Mathematics Education Research Journal
https://doi.org/10.1007/s13394-023-00447-z

ORIGINAL ARTICLE

Methodologies to reveal young Australian Indigenous students' mathematical proficiency

Bronwyn Reid O'Connor[1]

Received: 25 April 2022 / Revised: 18 January 2023 / Accepted: 25 January 2023
© The Author(s) 2023

Abstract

A 7-month mathematics proficiency program was conducted in a primary Australian Indigenous community school. This paper focuses on outlining the specific methodologies employed to explore how students' mathematical proficiency changed throughout the implementation of the program in Years 2 to 4 (~7 to 9 years old). A mixed methods research design was utilised, and findings will be presented to evidence how the combination of standardised tests, diagnostic tests, and Newman interviews were useful in capturing and making visible young Indigenous student's mathematical proficiency. Whilst standardised tests provided a useful and comparable measure of student achievement, diagnostic tests and Newman interviews gave space for Indigenous student voice and demonstrated their strengths and areas for improvement in relation to their conceptual understanding, procedural fluency, and strategic competence. From these findings, recommendations concerning the adjustment of data collection procedures for young students in this setting are presented. The findings question the accuracy of standardised tests in revealing young students' proficiency, and this has implications for the extent to which standardised tests are relied upon to inform educational reform particularly for Indigenous students. Striving for equitable educational outcomes is an important endeavour in Australia, and such undertakings must be driven by meaningful and accurate evidence of students' proficiency in mathematics.

Keywords Indigenous education · Mathematics assessment · Diagnostic assessment · Newman interviews

Introduction

The heart of this study, exploring young Indigenous students' mathematical proficiency, stemmed from the firm belief in young Indigenous students' capability in mathematics, a sentiment reiterated in other seminal work in the field (e.g.

Missing Open Access funding information has been added in the Funding Note.

✉ Bronwyn Reid O'Connor
bronwyn.reidoconnor@sydney.edu.au

[1] University of Sydney, Sydney, Australia

Matthews et al., 2003; Miller & Armour, 2021; Sarra & Ewing, 2014; Warren & DeVries, 2008). Mathematical proficiency encompasses the qualities that students should develop when studying mathematics at school. There are five interrelated strands that encompass mathematical proficiency including procedural fluency, conceptual understanding, strategic competence, adaptive reasoning, and productive dispositions (Kilpatrick et al., 2001). It is important to acknowledge that school mathematics is only one aspect of developing numerate students; however, it is an important form of knowledge that has equity connotations (Morris & Matthews, 2011; Thornton, 2020).

The aim of this paper is to present a discussion of methodological approaches that made visible the mathematical thinking of young Indigenous students. Thus, the research question answered in this paper is "what do various assessment forms reveal about young Indigenous students' mathematical proficiency?". Exploring methodological approaches that reveal young Indigenous students' mathematical thinking is important as measures of students' proficiency often drive reform and the adoption of pedagogical practices. The quality of early mathematics education is undeniably critical in supporting future mathematical success (Aubrey et al., 2006; Doig et al., 2003; van Tuijl et al., 2001), and it is important that approaches adopted by schools and teachers to improve outcomes for Indigenous students are accurately informed. Considering assessment practices that meaningfully and completely capture young Indigenous students' mathematical proficiency is important as "assessment in the mathematics classroom has the potential to alter the experiences of children in our schools at every level" (Serow et al., 2016, p. 235).

