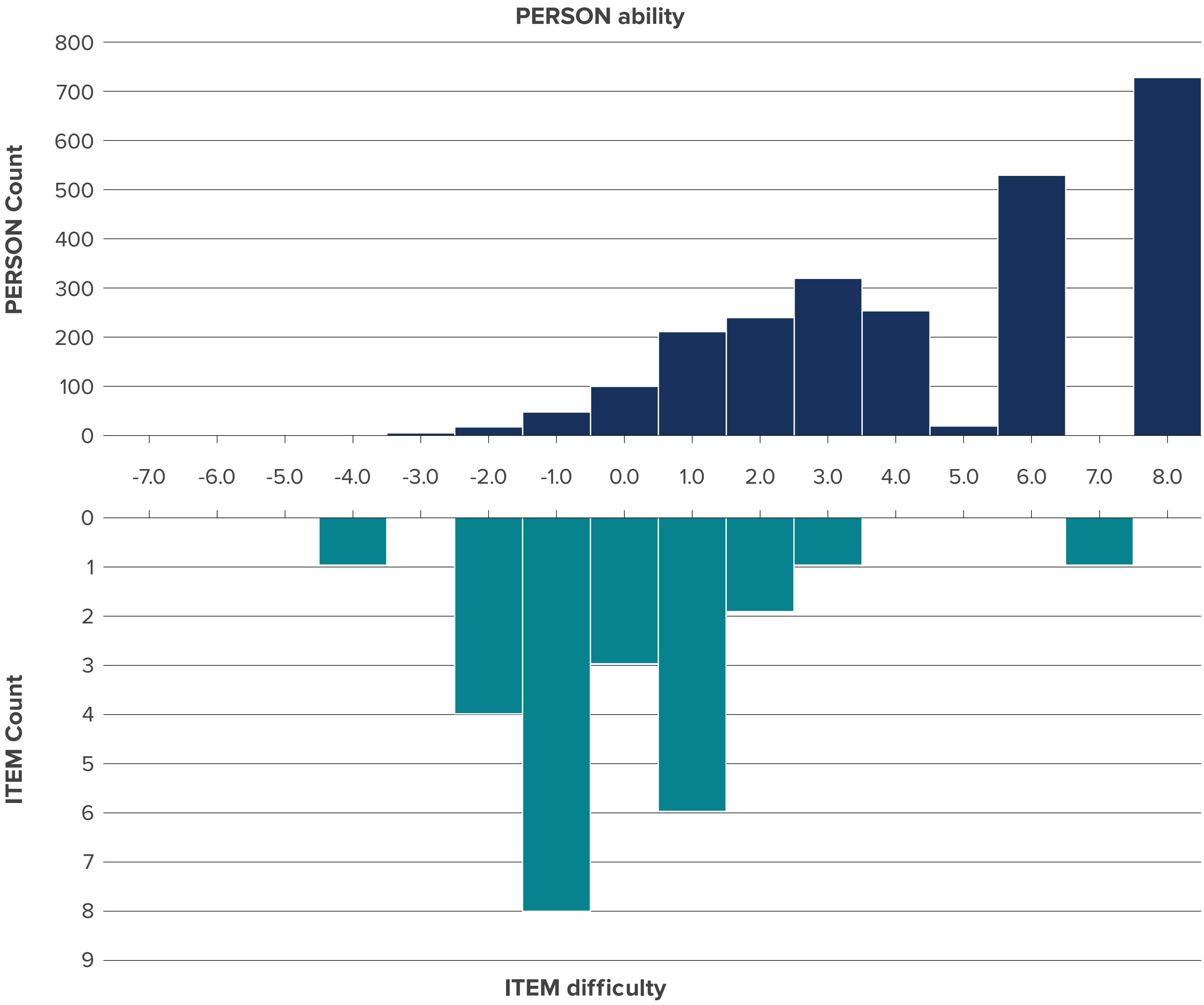
The reliability coefficient and person separation index of each AEDC domain are also reported in Table 1. Three domains (Social Competence, Emotional Maturity, and Communication Skills and General Knowledge) show a reasonable level of person separation, with the reliability coefficient greater than 0.8 and person separation index greater than 2. This shows that for these three domains the AEDC instrument can be expected to produce similar results under consistent conditions. The PSI results were above the minimum preferable value of 2. This meant the AEDC domains were able to distinguish acceptably well those children who could demonstrate the skills tested by the AEDC instrument, and those who could not.

On the other hand, both Language and Cognitive Skills, and Physical Health and Wellbeing domains show a relatively low person reliability coefficient around 0.7 and low person separation index around 1.5. This indicates that the instrument for these two domains was not sensitive enough to distinguish between different levels of abilities. The relatively low levels of separation and reliability for the Language and Cognitive Skills domain are most likely due to the targeting issue described in the next section (Section 4.2.2).

4.2.2. Person and item distributions for the Language and Cognitive Skills domain

Given that results from analyses in Stage 1 provided evidence of the key importance of the AEDC Language and Cognitive Skills domain, this domain was the focus of the remainder of the analyses conducted in Stage 2. A person-item distribution map for the Language and Cognitive Skills AEDC domain is shown in Figure 5. Here it can be seen that the person distribution (navy blue) did not align well with the item distribution (teal) and that there was a pronounced ceiling effect. That is, there were a large number of high performers at the top end of the developmental spectrum not being targeted by any indicators.

Figure 5: Person–item distribution map for the Language and Cognitive Skills domain



4.2.3. Item difficulty in the Language and Cognitive Skills domain

The items belonging to the AEDC Language and Cognitive Skills domain were assessed in terms of their relative difficulties, which provide evidence of the order of skills acquisition within the domain. Table 2 shows this sequence while providing the item difficulty estimates from the unidimensional Rasch model. The sequence of skill acquisition is also plotted in Figure 6, where one dot represents one indicator. The indicators with lower item difficulty, located on the left-hand side of the spectrum in Figure 6, assessed less advanced items pertaining to AEDC Language and Cognitive Skills, while indicators with higher item difficulty were situated on the right-hand side of the spectrum and assessed more advanced Language and Cognitive Skills. Item fit statistics are reported in the last 2 columns of Table 2 and a more detailed explanation of the item fit is provided in Appendix A.

Table 2: Item difficulty estimates from the unidimensional Rasch model for the AEDC Language and Cognitive Skills domain with item fit statistics

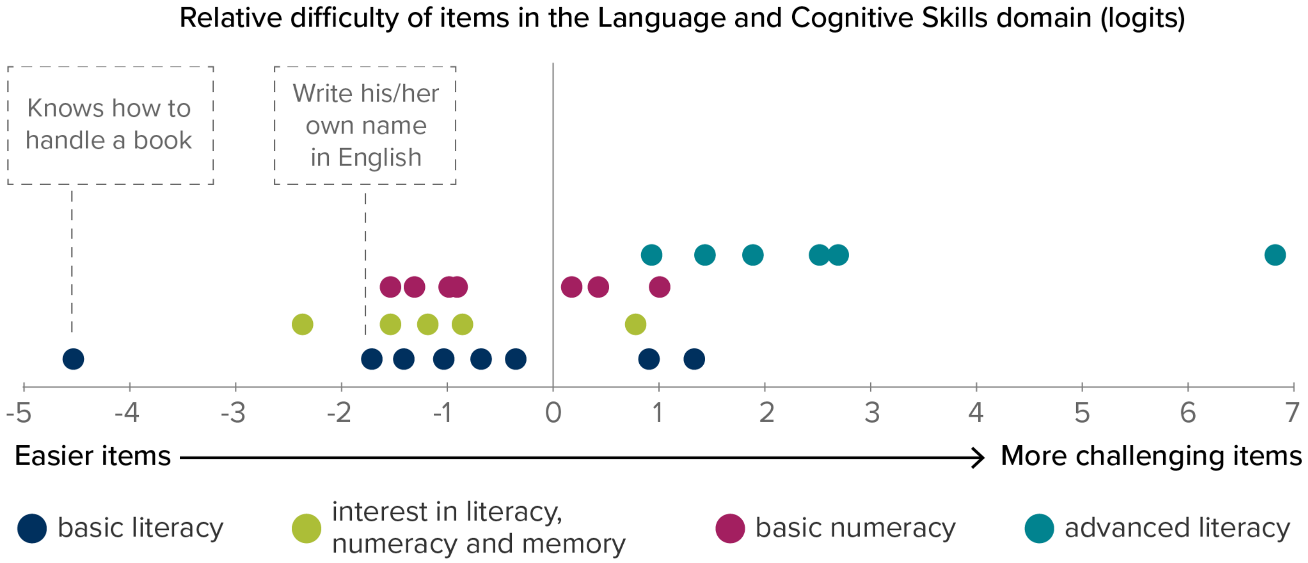
Indicator	Description	Subdomain	Item difficulty	Infit	Outfit
Lan_B8	Knows how to handle a book	Basic literacy	-4.470	0.972	0.140
Lan_B9	Generally interested in books	Interest in literacy, numeracy and memory	-2.290	1.096	0.497
Lan_B21	Write his/her own name in English	Basic literacy	-1.660	0.864	0.287
Lan_B27	Sort and classify objects by a common characteristic	Basic numeracy	-1.520	0.981	0.742
Lan_B26	Interested in games involving numbers	Interest in literacy, numeracy and memory	-1.510	1.039	1.268
Lan_B18	Experiment with writing tools	Basic literacy	-1.410	1.197	1.173
Lan_B32	Recognise geometric shapes	Basic numeracy	-1.330	1.008	1.120
Lan_B25	Interested in mathematics	Interest in literacy, numeracy and memory	-1.200	1.015	1.196
Lan_B14	Participate in group reading activities	Basic literacy	-1.040	1.042	0.743
Lan_B28	Use one-to-one correspondence	Basic numeracy	-1.000	0.902	0.515

Indicator	Description	Subdomain	Item difficulty	Infit	Outfit
Lan_B33	Understand simple time concepts	Basic numeracy	-0.950	1.045	**3.207**
Lan_B10	Interested in reading	Interest in literacy, numeracy and memory	-0.910	0.944	1.315
Lan_B19	Aware of writing directions in English	Basic literacy	-0.750	0.925	0.487
Lan_B11	Identify some letters of the alphabet	Basic literacy	-0.390	0.918	0.683
Lan_B30	Recognise numbers 1 to 10	Basic numeracy	0.130	0.932	0.575
Lan_B31	Say which number is bigger of the two	Basic numeracy	0.360	0.900	0.636
Lan_B24	Remembers things easily	Interest in literacy, numeracy and memory	0.820	1.071	1.336
Lan_B22	Write simple words	Advanced literacy	0.920	0.906	0.511
Lan_B12	Attach sounds to letters	Basic literacy	0.930	0.943	0.852
Lan_B29	Count to 20	Basic numeracy	0.970	0.970	0.981
Lan_B13	Show awareness of rhyming words	Basic literacy	1.250	1.084	1.239
Lan_B15	Read simple words	Advanced literacy	1.350	0.900	0.585
Lan_B20	Interested in writing voluntarily	Advanced literacy	1.860	1.293	**1.738**
Lan_B23	Write simple sentences	Advanced literacy	2.460	1.039	0.984
Lan_B17	Read simple sentences	Advanced literacy	2.560	0.902	0.730
Lan_B16	Read complex words	Advanced literacy	6.840	1.088	**1.779**

Inspection of Figure 6 shows that indicators belonging to the subdomains of Basic Literacy, Basic Numeracy, and Interest in Literacy, Numeracy and Memory subdomains are mostly located around the lower end of the spectrum. This means these indicators assessed more basic literacy and numeracy skills in the very early developmental stage. In contrast, the current analysis placed indicators of the Advanced Literacy subdomain on the right, implying that they assessed more challenging literacy skills.

One important result demonstrated in Figure 6 is that the indicators are not equally spread along the domain continuum. For instance, the least advanced indicator, 'knows how to handle a book,' is situated a significant distance from the second least advanced indicator, 'write his/her own name in English'. This suggests that a big leap in skill is required to reach the second milestone from the first. Similarly, the large gap in the proficiency of the second most advanced indicator, 'read simple sentences,' and the most advanced indicator, 'read complex words,' suggests a significant effort is required to progress between these 2 skills along the modelled scale.

Figure 6: Item difficulty distribution for the Language and Cognitive Skills AEDC domain



4.3. Stage 3: Multilevel modelling predicting Year 3 literacy and numeracy using individual AEDC indicators

The Stage 3 analysis presented here was conducted to understand the impact of individual indicators in the key AEDC domain of Language and Cognitive Skills on Year 3 NAPLAN Numeracy, Reading and Writing results.

Bivariate correlations were generated between each item, or indicator, from the AEDC Language and Cognitive Skills domain and each of the Year 3 NAPLAN scores (Reading, Writing and Numeracy). Table 3, below, presents these results. The strength of the correlations in Table 3 is generally low, ranging from 0.003 to 0.325 (average = 0.200). The 3 indicators with the strongest correlations with NAPLAN results were 'read complex words' (all r > 0.3), 'count to 20' (all r > 0.28) and 'recognise numbers 1 to 10' (all r > 0.28).

Table 3: Bivariate correlations between NAPLAN results (Numeracy, Reading, Writing) and AEDC indicators

AEDC indicator (order questions are provided)	AEDC subdomain	Numeracy	Reading	Writing
Knows how to handle a book	Basic literacy	0.039	0.030	0.040
Generally interested in books	Interest in literacy, numeracy and memory	0.104	0.121	0.089
Interested in reading	Interest in literacy, numeracy and memory	0.164	0.195	0.188
Identify some letters of the alphabet	Basic literacy	0.220	0.236	0.226
Attach sounds to letters	Basic literacy	0.276	0.277	0.253
Show awareness of rhyming words	Basic literacy	0.250	0.253	0.227
Participate in group reading activities	Basic literacy	0.169	0.173	0.165
Read simple words	Advanced literacy	0.264	0.248	0.225
Read complex words	Advanced literacy	0.325	0.306	0.300
Read simple sentences	Advanced literacy	0.266	0.263	0.239
Experiment with writing tools	Basic literacy	0.074	0.097	0.137
Aware of writing directions in English	Basic literacy	0.188	0.193	0.216
Interested in writing voluntarily	Advanced literacy	0.209	0.230	0.279
Write his/her own name in English	Basic literacy	0.166	0.165	0.207

AEDC indicator (order questions are provided)	AEDC subdomain	Numeracy	Reading	Writing
Write simple words	Advanced literacy	0.246	0.231	0.227
Write simple sentences	Advanced literacy	0.263	0.264	0.268
Remember things easily	Interest in literacy, numeracy and memory	0.244	0.240	0.246
Interested in mathematics	Interest in literacy, numeracy and memory	0.160	0.136	0.178
Interested in games involving numbers	Interest in literacy, numeracy and memory	0.126	0.115	0.134
Sort and classify objects by a common characteristic	Basic numeracy	0.108	0.126	0.135
Use one-to-one correspondence	Basic numeracy	0.183	0.163	0.213
Count to 20	Basic numeracy	0.309	0.280	0.289
Recognise numbers 1 to 10	Basic numeracy	0.301	0.285	0.288
Say which number is bigger of the two	Basic numeracy	0.267	0.246	0.268
Recognise geometric shapes	Basic numeracy	0.131	0.121	0.129
Understand simple time concepts	Basic numeracy	0.135	0.150	0.121

Multilevel analyses were performed next. This involved using defined steps for reducing the models (see Section 3.4.4) to identify the most relevant indicators that impacted on Year 3 NAPLAN performance while controlling for other child-level and school-level demographics. Likelihood ratio tests confirmed that the final reduced multilevel models provided a better fit to the data compared to the full model. Very similar findings were obtained for the NAPLAN Numeracy and Reading domains.

As shown in Table 4, the item 'recognise numbers 1 to 10' was the strongest predictor for Year 3 NAPLAN Numeracy and the second strongest for Reading and Writing results. The presence of this skill in early childhood contributed to a significant increase in Numeracy and Reading scores by more than 30 scale score units and had effect sizes of around 0.4, which show noteworthy effects.

The item 'read complex words' was the strongest predictor for NAPLAN Reading and Writing results. This was also the most advanced indicator on the AEDC Language and Cognitive Skills proficiency scale (reported in Section 4.2). Interestingly, being 'able to remember things easily' was among the top 5 predictors for NAPLAN Numeracy and Reading outcomes and ranked 6th for the NAPLAN Writing outcome. It was also the most advanced indicator in the Interest in Literacy, Numeracy and Memory subdomain. These findings provided some evidence that developed Language and Cognitive Skills in the first year of full-time school predict better learning outcomes in later years.

Table 4: Increase in NAPLAN scale score for the top 5 strongest and significant predictors of Year 3 NAPLAN performance

Indicator	Scale score	Effect size
Numeracy		
Recognise numbers 1 to 10	+31.37	0.43
Read complex words	+30.15	0.42
Count to 20	+20.77	0.29
Remember things easily	+19.40	0.27
Attach sounds to letters	+12.07	0.17
Reading		
Read complex words	+34.32	0.39
Recognise numbers 1 to 10	+34.03	0.39
Remember things easily	+20.97	0.24
Attach sounds to letters	+19.48	0.22
Count to 20	+17.49	0.20
Writing		
Read complex words	+22.58	0.34
Recognise numbers 1 to 10	+22.34	0.33
Use one-to-one correspondence	+20.53	0.31
Write his/her own name in English	+17.51	0.26
Aware of writing directions in English	+12.43	0.19

4.4. Stage 4: Aligning the AEDC item progression to the Australian Curriculum and the National Literacy and Numeracy Learning Progressions

The final stage of the analysis was conducted to anchor the earlier findings from Stages 1, 2 and 3 to the Australian context, and to use the evidence from this study to enhance understanding of the current functioning of curricula and learning progressions in Australia.

The Rasch fit analysis (see Section 4.2) confirmed that the 4 subdomains – Basic Literacy, Advanced Literacy, Basic Numeracy and Interest in Literacy, Numeracy and Memory – supported the measurement of the Language and Cognitive Skills AEDC domain. The item progression from Section 4.2 was reproduced in Table 5, together with each item's correspondence with the Australian Curriculum: English (AC:E), Australian Curriculum: Mathematics (AC:M) and the NLNLP. Examination of Table 5 shows:

- All indicators in Basic Literacy, Basic Numeracy and Advanced Literacy are aligned to the AC:E and AC:M and the NLNLP.
- Nearly all Basic Literacy, Advanced Literacy and Basic Numeracy indicators aligned to Australian Curriculum standards across Foundation and Year One. The exception to this was 'is the child able to read complex words?'. This indicator aligns to AC:E phonic and word knowledge content descriptors AC9E2LY10 and AC9E2LY11, which are both Year 2 level content descriptors.
- Nearly all Basic Literacy, Advanced Literacy and Basic Numeracy indicators aligned to the skills outlined by the NLNLP that are grouped as observable in Foundation and Year One (ACARA, 2022). The exceptions were 'is the child able to write simple sentences?' which aligned to Literacy sub-element CrT4/5, and 'is the child able to read complex words?' which aligned to Literacy sub-element PKW6, and is aligned to the Year 2 AC:E.