Currently, there is limited research that centres the voice of young Indigenous students' mathematical thinking and an overrepresentation of their mathematics achievement on standardised tests. The continuation of a reliance on standardised testing alone to ascertain Indigenous students' capabilities in mathematics is problematic (Rigney et al., 2020). Results from standardised testing have largely driven the deficit positioning of the mathematical capability of Indigenous students' that has persisted for the last three decades in Australia. The reliance on standardised testing alone is important to question given that there are numerous unresolved issues concerning what a good test looks like and what mathematics is privileged in being examined (Black & Wiliam, 2005; Niss, 2007; Suurtaam, et al. 2016). Furthermore, there are reported equity issues associated with standardised tests due to economic and cultural biases in test items (Carstairs et al., 2006; Eckermann et al., 2006; Hambleton, & Rodgers, 1995; Klenowski, 2014; Mellor & Corrigan, 2004; Tripcony, 2002). Vulnerable groups of students (e.g. girls, students from low socioeconomic backgrounds, culturally diverse students) are often reported as performing below majority groups in most standardised assessment in mathematics (e.g. TIMSS; Thomson et al., 2020). This phenomenon is not unique to Australian Indigenous students and has been noted in many other international contexts (e.g. Kanu, 2007; Lee, 2015; Özerk & Whitehead, 2012; Philpott, 2006; Solano-Flores et al., 2015; Trumbull & Nelson-Barbar, 2019). Using standardised testing alone as a methodological approach to explore young students mathematical thinking can limit our ability to recognise important success stories, particularly for Indigenous students. More importantly, it

conceals directions for teachers to productively capitalise on students' strengths based on their current mathematical understanding (Ginsburg, 2016; Wager et al., 2015). Thus, exploring ways in which other assessment methods can more meaningfully capture students' proficiency by giving space to students' voices revealing a child's viewpoint is important.

Drawing on sociocultural perspectives in this study, it is acknowledged that any assessment involves social negotiation (Preston & Claypool, 2021). The pursuit of equitable assessment for Indigenous students is "more of a sociocultural issue than a technical matter" (Klenowski, 2009, p. 89). In undertaking assessment, a sociocultural perspective recognises that the cultural socialisation of learners influences their thinking, communication, learning, and motivation (Nelson-Barber & Trumbull, 2007; Vygotsky, 1978). Thus, culture should be considered when developing and administering and assessment (He & van de Vijver, 2012), as well as what the assessor and student bring to the assessment task (Elwood, 2008). A sociocultural perspective shifts focus from what a student knows to what they can do, meaning that assessment moves from something being done to students to something that is being done with and for students (Klenowski, 2009). Therefore, in this paper the implications of assessment methods for investigating the mathematical thinking of Indigenous students are considered in light of these considerations.

This study examined the mathematical proficiency of young Indigenous students through the combination of a standardised test, diagnostic tests, and Newman's interviews. The ways in which the combination of these specific data sources made visible young Indigenous students' mathematical proficiency will be presented. Though this paper focuses on methodological approaches to examining students' proficiency, findings for each data source will be presented to demonstrate how the combination of data sources facilitated the triangulation of findings and made students' proficiency visible. The way in which mathematical proficiency is conceptualised in this study will be discussed in the following section.

Mathematical proficiency

The conceptualisation of mathematical proficiency in this study drew on Kilpatrick et al.'s (2001) foundational work describing mathematical proficiency as five interrelated strands. Kilpatrick et al.'s conceptualisation of mathematical proficiency supported the research design and data analysis in this study. Four of the five proficiency strands are cognitively focused (procedural fluency, conceptual understanding, adaptive reasoning, and strategic competence), and one is related to affective factors (productive dispositions). To be proficient in mathematics, students must have *procedural fluency*, encompassing the ability to apply mathematical procedures efficiently, appropriately and accurately (Hiebert & Lefevre, 1986). They must also possess a *conceptual understanding* of the related mathematical concepts and procedures. The complex and bi-directional relationship between the development of procedural fluency and conceptual understanding has been well discussed (e.g. Carpenter, 1986; Haapasalo & Kadijevich, 2000; Rittle-Johnson et al., 2015). Other strands critical to proficiency in mathematics include *adaptive reasoning,* referring

to a student's ability to justify and reflect on their understanding of mathematical concepts, and *strategic competence,* which encompasses the ability to formulate and represent mathematical problems appropriately (Kilpatrick et al., 2001). These four cognitive proficiency strands are the focus of mathematical proficiency as defined in the Australian curriculum (ACARA, 2022).