Together, the above points indicate that the estimated difficulty of the AEDC items is, for the most part, consistent with their perceived difficulty within the NLNLP.

There are some exceptions, however, which can be identified in Table 5. The item 'interested in reading', which is aligned to a level UnT1/2 on the Reading and Viewing sub-element 'Understanding Text', showed higher difficulty than 'can participate in group reading activities' (UnT3). This finding provides an interesting insight into the order of skills that children find challenging, particularly as it relates to engagement with reading. An unexpected order was also demonstrated for the item 'show awareness of rhyming words' (PhA2). This item was identified in the Rasch analysis to be a difficult item, yet the NLNLP indicates that it is a foundational skill within 'Phonological Awareness.' This skill is also identified within the AEDC as a Basic Literacy skill yet is closer in the Stage 2 ordered skill progression to the Advanced Literacy items. This finding may suggest that developing awareness of rhyming words is more difficult than other foundational literacy and numeracy skills.

Table 5: Progression of item difficulty

Item progression: Easier → More difficult

AEDC item	AEDC category	National Literacy and Numeracy Learning Progressions	Australian Curriculum: Mathematics	Australian Curriculum: English
Knows how to handle a book	**Basic Literacy skill**	UnT2	–	AC9EFLA07 AC9EFLA03 AC9EFLA04
Generally interested in books	**Interest in Literacy, Numeracy and Memory**	UnT1	–	AC9EFLA04
Writes his/her own name in English	**Basic Literacy skill**	CrT3	–	AC9EFLY14
Sort and classify objects by a common characteristic	**Basic Numeracy skill**	UGP2	AC9M1SP01	–
Interested in games involving numbers	**Interest in Literacy, Numeracy and Memory**	N/A	–	–
Experiment with writing tools	**Basic Literacy skill**	HwK1	–	AC9EFLY06 AC9EFLY08
Recognise Geometric Shapes	**Basic Numeracy skill**	UGP2	AC9MFSP01	–
Interested in Mathematics	**Interest in Literacy, Numeracy and Memory**	N/A	AC9EFLA02	–
Participate in group reading activities	**Basic Literacy skill**	UnT3	–	AC9EFLE02
Use one-to-one correspondence	**Basic Numeracy skill**	CPr3	AC9MFN03	–
Understands simple time concepts	**Basic Numeracy skill**	MeT1	AC9MFM02	–
Interested in reading	**Interest in Literacy, Numeracy and Memory**	UnT1/2	–	–
Aware of writing direction in English	**Basic Literacy skill**	CrT3	–	AC9EFLA04
Identify some letters of the alphabet	**Basic Literacy skill**	PKW1/2	–	AC9EFLY11
Recognise numbers 1 to 10	**Basic Numeracy skill**	NPV2	AC9MFN01	–

(Table 5 continues on the following page)

(Table 5 continued)

Item progression: Easier → More difficult

AEDC item	AEDC category	National Literacy and Numeracy Learning Progressions	Australian Curriculum: Mathematics	Australian Curriculum: English
Say which number is bigger of the two	**Basic Numeracy skill**	NPV3	AC9MFN01 AC9MFN03	–
Remembers things easily	**Interest in Literacy, Numeracy and Memory**	N/A	–	–
Write simple words	**Advanced Literacy skill**	CrT4	–	AC9EFLY12 AC9EFLY14
Attach sounds to letters	**Basic Literacy skill**	PKW3	–	AC9EFLY10 AC9EFLY11
Count to 20	**Basic Numeracy skill**	CPr5	AC9MFN03	–
Show awareness of rhyming words	**Basic Literacy skill**	PhA2	–	AC9EFLE04 AC9EFLY09
Read simple words	**Advanced Literacy skill**	PKW3/4	–	AC9EFLY14
Interested in writing voluntarily	**Advanced Literacy skill**	CrT1-CrT3	–	AC9EFLE02 AC9EFLY06
Write simple sentences	**Advanced Literacy skill**	CrT4/5	–	AC9E1LY06
Read simple sentences	**Advanced Literacy skill**	UnT4	–	AC9EFLY04 AC9EFLY06
Read complex words	**Advanced Literacy skill**	PKW6	–	AC9E2LY10 AC9E2LY11

The AEDC indicators found in Stage 3 analyses to predict the greatest shifts in NAPLAN scores are summarised in Table 4. Their alignment to the NLNLP is shown in Table 6. Of all the AEDC items, 8 predict the largest shifts across one, 2 or all 3 NAPLAN domains. These are Recognise numbers 1 to 10 (NPV2); Read complex words (PKW6); Count to 20 (CPr5); Remember things easily; Attach sounds to letters (PKW3); Use one-to-one correspondence (CPr3); Write his/her own name in English (CrT3); Aware of writing directions in English (CrT3).

Table 6: Linking the NLNLP with the top 5 strongest and significant AEDC predictors of Year 3 NAPLAN performance, ordered from greatest to least impact on the performance

Five strongest indicators of future performance	NLNLP
Numeracy	
Recognise numbers 1 to 10	NPV2 (Foundation)
Read complex words	PKW6 (Year 2)
Count to 20	CPr5 (Year 1)
Remember things easily	N/A
Attach sounds to letters	PKW3 (Foundation)
Reading	
Read complex words	PKW6 (Year 2)
Recognise numbers 1 to 10	NPV2 (Foundation)
Remember things easily	N/A
Attach sounds to letters	PKW3 (Foundation)
Count to 20	CPr5 (Year 1)
Writing	
Read complex words	PKW6 (Year 2)
Recognise numbers 1 to 10	NPV2 (Foundation)
Use one-to-one correspondence	CPr3
Write his/her own name in English	CrT3 (Foundation)
Aware of writing directions in English	CrT3 (Foundation)

The ability to recognise numbers 1 to 10 was the strongest predictor for Year 3 NAPLAN Numeracy results and was the second strongest predictor for Reading and Writing. The presence of this skill in early childhood contributed to a significant increase in Numeracy and Reading scores by more than 30 scale score points. This skill maps to a foundation-level skill on the NLNLP (NPV2). However, the ability to read complex words, captured by the AEDC, was the strongest predictor for NAPLAN Reading and Writing results. This was also the most advanced indicator on the Language and Cognitive Skills AEDC scale and mapped to Year 2 level skills on the NLNLP (PKW6). Therefore, a mix of both foundational and more complex skills within the AEDC are among the strongest predictors of Year 3 NAPLAN results.

4.5. Limitations of the study

There are several limitations to note regarding this study. First, the dataset did not include a complete cohort of Australian children, but a sample of approximately 2,000 children who were assessed in the 2009 AEDC collection, with matched NAPLAN and AEDC data. Despite the fact that this sample was drawn from a nationally representative sample of LSAC data, results appear to be not completely representative of the population starting school in 2009. Specifically, caution should be used in interpreting the findings regarding the less represented child groups in the sample. These are those with lower achievement in NAPLAN, Aboriginal and Torres Strait Islander children, children speaking English as a second language, and those from the government school sector.

A second point to note is that since the collection of the AEDC data began in 2009, some changes may have been made to the AEDC instrument. This means the implications of the current findings may require some adjustment to apply to the populations of children currently at school.

A third limitation is that some data appeared to be missing not at random from the LSAC results (Daraganova & Sipthorp, 2011). Missing not at random means some children with certain characteristics that are not available in the data are more/less likely to have missing data. Our effort to address this using the MICE data imputation method reduced bias in model estimates to some extent but may not have fully removed the effects of this limitation.

Fourth, the number of children per school was small, while the interdependence of the data within schools was still substantial. As a result, both multilevel modelling and single-level modelling impose statistical instability. We employed both modelling strategies for the SEM mediation analyses. The comparisons of the two sets of results did not present serious differences, supporting the validity of the findings. Future research may further explore full multilevel analysis with larger numbers of children per school.

Finally, it should be noted that the AEDC data is subject to some uncertainty from teachers' judgements as the teachers observe the children's skills and respond to the questions. Moreover, the Early Development Index was designed primarily to differentiate whether children were meeting certain thresholds of development, rather than to measure the full spectrum of early development. As such, it may not be the most ideal instrument for the purpose of understanding the full range of skills and knowledge children possess as they start school. However, the use of the AEDC data in this analysis has provided some meaningful and applicable results, and this potentially points to the usefulness of replicating the current approach with a more targeted measure of early childhood development at the start of primary school. Replication of our approach with different cohorts could also serve to address some of the limitations noted here.

5. Discussion

5.1. Summary of results

This study focussed on 3 key research questions (see Section 1.3). The first question asked whether a child's ability to successfully complete Basic and Advanced Literacy and Basic Numeracy tasks provided an indication of future achievement in literacy and numeracy assessments. Stage 1 analyses used Structural Equation Modelling to examine AEDC domains in relation to NAPLAN results and found that the Language and Cognitive Skills domain was the strongest predictor of later NAPLAN results. Stage 3 analyses showed the additional finding of the relevance of individual items from this AEDC domain, corresponding to specific skills, that were strongly predictive of Year 3 NAPLAN results.

The second research question asked how teachers' responses to questions from the Language and Cognitive Skills AEDC domain could demonstrate a sequenced progression of early literacy and numeracy skills on a progressive scale. This question was answered by Stage 2 analyses, in which Rasch modelling showed the sequence of skill acquisition captured by the AEDC Language and Cognitive Skills domain.

The third question was does the progression of Language and Cognitive Skills items align with the Australian Curriculum (for English and Mathematics) and National Literacy and Numeracy Learning Progressions? This was answered in Stage 4 analyses, where the direct mapping of the AEDC items to the Australian Curriculum and the NLNLP was conducted.