In addition to the four cognitively focused strands, a *productive disposition* is a critical affective strand of mathematics proficiency. The presence of productive dispositions could be observed through students' behavioural engagement in the mathematical tasks administered. Productive dispositions refer to a student's ability to perceive mathematics as useful and worthwhile, and to envisage themselves as an effective and capable learner of mathematics (Philip & Siegfried, 2015; Woodward et al., 2017). Affective factors are important as there is a relationship between academic achievement and students' attitudes towards mathematics (i.e. productive dispositions) (Ma, 1997; Ma & Xu, 2004). This is because students' productive dispositions influence their motivation, specifically their willingness to persevere with tasks until completion (Wigfield et al., 2015). Lacking a productive disposition towards mathematics can manifest as avoidance (Woodward et al., 2017). The ways in which particular methods of assessment reveal students' proficiency, in relation to the five strands, will be discussed in the following section.

Exploring young Indigenous students' mathematical proficiency: literature review

Internationally, there are gaps in current research surrounding assessment practices for Indigenous students (Preston & Claypool, 2021). In a systematic review, Miller and Armour (2021) analysed the methodologies of 28 empirical studies which examined the impact of teaching and learning strategies on Australian Aboriginal and Torres Strait Islander student outcomes. Of these 28 studies, 17 collected data from students and four focused solely on assessing young students mathematical thinking. It is important to note here that Miller and Armour stress that the limitations of a systematic review of empirical studies is that at times it can privilege the voice of non-Indigenous researchers. This can also have impacts on research design for these studies as it is likely a Western frame has been used for the research design and the analysis of data. However, this review reveals that there is a gap in research focusing on young Indigenous students' mathematical thinking. Miller and Armour point out that "many studies only reported teachers' perceptions of students learning" (2021, p. 74). In particular, only three of these studies gave a balance of voice by including interviews with young Indigenous students. This points to a large gap in research over the past 30 years and, if we continue to seek equitable educational outcomes in Australia, this is problematic.

Standardised tests

Despite this identified gap in research, the limitations of utilising standardised tests alone to measure young students mathematical thinking are often acknowledged. In

addition to the limitations discussed earlier, literacy and test taking skills also need to be considered when examining the ability of a standardised test to accurately reflect any young student's proficiency in mathematics. Taking a sociocultural perspective, this is particularly relevant for Indigenous students who may bring to school different cultural capital influencing test taking skills (Klenowski, 2009; Sarra, 2011; Tripcony, 2002; Zevenbergen, 2004). Thus, it is understood that reliance on standardised testing alone to ascertain Indigenous students' capabilities in mathematics can be problematic (Rigney et al., 2020), though they have been used in research to evaluate the effectiveness of interventions with Indigenous learners in Australia (e.g. Pegg & Graham, 2013). In addition to the above considerations, standardised assessment in mathematics, particularly multiple-choice assessment, also has limitations in its ability to distinguish whether an answer was a fortunate guess. Through such assessment it is difficult to examine students' reasoning, their confidence, or their disposition towards mathematics or the specific tasks (Perso, 2009). This indicates that standardised tests alone allow limited identification of students' mathematical strengths in relation to the five strands of mathematical proficiency. For these reasons, using standardised testing alone can limit our ability to meaningfully evaluate students' mathematical proficiency, particularly for Indigenous students. Despite these concerns, standardised tests in mathematics do have the capacity to provide a comparable pre- and post-program measure of changes in students' mathematics achievement, and such tests can "provide data about the transfer of basic fact knowledge to more complex academic and cognitive tasks" (Pegg & Graham, 2013, p. 12). Therefore, acknowledging the problematic nature of standardised tests in conjunction with its potential uses, it is argued that it is important to combine standardised tests with other assessment tools to meaningful gauge Indigenous learners mathematical thinking. That is, additional alternative assessment strategies are likely to give further useful insight when striving to gain a deeper understanding of Indigenous students' mathematical proficiency (English, 2016; Klenowski, 2009, 2014).