Seen together, the results of this study showed that while all 5 AEDC domains were important, the Language and Cognitive Skills domain had the strongest positive association with later Year 3 and Year 5 NAPLAN performance. The development of early literacy skills was shown to be indicative of future success in both literacy and numeracy. Further, we were able to demonstrate the value of mastering specific basic literacy and numeracy skills. For example, the skill of 'attaching sounds to letters' is a basic literacy skill that was shown here to be an indicator of future success in both Reading (with an increase in NAPLAN scale score of +19.48) and Numeracy (with an increase in NAPLAN scale score of +12.07; see Table 4). This suggests that understanding letters and sounds is not only essential for the development of reading, but also important for the development of numeracy.

Linking these findings to skill progressions added another element to the current investigation. As discussed, the further along the difficulty scale a skill appears, the more advanced the skill. To demonstrate competence in the most advanced item identified here, 'read complex words', children may need to learn the prior skills along the progression to build the foundation for reading complex words. Of course, young children's learning is not a uniform progression; some children may leap ahead, skipping skills, plateau, regress, and speed up again in their learning. Nonetheless, a child's ability to 'read complex words', developed through their ability to understand the phonetic properties of words, influenced all 3 areas of Year 3 NAPLAN. There were increases in NAPLAN scale scores of +30.15 on Numeracy scores, +34.32 on Reading scores and +22.58 on Writing scores. The alignment of this item to the NLNLP showed that the acquisition of each NLNLP skill by a child, as they work towards being able to read complex words, may influence not only the child's reading potential, but writing and mathematical capability too.

The Rasch analysis investigated 5 skills in the Interest in Literacy, Numeracy and Memory subdomain. These items, in ascending order of difficulty, are:

1. Generally interested in books
2. Interested in games involving numbers
3. Interested in mathematics
4. Interested in reading
5. Can remember things easily.

Although the Interest in Literacy, Numeracy and Memory subdomain could not easily be aligned to the AC:E, AC:M nor the NLNLP, the results from the multidimensional item response model suggested that overall, the children showing interest in literacy and numeracy, and who can also remember things easily, demonstrated stronger early literacy and numeracy skills. This links to insights from cognitive science, which suggest that interest in learning any content is facilitated by exposure to the learning domain (Sweller et al., 1998, 2019). Consequently, childhood interest in reading is likely to increase as the child is exposed to reading. This study highlights the importance of facilitating young children's interest in text in addition to the cognitive skills that are described in the AC:E, AC:M and NLNLP.

Unsurprisingly, 'remember things easily' was shown to be a strong indicator of future NAPLAN success across Numeracy (+19.40) and Reading (+20.97). Literacy and mathematics experts point out that working memory plays a critical role in learning mathematics and in learning to read. Cognitive science explains how learners process new information in working memory and then transfer the knowledge gained to long-term memory, where it can be stored and used. This process highlights the importance of consolidating learning by building networks of information in long-term memory that can be accessed easily and automatically. Building long-term memory requires the working memory to initially process new information effectively without being overloaded (See CESE 2017 for a literature review of the 'Cognitive Load Theory'). Therefore, 'remembers things easily' is related to learning things easily.

It is also important to view 'remembers things easily' in the context of environmental factors that influence memory function. For example, learning environments that manage child stress, support emotional regulation and help children to develop effective learning habits and behaviours can facilitate memory being used to effectively consolidate learning (Montoya et al., 2019). This highlights the importance of supportive learning environments for closing gaps in early learning and fostering later learning for all children, including children with specific difficulties processing new information in short-term memory or recalling information from long-term memory.

The results found here suggest several key takeaway messages. These are discussed in more detail in the following section.

5.2. Early mastery of Language and Cognitive Skills sets children up for later academic success

A clear outcome of the analysis was demonstrating that the skills that children develop in their first year of schooling are associated with NAPLAN Reading and Writing scores in Year 3, with sustained and additional impact into Year 5. In particular, early mastery of language and cognitive skills has a lasting effect and sets children up for later academic success. Conversely, children who enter full-time school without these skills are more likely to experience prolonged challenges as they progress through their schooling journey. Some of the key findings from the data analysis were that:

- Language and Cognitive Skills development levels had a strong and direct impact on children's academic performance in Year 3. This was followed by Communication Skills and General Knowledge, which also had a direct but smaller effect on Year 3 results.
- The direct or indirect effect of the Language and Cognitive Skills domain from the AEDC on Reading and Numeracy scores until Year 5 was sustained, with aspects of this influence becoming evident after Year 3. In this way, school ICSEA showed similarly sustained impact.
- For Numeracy outcomes, a direct effect of gender on both Year 3 and Year 5 Numeracy (favouring boys) was also significant.
- Of the 4 subdomains, Basic Literacy and Basic Numeracy success significantly explained Year 3 Numeracy outcomes.

Other AEDC domains of Physical Health and Wellbeing, Emotional Maturity, and Social Competence also indirectly impacted later learning outcomes. Hence, more tailored and targeted support in both ECEC and school contexts is called for, particularly for children with lower development in these domains.

Children with more disadvantaged backgrounds were less likely to participate in NAPLAN tests or to be tracked through the LSAC period. This means systems, researchers and schools have less data to check the progress of these children who were more likely to need extra support.

Our analysis also shed further light on the impact of demographic factors on children's achievement. Children and school demographic variables were consistently significant in explaining academic achievement in the SEM models with AEDC predictors. These results mean that children with disadvantaged backgrounds and conditions were more likely to have lower levels of early development skills on entry to school. They also mean that these initial equity gaps persisted as children progressed to Year 3, in some cases exacerbating in Year 5. The widening of the performance gap as children progress through school is worrying as it indicates that it is increasingly difficult for children to catch up to their peers (Gonski, 2018).

5.3. Supporting children to progressively develop their early literacy and numeracy skills is a key to the success of later academic achievement

The SEM analysis provided evidence that the acquisition of AEDC Language and Cognitive Skills in early years of school is associated with better learning outcomes in later years, as indicated through large-scale assessments such as NAPLAN. Of the 26 skills tested within the AEDC Language and Cognitive Skills domain, 8 were shown to have a strong correlation with successful NAPLAN performance in Year 3 across Numeracy, Reading and Writing. These 8 items are: Recognise numbers 1 to 10; Read complex words; Count to 20; Remember things easily; Attach sounds to letters; Use one-to-one correspondence; Write his/her own name in English; Aware of writing directions in English.

Our analysis showed that children in their foundation year who demonstrated these skills were more likely to experience future success throughout the primary years. Importantly, this does not demonstrate a causal relationship.

These skills may be indicators of a range of experiences. For example, at home, foundational literacy and numeracy skills are supported through responsive adult-child interactions and shared activities such as reading together, which simultaneously supports children's sense of wellbeing and identity as a capable learner. In ECEC, early childhood educators and teachers support children's early literacy and numeracy skills through play-based learning and intentionality that builds on each child's strengths, curiosities and capabilities. Further analysis is warranted to understand how children's ECEC experiences contribute to their learning and development at school entry.

In the foundation year, teachers can support children to learn these skills by providing targeted teaching as part of a broader approach to teaching literacy and numeracy, to ensure fundamental skills are embedded.

The ability to recognise numbers 1 to 10 was the strongest predictor for Year 3 NAPLAN Numeracy and second strongest predictor for Reading and Writing. The presence of this skill by the start of primary school contributed to a significant increase in Numeracy and Reading scores by more than 30 scale score points (effect size 0.4). This finding aligns with Geary's observations that the 'risk of long-term difficulties with mathematics or at least starting school significantly behind one's peers in fundamental numerical knowledge – can be determined by 3 and a half to 4 years of age by the length of children's count list (how far they can count without error)' (Geary, 2022, p. 7).

The ability to read complex words was the strongest predictor for NAPLAN Reading and Writing results. This was also the most advanced indicator on the Language and Cognitive Skills AEDC domain. Interestingly, being able to remember things easily was among the top 5 predictors for Numeracy and Reading domains. This was also the most advanced indicator in the Interest in Literacy, Numeracy and Memory subdomain. These findings provide some evidence that the acquisition of skills towards more advanced Language and Cognitive Skills would be one indicator of better learning outcomes in later years.

While some children will enter primary school having acquired foundational skills in the home and ECEC settings, others will not have had this opportunity. Teachers of the early years of school have an opportunity to identify children in their classrooms who have not yet acquired these skills, and provide the support required. The NLNLP indicates that children develop knowledge and skills progressively, building on foundations to develop higher-order skills over time.

Where children are unable to meet AEDC indicators, teachers can use the NLNLP to identify and actively teach the observable and measurable literacy and numeracy skills to each of these indicators. For example, 'able to recognise numbers 1 to 10' was found to align with NPV2 on the numeracy progressions. To achieve this skill, the numeracy progressions outline that the child must first identify and produce familiar number names and numerals that relate to their lives without determining quantity or ordinal position (NPV1). Another example, from AEDC basic literacy skills, is 'attach sounds to letters,' which aligns to PKW3 on the literacy progressions.

Teachers can focus on teaching the precursor skills, such as indicating letters and words in a variety of situations in the environment (PKW1) and identifying familiar letters and words in environmental print (PKW2), to support the child to recognise and actively attach sounds to individual letters. Interestingly, 'attach sounds to letters' is not a difficult skill within the AEDC, yet the ability of a child to demonstrate this skill predicts a shift in NAPLAN Reading by 19 scale scores (effect size 0.22) and Numeracy by 12 scale scores (effect size 0.17). This suggests that supporting children, through the teaching of precursor skills, to achieve a seemingly easy skill within the AEDC can have a meaningful effect on future academic success.