Diagnostic assessments

One assessment type that may compliment standardised tests in examining students' proficiency is diagnostic assessment. Diagnostic testing has long been recognised as an important component of teaching in mathematics, forming a cyclical approach to teaching and learning (Ashlock, 1976). Such testing has also been drawn upon in studies with young Australian Indigenous learners in other contexts (e.g. Miller, 2015; Warren & Miller, 2013). The importance of including a diagnostic tool is that, in addition to determining achievement levels (e.g. *what* students know), it is valuable to also determine students' understanding of concepts (e.g. *how* students know), and diagnostic assessment fulfils this role by enabling error analysis (Booker, 2011; Brown & Burton, 1978). Therefore, such assessment complements what a standardised test might relevel about what students know. Diagnostic assessment has the potential to reveal students' strengths in the four cognitively focused proficiency strands, depending on the test items, and can also shed light on the affective strand in relation to their behavioural engagement in the mathematical tasks, which is

related to productive dispositions. For example, Kilpatrick et al. (2001) propose that if a method of solving a particular problem (e.g. applying the addition algorithm) is incorrect, or misapplied, it could be considered that a student is still developing conceptual understanding of the associated concept. Also, if a student has chosen an incorrect method of solving (e.g. confusing addition for subtraction when solving worded problems), then that might indicate related difficulties associated with both strategic competence and conceptual understanding. Other conceptual errors might also relate to place value difficulties as conceptual understanding of multi-digit arithmetic is exemplified by fluid and flexible understanding of place value manipulations (Hiebert & Wearne, 1996). Analysing student scripts and distinguishing between conceptual understanding and procedural fluency difficulties is not always possible due to the interrelations between both (Kilpatrick et al., 2001). Furthermore, correct application of a procedure to solve an operation does not always imply that there was conceptual understanding. Schwab also surmises that written testing alone can be problematic as "not only that it does not work, but it confuses problem with symptom" (2012, p. 12). Therefore, there is a need to further explore and confirm students' proficiency, and this can be done by making students' voice central through other alternative assessments such as Newman interviews.

Newman interviews

Newman interviews are a meaningful form of alternative assessment in mathematics, particularly for young students. In other studies with young Indigenous students, interviews of some form were drawn upon as a feature of the methodological approach (e.g. Grootenboer & Sullivan, 2013; Papic, 2015). Leder and Forgasz (2006) note the need for research to obtain rich data beyond performance tests to ensure that the reflected behaviours adequately capture what happens in a classroom setting. Newman interviews provide a mechanism of providing such data but have not yet been drawn upon specifically in exploring the mathematical thinking of young Indigenous learners in Australia as studies have not focused on problem-solving processes. The first benefit of interviews is that it subverts some potential language and reading issues, and issues concerned with engaging in a formal testing discourse noted in earlier sections. The second is that the motivational impact of positive and timely affirmation of students' thinking has been well documented (e.g. Hattie, 2008), and interviews are one way of facilitating this. Thirdly, research in Indigenous education has indicated how young Indigenous learners are person-orientated, and interviews with students facilitate a direct social exchange between the assessor and student (Nichol & Robinson, 2000; Sarra et al., 2018; Warren & Miller, 2013). Nichol and Robinson (2000) describe how "students who feel personal connection with the teacher will be more cooperative, interested in learning, willing to take risks and attempt new tasks" (p. 500). The importance of the interviewer having an established relationship with the school and students has been noted in studies with young Indigenous students (e.g. Grootenboer & Sullivan, 2013; Miller, 2015; Papic, 2015). Students interview responses provide a mechanism to explore students' mathematical proficiency beyond what a written diagnostic test alone can provide. It also provides the opportunity to confirm students' strengths in