The above examples show how AEDC data can feedforward by highlighting the literacy and numeracy skills that indicate a significant impact on children's future literacy and numeracy success. By coupling AEDC data with alignment to key documents such as the AC:E, AC:M as a starting point and then engaging with the NLNLP for more nuanced and identified steps, teachers can better understand where children are in their learning. This study also demonstrates that while some items predictive of later success were identified as advanced on the ordered skill progression ('count to 20' and 'read complex words'), the other items are more foundational skills and highlight the importance of ensuring these skills are mastered to provide a pathway to achieving more advanced literacy and numeracy skills. In other words, being able to demonstrate difficult skills early on is not always associated with future literacy and numeracy success. This analysis highlights the need for achieving basic literacy and numeracy skills such as 'recognising numbers 1 to 10' and 'writing his/her name in English' in order to progress to advanced literacy and numeracy skills.

5.4. Understanding children's strengths and vulnerabilities facilitated through the alignment between AEDC, the Australian Curriculum and NLNLP provides feedforward opportunities for targeted teaching and support

Our analysis showed that the progression of early literacy, numeracy and cognitive skills as derived from the analysis of AEDC Language and Cognitive Skills domain items align with the Australian Curriculum and the NLNLP. This highlights that the acquisition of the foundational skills identified by the AEDC follows the same pattern identified by the Australian Curriculum and the NLNLP. There were some exceptions: 'interested in reading' and 'show awareness of rhyming words' were found by our analysis to be more difficult than other items placed as more basic or foundational skills in the NLNLP.

Despite these exceptions, the general sequencing of the AEDC indicators to follow the progression of observable skill development outlined in the NLNLP indicates the alignment of the AEDC census items and curriculum documents (NLNLP) for the early years of school. It is also reassuring and supportive of the work of ECEC and school teachers. These tools, which are aligned, may be used by teachers in the following ways:

- Where a child is not yet achieving a skill identified in the AEDC, the teacher may locate the skill within the NLNLP. There is an opportunity to look to earlier skills in the NLNLP progressions, where the teacher can focus on teaching earlier precursor knowledge and skills until the child can demonstrate the AEDC skill. For example, a teacher might want to support a child to identify and produce familiar number names and numerals that relate to their lives (NPV1). Other related skills include matching quantity of items in a collection to the correct number name or numeral from 1 to 10, or identifying standard number configurations such as those on dominos (NPV2).
- When a child is consistently demonstrating the skills described in the AEDC, the teachers may use the NLNLP to identify the next step in the child's learning.

In this way, the AEDC – whose primary purpose is to provide a population-level census of how Australia's children are faring – can also provide informal information for foundation year teachers completing the census, supporting ongoing planning for each child's learning. The creation of a language and cognition skill progression scale and the subsequent alignment to the AC:E, AC:M and NLNLP is an important acknowledgement of the potential of this data to *feedforward* to inform teaching and learning.

There is an opportunity in the early years in school to target the achievement of the 8 essential skills assessed through AEDC that predict the highest future success and use the NLNLP and Australian Curriculum to backwards-map a child's learning where the child is not meeting these essential early learning skills. School teachers may also choose to draw on the AERO Early Childhood Learning Trajectories, which link to the Early Years Learning Framework and describe how children learn and develop in 5 key domains, including mathematical thinking, language and communication. Using these resources can also support continuity for children transitioning from ECEC into school.

The alignment of the AEDC Language and Cognitive Skills domain items to the AC:E and the AC:M provides opportunities for schools to target and ensure children are able to achieve these core skills. Teachers may also require support at the school and system levels to use data to ensure each child is able to progress to achieving improvement in literacy and numeracy skills. Further research may be required to assess the school and system-level supports that would best enable teachers to cater for early childhood learning diversity and the potential to use the Australian Curriculum and the NLNLP as aligned to AEDC items.

This research has provided a first step in identifying the foundational skills that are indicators of the skills children need to support future literacy and numeracy success. The development of teaching resources and learning programs focused on these early developmental skills may be helpful for targeting these key areas to set children up for future literacy and numeracy success. Coupling high-quality Tier 1 instruction with intervention, where required, will ensure teachers provide targeted support for all children (Berkeley et al., 2009).

5.5. Opportunity for more granular understanding of advanced literacy and numeracy skills on entry to primary school

The analysis from Stage 2 included an examination of whether the AEDC instrument can distinguish between children of high and low developmental levels. It was found that while the other domains had greater reliability, the Language and Cognitive Skills domain had the lowest person reliability. This means the skill indicators in this domain do not distinguish high and low developmental levels as well as those in the other domains.

Further analysis from this research highlights that, for children at the high end of the developmental scale, the AEDC is a 'blunt instrument' as it does not provide a granular indicator of more advanced literacy and numeracy skills. This is an important finding and relates to a recently funded study – Measuring Children's Developmental Strengths Pilot – the focus of which was to address the 'ceiling effect in the AvEDI (AEDC)' (Social Research Centre, 2019, p. 20). This study also confirmed that there needs to be an introduction of additional items for children 'who receive the highest possible score of 10 for a specific domain' (Social Research Centre, 2019, p. 20).

The research presented here acknowledges that the AEDC is not intended as an identifier of children of high developmental levels. However, there is opportunity for greater understanding of how to enable future literacy and numeracy achievement for children who start school with a comparatively higher level of literacy and numeracy development. One avenue for addressing this may be through the expansion of focus in assessment of skills at the start of primary school. The consideration of new, more difficult items, anchored to the AC:E, AC:M and NLNLP, may address the gap found in this study between 'read a simple sentence' and 'read complex words', which demonstrates a significant leap in progression. Another avenue for consideration by future researchers is whether the large proportion of high performers suggests that the AEDC should be run in early learning centres rather than in the first year that children start school.

It is also important to acknowledge that while there is potential for a nationally consistent assessment of children when they start school, many jurisdictions currently implement comprehensive entry to school assessment.

6. Conclusion

The aim of this study was to support continuity of learning and development by seeking to understand how children's learning and development progresses in the early years of school. The study has achieved this through identifying the early skills that predict later educational achievement. We have also investigated these skills at a granular level to identify an ordered skill progression that ranges from the skills children find most easy to those they find the most difficult. This ordered skill progression subsequently allowed the alignment of key curriculum documents (AC:E, AC:M and NLNLP) to these items and examination of the connection between item difficulty and curriculum difficulty. The analysis found that there was definite alignment; however, some items that children found easier were considered more challenging in the curriculum documents. There were also some items tagged as basic literacy items that were situated next to those intended to be more advanced literacy items in the ordered skill progression.

One particularly significant finding related to the top 5 skills that predicted future achievement in each NAPLAN domain of Numeracy, Reading and Writing. Demonstration of these 5 key skills was associated with greater likelihood of future success in literacy and numeracy after their foundation year at school. The items that were identified as top 5 skills were not necessarily the most difficult items. This highlights that the mastery of basic skills creates important incremental steps towards more advanced skills.

The study's findings regarding the impact of demographic factors as well as the AEDC domains on future Year 3 and Year 5 literacy and numeracy scores led to another important conclusion. This was that equity gaps not only started from the early years of development, but also remained persistent and showed new avenues of impact in later years. Young children with less development in various subdomains will benefit from more tailored and targeted support to achieve their potential in the longer term.

There is an opportunity for continued collaboration with agencies and groups around Australia who are driving exploration and development of resources supporting literacy and numeracy achievement. Finally, further research into the other AEDC domains and their association with later academic achievement may be beneficial, with a particular focus on children who face greater disadvantage.

References

Australian Bureau of Statistics. (2010). *Australian statistical geography standard (ASGS): Volume 1 – Main structure and greater capital city statistical areas, July 2011*. Retrieved May 15, 2023, from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202011?OpenDocument

Australian Children's Education & Care Quality Authority [ACECQA], 2020.

Australian Curriculum, Assessment and Reporting Authority. (2014). *What does the ICSEA value mean?* https://docs.acara.edu.au/resources/About_icsea_2014.pdf

Australian Curriculum, Assessment and Reporting Authority. (2015). *Myschool fact sheet: Understanding the 'school profile' page*. https://docs.acara.edu.au/resources/Understanding_school_profile_page.pdf

Australian Curriculum, Assessment and Reporting Authority. (2016). *NAPLAN – general*. Retrieved February 1, 2022, from https://www.nap.edu.au/naplan/faqs/na—an--general

Australian Curriculum, Assessment and Reporting Authority. (2017). *The Australian National Assessment Program Literacy and Numeracy (NAPLAN) Assessment Framework: NAPLAN Online 2017-2018*. Retrieved May 29, 2023, from https://nap.edu.au/docs/default-source/default-document-library/naplan-assessment-framework.pdf

Australian Curriculum, Assessment and Reporting Authority. (2020). *National literacy and numeracy learning progression (Version 3.0)*. Retrieved March 31, 2022, from https://www.australiancurriculum.edu.au/resources/national-literacy-and-numeracy-learning-progressions/version-3-of-national-literacy-and-numeracy-learning-progressions/

Australian Curriculum, Assessment and Reporting Authority. (2022). *General capabilities in the Australian curriculum (Version 9)*. Retrieved March 31, 2022, from https://v9.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/literacy?element=0&sub-element=0

Australian Curriculum, Assessment and Reporting Authority. (n.d). General Capabilities Literacy (Version 8.4). Retrieved May 25, 2022 from https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/literacy/

Australian Education Research Organisation. (2022). *Early childhood data in Australia scoping report: Full publication*. https://www.edresearch.edu.au/resources/early-childhood-data-australia-scoping-report/early-childhood-data-australia-scoping-report-full-publication

Australian Government Department of Education. (2022). *Belonging, being and becoming: The early years learning framework for Australia*. (V2.0, 2022). https://www.acecqa.gov.au/sites/default/files/2023-01/EYLF-2022-V2.0.pdf

Australian Government Department of Education, Skills and Employment. (2021). *Australian early development census national report 2021*. https://www.aedc.gov.au/resources/detail/2021-aedc-national-report

Australian Government Department of Education, Skills and Employment. (2022). *Australian early development census (AEDC) data guidelines*. https://www.aedc.gov.au/resources/detail/aedc-data-guidelines

Australian National Audit Office. (2012). *Performance audit: National partnership agreement on literacy and numeracy*. https://www.anao.gov.au/work/performance-audit/national-partnership-agreement-literacy-and-numeracy

Berkeley S, Bender WN, Gregg Peaster L and Saunders L (2009) 'Implementation of Response to Intervention: A Snapshot of Progress', Journal of Learning Disabilities, 42(1):85–95, doi:10.1177/0022219408326214.

Björklund, C. (2012). What counts when working with mathematics in a toddler-group. Early Years, 32(2), 215–228. https://doi.org/10.1080/09575146.2011.652940

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NJ: Routledge.

Brinkman, S., Gregory, T., & Braunack-Mayer, L. (2015). *The AEDC predicts NAPLAN results nine years later*. AEDC National Conference. https://www.aedc.gov.au/resources/detail/the-aedc-predicts-naplan-results-nine-years-later

Brinkman, S., Gregory, T., Harris, J., Hart, B., Blackmore, S., & Janus, M. (2013). Associations between the early development instrument at age 5, and reading and numeracy skills at Ages 8, 10 and 12: A prospective linked data study. *Child Indicators Research, 6*(4), 695–708. https://doi.org/10.1007/s12187-013-9189-3

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Campbell, S. (2020). Teaching phonics without teaching phonics: Early childhood teachers' reported beliefs and practices. *Journal of Early Childhood Literacy, 20*(4), 783–814. https://doi.org/10.1177/1468798418791001

Chaitow, L., McCabe, P., Munro, N., and & Purcell, A. (2022). Language and early literacy professional development: A complex intervention for early childhood educators and speech-language pathologists. *International Journal of Speech-Language Pathology*. Advance online publication. https://doi.org/10.1080/17549507.2022.2115136

Chen, J., Hynes-Berry, M., Abel, B., Sims, C., & Ginet, L. (2017). Nurturing mathematical thinkers from birth: The why, what and how. *Zero to Three, 37*(5), 31–33. https://www.zerotothree.org/resource/vol-37-no-5-building-stem-science-technology-engineering-math-skills-from-birth/

Clerkin, A., & Gilligan, K. (2018). Preschool numeracy play as a predictor of children's attitudes towards math at age 10. *Journal of Early Childhood Research, 16*(3), 319–334. https://doi.org/10.1177/1476718X18762238

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd edn). Routledge.

Daraganova, G., & Sipthorp, M. (2011). *Wave 4 weights: The longitudinal study of Australian children* (LSAC Technical Paper No. 9). Australian Government, Australian Institute of Family Studies. https://growingupinaustralia.gov.au/sites/default/files/tp9.pdf

Department of Education and Early Childhood Development. (2013). *Early bird catches the worm: The causal impact of pre-school participation and teacher qualifications on year 3 national NAPLAN cognitive tests*. https://www.education.vic.gov.au/Documents/about/research/preschoolparticipationandqualissummarypaper2013.pdf

Department of Education and Training. (2019). *2018 AEDC national report*. https://www.aedc.gov.au/resources/detail/2018-aedc-national-report

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duncan, R. J., Duncan, G. J., Stanley, L., Aguilar, E., & Halfon, N. (2020). The kindergarten Early Development Instrument predicts third grade academic proficiency. *Early Childhood Research Quarterly, 53*(4), 287–300. https://doi.org/10.1016/j.ecresq.2020.05.009

Education Council. (2019). *Alice Springs (Mparntwe) education declaration*. https://www.education.gov.au/alice-springs-mparntwe-education-declaration/resources/alice-springs-mparntwe-education-declaration

Elek, C., Gray, S., West, S., & Goldfeld, S. (2022). Effects of a professional development program on emergent literacy-promoting practices and environments in early childhood education and care. *Early Years*, 42(1), 88–103. https://doi.org/10.1080/09575146.2021.1898342

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Studies, 13*(2), 241–254. https://doi.org/10.1111/1467-9280.00427

Fisher, W. P., Jr. (1992). Reliability statistics. Rasch Measurement Transactions, 6(3), 238. www.rasch.org/rmt/rmt63i.htm

Fisher, W.P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions, 21(1)*, 1095.

Geary, D. (2022). *Setting the preschool foundation for success in mathematics*. https://www.cis.org.au/publication/setting-the-preschool-foundation-for-success-in-mathematics/

Goldfeld, S., O'Connor, E., O'Connor, M., Sayers, M., Moore, T., Kvalsvig, A., & Brinkman, S. (2016). The role of preschool in promoting children's healthy development: Evidence from an Australian population cohort. *Early Childhood Research Quarterly, 35*, 40-48.

Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, L., Perry, L-A., Roberts, M. (2018). *Through growth to achievement: The report of the review to achieve educational excellence in Australian schools*. Commonwealth of Australia. https://www.education.gov.au/quality-schools-package/resources/through-growth-achievement-report-review-achieve-educational-excellence-australian-schools

Goss, P., Sonnemann, J., Chisolm, C., & Nelson, L. (2016). *Widening gaps: What NAPLAN tells us about student progress*. Grattan Institute. https://grattan.edu.au/report/widening-gaps/

Harris, F., Smith, M., Laurens, K. R., Green, M. J., Tzoumakis, S., Kariuki, M., & Carr, V. J. (2018). Validation of a two-factor model of the Best Start Kindergarten Assessment of literacy and numeracy. *Australian Journal of Education, 62*(1), 36–48. https://doi.org/10.1177/0004944118761381

He, V. Y., Nutton, G., Graham, A., Hirschausen, L., & Su, J.-Y. (2021). Pathways to school success: Self-regulation and executive function, preschool attendance and early academic achievement of Aboriginal and non-Aboriginal children in Australia's Northern Territory. *PLOS ONE, 16*(11), e0259857. https://doi.org/10.1371/journal.pone.0259857

Heckman, J. J. (2011). The economics of inequality: The value of early childhood education. *American Educator, 35*(1), 31.

Huo, S., Zhang, X., & Law, Y. K. (2021). Pathways to word reading and calculation skills in young Chinese children: From biologically primary skills to biologically secondary skills. *Journal of Educational Psychology, 113*(2), 230–247. https://doi.org/10.1037/edu0000647

Kral, I., Fasoli, L., Smith, H., Meek, B., & Phair, R. (2021). A strong start for every Indigenous child (OECD Education Working Papers, No. 251). OECD Publishing. https://dx.doi.org/10.1787/ebcc34a6-en

Lee, S. (2012). Toddlers as mathematicians? Australasian Journal of Early Childhood, 37(1), 30–37. https://doi.org/10.1177/183693911203700105

Lennox, M., & Westerveld, M. (2014). Enhancing emergent literacy performance of Australian students from disadvantaged backgrounds in their first year of school. *Journal of Clinical Practice in Speech-Language Pathology, 16*(2), 58–65.

Linacre, J.M. (2009). *A User's Guide to WINSTEPS*. Chicago Il: Winsteps.com.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis With Missing Data*. New York: Wiley.

Lonigan, C. J., & Shanahan, T. (2008). *Developing early literacy: Report of the National Early Literacy Panel executive summary*. https://lincs.ed.gov/publications/pdf/NELPSummary.pdf

MacDonald, A. (2018). *Mathematics in early childhood education*. Oxford University Press.

Montoya, M. F., Susperreguy, M. I., Dinarte, L., Morrison, F. J., San Martín, E., Rojas-Barahona, C. A., & Förster, C. E. (2019). Executive function in Chilean preschool children: Do short-term memory, working memory, and response inhibition contribute differentially to early academic skills? *Cross-Domain Development of Academic and Cognitive Skills, 46*, 187–200. https://doi.org/10.1016/j.ecresq.2018.02.009

Neumann, M. M. (2016). A socioeconomic comparison of emergent literacy and home literacy in Australian preschoolers. *European Early Childhood Education Research Journal, 24*(4), 555–566. https://doi.org/10.1080/1350293X.2016.1189722

Nicholas, M. & Rouse, E. (2021). Learning to read: Where should early childhood educators begin? *Literacy, 55*(1), 3–13. https://doi.org/10.1111/lit.12229

O'Connor, M., Chong, S., Quach, J., & Goldfield, S. (2019). Learning outcomes of children with teacher-identified emerging health and developmental needs. *Child: Care, Health and Development, 46*(2), 223–231. https://doi.org/10.1111/cch.12737

Organisation for Economic Co-operation and Development. (2018). *Early Learning Matters*. https://www.oecd.org/education/school/Early-Learning-Matters-Project-Brochure.pdf

Organisation for Economic Co-operation and Development. (2020). *Early learning and child well-being: A study of five-year-olds in England, Estonia, and the United States*. OECD Publishing. https://dx.doi.org/10.1787/3990407f-en

Perry, B., Gervasoni, A., Hampshire, A., & O'Neill, W. (2016). 'Let's count': Improving community approaches to early years mathematics learning, teaching and dispositions through noticing, exploring and talking about mathematics. In B. White, M. Chinnappan, & S. Trenhom (Eds.), *Proceedings of the 39th Annual Conference of the Mathematics Education Research Group of Australasia*. Mathematics Education Research Group of Australasia. https://merga.net.au/Public/Public/Publications/Annual_Conference_Proceedings/2016_MERGA_Conference_Proceedings.aspx