relation to the cognitively focused strands of proficiency, which may have been only speculative from diagnostic test responses. The affective proficiency strand relating to productive dispositions can also be observed through students' engagement with the questions. This study chose to draw on Newman interviews as it provided a framework for exploring students proficiency specifically when solving worded mathematics problems, complementing a written diagnostic test exploring students' proficiency with the same tasks. In Newman's seminal work on mathematical error analysis, it was determined that there are five distinct stages that a student must work through to successfully complete a worded problem (Newman, 1977, 1983). Newman's problem-solving stages are hierarchical, and success is necessary at each stage to accurately complete a problem-solving task (see Fig. 1).

The five stages of Newman's problem-solving hierarchy align and integrate with the definitions of the five strands of mathematical proficiency. That is, a student's interview may indicate that they are working towards developing *procedural fluency* with number facts if they demonstrate difficulties at the process skills stage. Alternatively, a student may be working towards developing their comprehension of the task or their skills in transforming the task into a suitable operation, which are both components of *strategic competence* and *conceptual understanding*. Therefore, error analysis of diagnostic assessment in conjunction with Newman interviews is likely to yield complimentary insight into students' mathematical proficiency.

Connecting assessment forms

The ways in which standardised tests, diagnostic tests, and Newman interviews work together to make visible students' mathematical proficiency is detailed in Fig. 2. Each assessment form has its own specific benefits and is valuable in revealing students' proficiency. It is the combination of assessment forms that is powerful in capturing students' proficiency.

Research design

Overview of methodology

The methodological approach for this study was an explanatory mixed methods design (Creswell & Plano Clark, 2017), aligning with pragmatist philosophies (Tashakkori & Teddlie, 2010). The mixing of methods has been proposed as a way of conducting

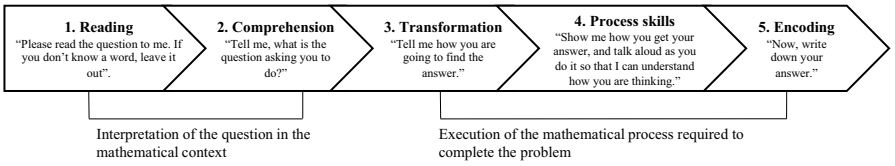


Fig. 1 Newman's hierarchy of solving a worded mathematics problem. Figure adapted from (Reid O'Connor & Norton, 2020) and questions adapted from White (2005)

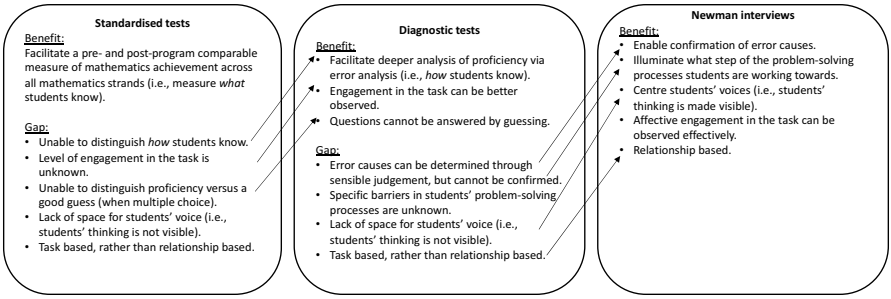


Fig. 2 Complimentary relationship between standardised tests, diagnostic tests, and Newman interviews in revealing student's' mathematical proficiency

research that supports an understanding of Indigenous ways of knowing, transcending what qualitative methods alone can provide (Botha, 2011). Quantitative data concerning students' proficiency was collected via two mathematics tests (a standardised assessment, and a diagnostic test) and an interview. Qualitative data included analysis of student errors on diagnostic tests, and student responses from interviews.

To ensure the values of reciprocity, respect, equality, responsibility, survival and protection, and spirit and integrity (NHMRC, 2018) were part of the design of the study, the research study was conducted with the first accountability being to the Indigenous school community (Denzin et al., 2008). Reciprocity was central to the study, therefore the school had first access to research findings, and the results were disseminated to teachers in a timely manner throughout the study so that they could be used by teachers to tailor learning experiences for the benefit of the students. Ethical clearance was obtained for this study and the school and parents provided informed consent prior to participation. Students and/or parents were able to withdraw at any time without penalty. Prior to gaining school consent, the school leadership team was consulted and drove the development of the aims of the study.

To guard against any potential author bias, the trustworthiness criteria (Guba & Lincoln, 1985) were applied. In particular, the criteria of confirmability (i.e. the degree to which findings are neutral and free of biases) was carefully considered due to the potential biases that may result from the researcher's unique background, experiences, and perspectives (Creswell, 2012). Triangulation through integration of both qualitative and quantitative data sources contributed to establishing the trustworthiness of this study as quantitative findings were supported by qualitative observations. The original sources of data that gave rise to particular assertions provided rich description and were made traceable by including original samples of students' work and verbatim quotes in the reporting of results where relevant. Though it is impossible to eliminate all researcher bias, the thorough documentation and multiple data collection opportunities assisted in recognising and minimising potential biases (Creswell, 2012).

Context of the study

The findings reported on in this paper are from a large study which evaluated the effectiveness of a mathematics proficiency program in one Indigenous community school. The purpose of this paper is not to present the impact of this program, but it is important to note the context in which data was collected (further detail on the program is detailed in Reid O'Connor, 2020, 2021; Reid O'Connor & Norton, 2020). Driven by a need identified and driven by the school community, the mathematics proficiency program developed for this study centred on a shared belief in aims and approach. The mathematics proficiency program was informed by a review of literature surrounding effective pedagogies in mathematics education. The literature was summarised into practice recommendations that were provided to teachers at the beginning of the study. The intention of review of effective pedagogies was to support teachers by providing direction for them to establish mathematics programs that were structured in a manner to provide opportunities to thoughtfully develop foundational mathematics concepts.

The practice recommendations, in brief, firstly included the suggestion to build the school mathematics programs in a way that involved high levels of teacher collaboration to develop a shared, school-wide approach (Boaler & Staples, 2008; Jorgensen, 2018; Stronger Smarter, 2014). In relation to broad pedagogical suggestions, the practice recommendations drew on a mastery learning approach where learning is recommended to be structured in small units of work based on pre-testing and concluded by post-testing (Kulik et al., 1990). Mastery learning involves teaching a topic until mastery is achieved and providing continuous feedback to students, resulting in mathematics learning that is not specifically time-bound, but rather progressed at a rate suitable for each cohort. Effective explicit instruction was also captured in the practice recommendations due to its alignment with mastery teaching approaches. Explicit instruction features logical sequencing of skills, and breaking down complex skills and strategies until mastery is achieved and units can be synthesised (Rosenshine, 2012). The importance of promoting initial success through guided and supported practice and immediate affirmative and corrective feedback to build confidence is also important when teaching explicitly (Archer & Hughes, 2011). Scaffolding learning using explicit teaching has also been described as a powerful pedagogy for Indigenous learners (Howard & Perry, 2011). The practice recommendations also highlighted the efficacy of teachers having high mathematical expectations of students and focusing on number as a priority (Jorgensen, 2018), and concrete-representational-abstract teaching sequences (Sarra et al., 2016).

Context of the study sample

The students in this study were of Aboriginal and/or Torres Strait Islander heritage, which encompasses a diverse range of people. The sample school in this study was an urban, community run school for Indigenous students in Australia. The school had a population of approximately 200 students with 93% identifying as Indigenous at the time of the study. All students spoke, and were taught in, English. Being community run, the school presented a unique educational opportunity for students.