Raghubar, K. P., & Barnes, M. A. (2017). Early numeracy skills in preschool-aged children: A review of neurocognitive findings and implications for assessment and intervention. *The Clinical Neuropsychologist, 31*(2), 329–351. https://doi.org/10.1080/13854046.2016.1259387

Sacks, D., Bayles, K., Taggart, A., & Noble, S. (2020, June). COVID-19 and education: How Australian schools are responding and what happens next. *Government Matters*. https://www.pwc.com.au/government/government-matters/covid-19-education-how-australian-schools-are-responding.html

Sanson, A., Nicholson, J., Ungerer, J., Zubrick, S., Wilson, K., Ainley, J., Berthelsen, D., Bittman, M., Broom, D., Harrison, L., Rodgers, B., Sawyer, M., Silburn, S., Strazdins, L., Vimpani, G., & Wake, M. (2002). *Introducing the longitudinal study of Australian children* (LSAC Discussion Paper No. 1). Australian Institute of Family Studies. https://growingupinaustralia.gov.au/sites/default/files/discussionpaper1.pdf

Saunders, V., Beck, M., McKechnie, J., Lincoln, M., Phillips, C., Herbert, J., & Davey, R. (2022). A good start in life: Effectiveness of integrated multicomponent multisector support on early child development – Study protocol. *PLOS One, 17*(8), Article e0267666. https://doi.org/10.1371/journal.pone.0267666

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research, 8(1)*, 3-15.

Serpell, S., Baker, L., & Sonnenschien, S. (2004). *Becoming literate in the city*. Cambridge University Press.

Social Research Centre. (2019). *2018 AEDC data collection technical report*. https://www.aedc.gov.au/resources/detail/2018-aedc-data-collection-technical-report

Soloff, C., Lawrence, D., & Johnstone, R. (2005). *The longitudinal study of Australian children: Sample design* (LSAC Technical Paper No. 1). https://doi.org/10.13140/2.1.2587.8728

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/a:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tayler, C., Cloney, D., & Niklas, F. (2015). A bird in the hand: Understanding the trajectories of development of young children and the need for action to improve outcomes. *Australasian Journal of Early Childhood, 40*(3), 51–60. https://doi.org/10.1177/183693911504000308

Taylor, C. L., Christensen, D., Venn, A. J., Preen, D. B., Stafford, J., Hansen, E., Jose, K., & Zubrick, S. R. (2022). Use of administrative record linkage to examine patterns of universal early childhood health and education service use from birth to Kindergarten (age four years) and developmental vulnerability in the Preparatory Year (age five years) in Tasmania, Australia. International Journal of Population Data Science, 6(3). https://doi.org/10.23889/ijpds.v6i3.1681

Viner, R., Russell, S., Crocker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *The Lancet, 4*(5), 397–404. https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(20)30095-X/fulltext

Warren, D., Daraganova, G. & O'Connor, M. (2018). *LSAC Annual Statistical Report 2017 chapter*. Australian Institute of Family Studies. https://growingupinaustralia.gov.au/research-findings/annual-statistical-report-2017/preschool-and-childrens-readiness-school

Williams, K., White, S., & MacDonald, A. (2016). Early mathematical achievement of boys and girls: Do differences in early self-regulation pathways explain later achievement? *Learning and Individual Differences, 51*, 199–209. https://doi.org/10.1016/j.lindif.2016.09.006

Williams, K. E., Berthelsen, D., & Laurens, K. R. (2022). Academic resilience from school entry to third grade: Child, parenting, and school factors associated with closing competency gaps. *PLOS One, 17*(11). https://doi.org/10.1371/journal.pone.0277551

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Mesa Press.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month old infants. *Cognition, 74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Yelland, N., Diezmann, C., & Butler, D. (2014). *Early mathematical explorations.* Cambridge University Press.

Zubrick, S. R., Taylor, C. L., & Christensen, D. (2015). Patterns and predictors of language and literacy abilities 4-10 years in the Longitudinal Study of Australian Children. *PLOS One, 10*(9). https://doi.org/10.1371/journal.pone.0135612

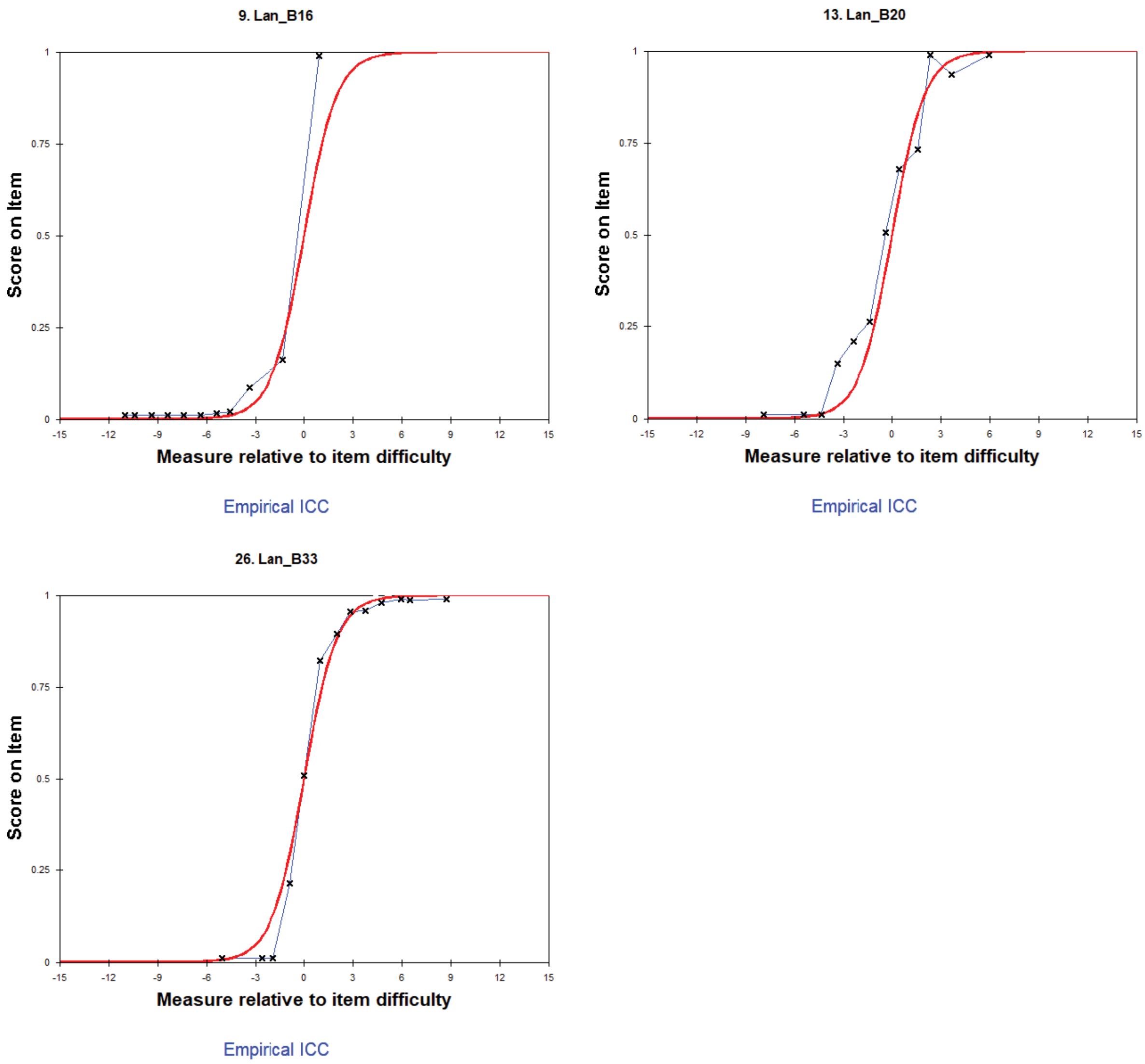
Appendices

Appendix A – Item fit analysis

Item fit analysis was performed to determine whether responses associated with each indicator reflect model expectations. Infit and outfit mean square statistics were reported in Table 2. Results showed that all indicators had Infit statistics within the acceptable range of 0.6 to 1.4. Notably, 3 indicators (one from Basic Numeracy, 2 from Advanced Literacy subdomains) had high outfit (>1.4, bold in Table 2), falling outside the acceptable range. This indicates that these indicators may be under-fit and potentially did not meet the model expectation well. The item characteristics curve (ICC) was generated for each of these 3 indicators to provide more insight for the high outfit statistics.

This shows that these indicators were not under-fit, meaning that they were able to distinguish between high and low performers. The high outfit statistics were most likely contributed by the unmodelled noise in the response data.