Indigenous ownership for the school meant that there was an understanding of students' cultural experiences, with the expectation that teaching could build from these experiences, making instruction culturally responsive and culturally inclusive (Rigney et al., 2020; Sarra & Ewing, 2021). In the larger study, the sample consisted of four classes comprising Grade 2/3 ($n=12$), Grade 3/4 ($n=12$), Grade 4/5 ($n=12$), and Grade 5/6 ($n=14$). This paper reports on findings from the Year 2/3 ($n=12$) and Year 3/4 ($n=12$) cohort for brevity; however, trends in findings reported in this paper are consistent throughout all four cohorts.

The role of the researcher

As the researcher does not play a neutral role in the collection and interpretation of data or the delivery of the program, it is important to consider the researcher's role in the study. Research into the ways in which non-Indigenous researchers conduct research in Indigenous spaces, in both national and international context, notes that relationships built on mutual respect, and reciprocal exchange with Indigenous communities are foundational (e.g. Aveling, 2013; Kilian et al., 2019; Martin, 2002; Skille, 2021). The researcher was a non-Indigenous mathematics teacher who had taught at the school for several years prior to the study. The researcher's role was highly interactive with students throughout the study. As they had previously taught all students involved in the study, the researcher had developed relationships with students participating in the study.

The researcher's role was also highly interactive with teachers throughout the program's implementation. The program was guided by the researcher who delivered professional development, supported classroom teachers in the planning of mathematics programs, and provided ongoing mentorship throughout. The established relationship between the teachers and the researcher meant that the study began from a position of mutual respect, and reciprocal exchange. Overall, the researcher's history with the school and the study design focusing on teacher's autonomy ensured that this study fit the optimum scenario identified by Riley (2021). That is, the school community and researcher had a strong awareness of each other and were engaged in open and regular communication, and the core delivery of the program was community led.

Data sources

The mathematics proficiency program was conducted over a seven-month period from March to October of the school year in 2017. The PAT-M assessment was administered by the researcher, followed by the diagnostic tests, and then the Newman interviews to ascertain changes in students' mathematical proficiency (see timeline in Fig. 3). There was a two-week period between the PAT-M test and the diagnostic test, and a one-week period between the diagnostic test and Newman interviews.

Standardised achievement tests: the Progressive Achievement Test-Mathematics (PAT-M)

The PAT-M provides a measure of students' skills in, and understanding of, school mathematics (Stephanou & Lindsey, 2013). It is a widely used achievement test in Australia developed and published by the Australian Council for Educational Research. These tests have been designed to provide a snapshot of students' current achievement in mathematics, and to monitor student improvement over time (Stephanou & Lindsey, 2013). The PAT-M has been used in other studies with Indigenous students to measure program efficacy (e.g. Pegg & Graham, 2013), and the sample school was drawing on PAT-M test data prior to the study. Each year level has a specific test paper consisting of 30–40 multiple choice items across the mathematics strands of number, algebra, geometry, measurement, statistics, and probability. The results of the PAT-M are measured quantitatively on a numerical Rasch measurement scale allowing for tests for different year levels to be compared. The changes in each cohort's mean score from the beginning to the end of the program were calculated and compared to the national norming sample.

In administering the tests, it was understood by teachers in the context of the sample school that completing a standardised (or diagnostic) test as a whole class under test conditions for the students would not be possible. This was for many logistical reasons, including students needing support (e.g. not knowing how to attempt the test and asking clarifying questions, needing encouragement, or needing support in writing and recording answers in appropriate places), and issues with student focus for sustained periods of time under formal test conditions. To obtain responses from students that most accurately reflected their proficiency, the adjustment was made to administer the standardised and diagnostic tests in small groups (2–5 students) in a quiet, small group working space at the school with the researcher. According to test instructions, all tests require 40 min of testing time; however, students were allowed to continue until they had completed testing. For the PAT-M, all students were able to complete the test within the 40-min testing time.