Figure A1: Item characteristics curves for B16, B20 and B33 indicators in the Language and Cognitive Skills domain



Appendix B – Impact of indicators on later years' learning outcomes

Table B1: Regression coefficients for the multilevel model on Year 3 NAPLAN Numeracy results

Variable	Coefficient	Standard error	P-value
AEDC indicator			
Attach sounds to letters	12.073	3.813	0.002
Show awareness of rhyming words	11.015	4.250	0.010
Read complex words	30.154	3.261	<0.001
Interested in writing voluntarily	10.530	5.732	0.066
Write simple sentences	7.289	2.240	0.001
Remember things easily	19.399	2.799	<0.001
Count to 20	20.767	3.756	<0.001
Recognise numbers 1 to 10	31.370	5.876	<0.001
Children-level demographics			
Age at Year 3 NAPLAN test	0.002	0.542	0.997
Female (ref: Male)	-19.819	2.960	<0.001
Aboriginal (ref: Non-Aboriginal)	-18.936	6.872	0.006
LBOTE (ref: Non-LBOTE)	15.867	3.815	<0.001
Special needs (ref: Not special needs)	-25.542	7.180	<0.001
School-level demographics			
School ICSEA	0.185	0.017	<0.001
Catholic sector (ref: Government)	-8.145	4.156	0.050
Independent sector (ref: Government)	-1.622	4.769	0.734
Other Urban population >1,000 and <99,999 (ref: Major Urban population >=100,000)	-8.318	2.640	0.002
Bounded Locality (ref: Major Urban population >=100,000)	3.218	3.439	0.349
Rural Balance the remainder of State/Territory (ref: Major Urban population >=100,000)	3.601	5.627	0.522
Constant	120.845	63.798	0.058

*P-value highlighted in teal indicates statistical significance of the coefficient at 5% significance level

Table B2: Regression coefficients for the multilevel model on Year 3 NAPLAN Reading results

Variable	Coefficient	Standard error	P-value
AEDC indicator			
Identify some letters of the alphabet	15.776	5.911	0.009
Attach sounds to letters	19.476	6.015	0.001
Read complex words	34.318	6.774	<0.001
Interested in writing voluntarily	14.446	6.387	0.024
Write simple sentences	11.271	5.091	0.027
Remember things easily	20.971	4.053	<0.001
Count to 20	17.491	6.778	0.010
Recognise numbers 1 to 10	34.029	6.920	<0.001
Children-level demographics			
Age at Year 3 NAPLAN test	0.448	0.571	0.432
Female (ref: Male)	-1.702	3.079	0.580
Aboriginal (ref: Non-Aboriginal)	-36.353	9.847	<0.001
LBOTE (ref: Non-LBOTE)	1.036	5.202	0. 842
Special needs (ref: Not special needs)	-47.346	10.246	<0.001
School-level demographics			
School ICSEA	0.234	0.023	<0.001
Catholic sector (ref: Government)	-7.658	5.383	0.155
Independent sector (ref: Government)	-2.123	5.210	0.684
Other Urban population >1,000 and <99,999 (ref: Major Urban population >=100,000)	-11.160	4.512	0.013
Bounded Locality (ref: Major Urban population >=100,000)	13.118	5.367	0.015
Rural Balance the remainder of State/Territory (ref: Major Urban population >=100,000)	3.983	6.171	0.519
Constant	25.635	76.624	0.738

*P-value highlighted in teal indicates statistical significance of the coefficient at 5% significance level

Table B3: Regression coefficients for the multilevel model on Year 3 NAPLAN Writing results

Variable	Coefficient	Standard error	P-value
AEDC indicator			
Identify some letters of the alphabet	12.519	6.417	0.052
Read complex words	22.582	1.278	<0.001
Aware of writing directions in English	12.433	5.580	0.026
Interested in writing voluntarily	10.559	4.109	0.010
Write his/her own name in English	17.508	8.007	0.034
Write simple sentences	7.359	4.618	0.111
Remember things easily	12.361	4.010	0.002
Use one-to-one correspondence	20.532	7.023	0.004
Recognise numbers 1 to 10	22.342	8.042	0.005
Children-level demographics			
Age at Year 3 NAPLAN test	-0.196	0.246	0.426
Female (ref: Male)	18.190	1.502	<0.001
Aboriginal (ref: Non-Aboriginal)	-24.191	5.416	<0.001
LBOTE (ref: Non-LBOTE)	14.843	3.711	<0.001
Special needs (ref: Not special needs)	-19.103	7.974	0.017
School-level demographics			
School ICSEA	0.115	0.017	<0.001
Catholic sector (ref: Government)	0.344	3.436	0.920
Independent sector (ref: Government)	5.121	3.435	0.136
Other Urban population >1,000 and <99,999 (ref: Major Urban population >=100,000)	-9.097	1.946	<0.001
Bounded Locality (ref: Major Urban population >=100,000)	-5.581	4.053	0.169
Rural Balance the remainder of State/Territory (ref: Major Urban population >=100,000)	-4.554	3.837	0.235
Constant	201.911	30.562	<0.001

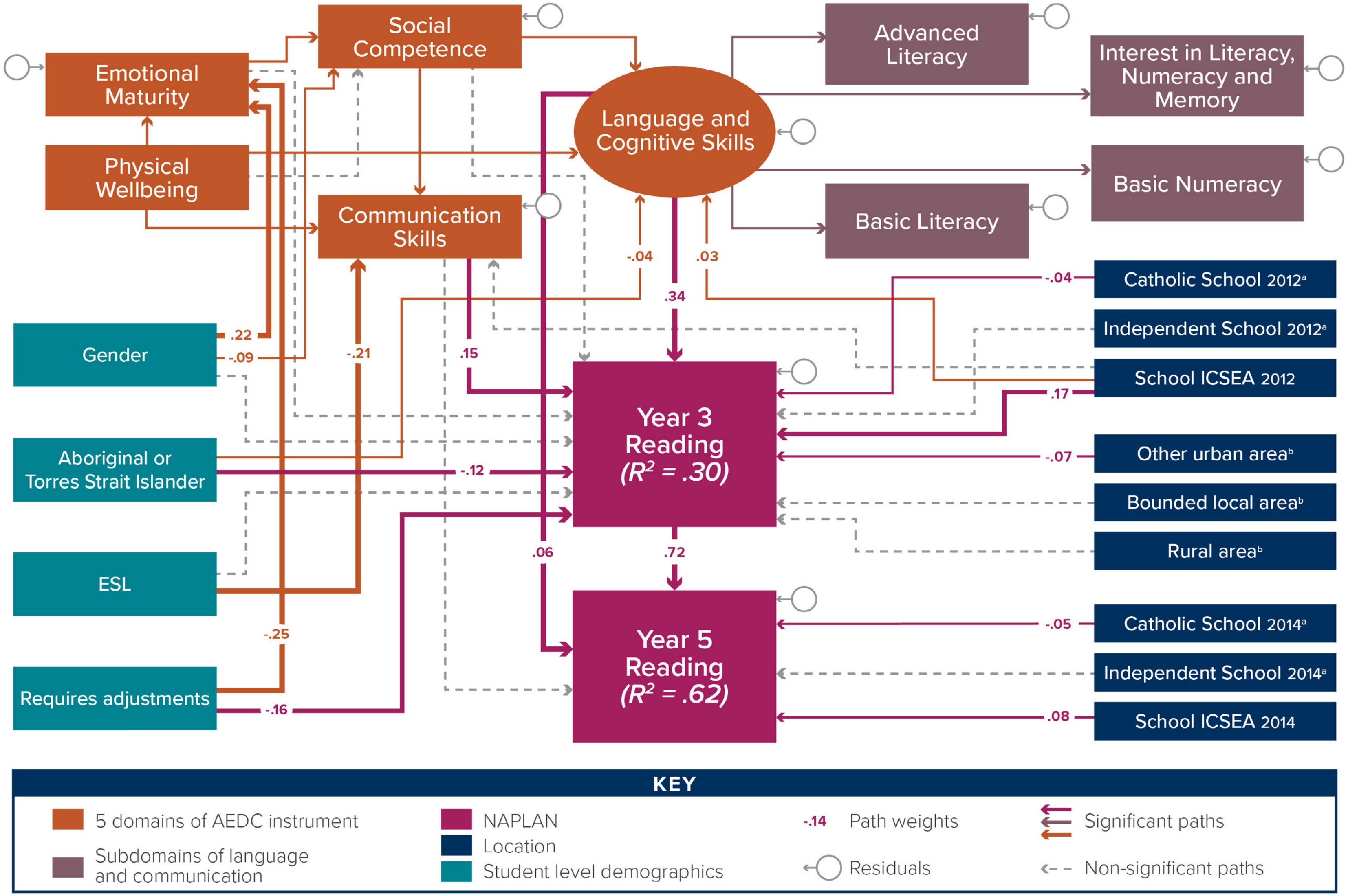
*P-value highlighted in teal indicates statistical significance of the coefficient at 5% significance level

Appendix C – Replicating the SEM model results using a latent factor of Language and Cognitive Skills

This set of models replicated the model results that were reported above (Section 4.1.2), with the key difference of measuring the Language and Cognitive Skills domain as a latent variable (see Section 3.3.1) defined by the observed scores on the 4 subdomains of Basic Literacy, Advanced Literacy, Basic Numeracy, and Interest in Literacy, Numeracy and Memory.

When the subdomains of AEDC Language and Cognitive Skills were modelled as indicators of a latent skills variable, the models showed consistent results to those reported above already. As shown in Figure 5 the explained variance was 30% and 62% for Year 3 and Year 5 Reading with significant direct impact shown from the latent Language and Cognitive Skills factor and indirect effects shown from Communication Skills and General Knowledge. These 2 domains' mediator roles were also significant in carrying impact indirectly from the other 3 AEDC domains to NAPLAN outcomes. This model presents a rigorous statistical analysis by examining the AEDC Language and Cognitive Skills domain as a latent construct. The consistency of these results with the other SEM results provides evidence for the robustness of the modelling that has been used.

Figure C1: Imputed data model with subdomains as indicators of latent Language and Cognitive Skills and other AEDC factors explaining Year 3 and Year 5 Reading scores when demographic variables were controlled: standardised single-level SEM coefficients



Note. Model fit: χ^2 (124, N = 2,457) = 380.65, p < 0.001; RMSEA = .03; CFI = .94; TLI = .92; SRMR = .03.

Australian Education Research Organisation

For more information visit
edresearch.edu.au

040923.EXT.AA.V1