Diagnostic tests

The diagnostic tests used in this study were from the Booker diagnostic tests (Booker, 2011). The Booker tests are designed for use in the classroom to diagnose error patterns in the domains of numeration (number and place value),

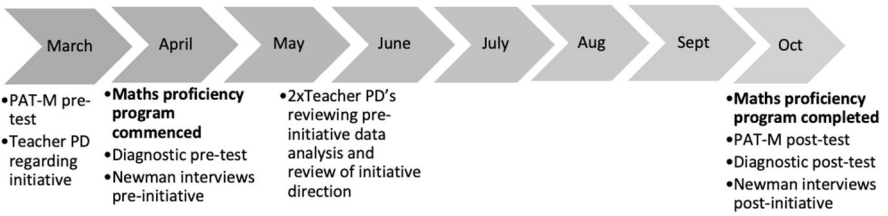


Fig. 3 Timeline for the study

addition, subtraction, multiplication, and division. There are five individual test papers assessing each of these domains. In selecting an appropriate diagnostic test for this study, the cultural relevance of the mathematics test items was considered as student's understanding of the test items would likely influence achievement. For example, studies in remote Indigenous contexts have demonstrated that context-based problem-solving tasks may be irrelevant for students if the questions are far removed from what is valued or experienced within their social and cultural environment (Grootenboer & Sullivan, 2013). The relevance of the specific test items was evaluated considering what was familiar to the everyday classroom discourse of the students in the study school. Student understanding of question contexts was not assumed during administration of the diagnostic tests, and the administering researcher remained available during testing time to provide clarity in relation to the context of the task if required. No such difficulties with interpreting contexts were found during administration of the written tests or interviews. This was anticipated as teachers in the study were already utilising Booker's tests prior to the study as the questions were deemed similar to what was used in mathematics classes already.

In this paper, for brevity, the findings of the addition diagnostic test only are reported as it provides sufficient detail to describe how this data sources provided insight into students' proficiency, and broad trends in findings were consistent across each diagnostic tests (see Reid O'Connor, 2020). The addition test begins by examining students' proficiency with single-digit addition facts, followed by asking the students to "write in words how you would read [4 + 3]" and "write a story to match this [4 + 3] addition". The test then assesses students' proficiency with addition computations in a progressive way. Worded problems are also mixed with algorithm questions in the test, preventing fatigue. Examples of the computation and worded problem questions are outlined in Table 1.

During administration of the tests, it was found that students in the Year 2/3 cohort were able to answer no more than one of the first four computation questions. Due to this, the remainder of the computation questions were not administered to Year 2/3. All Year 2/3 students were encouraged to attempt the worded problems; however, none did so.

When administering the test items that asked the students to "write in words how you would read 4 + 3" (for example) and "write a story to match this (4 + 3) addition",

Table 1 Addition computation and worded question examples

1) 73 + 37	2) 50 + 48	3) 537 + 256	4) 574 + 395	5) 326 + 578
6) 567 + 908	7) 6 + 9 + 4 + 8 + 5	8) 798 + 826 + 134 + 957	9) 7654 + 4396	10) 39,286 + 56,906

Worded problem: 1) Jackie had 138 marbles. She played with Cathy and Kylie and won 65 more marbles. How many marbles does she have now?

Worded problem: 2) Gabriel's parrot lost 7 black feathers and 6 blue feathers. How many feathers did his parrot lose?

Worded problem: 3) At the zoo, the giraffes each had 768 bales of hay. The seals eat 392 buckets of fish. The elephants eat 695 bales of hay. How many bales of hay were eaten?

NB: Computation questions were formatted vertically on the diagnostic test paper as Booker et al. determines that "a vertical recording of addition is preferable from the outset…to match the way materials are manipulated" (2014, p. 202–203)

the researcher chose to modify both questions to an oral form where students were asked to say the sum (e.g. 4 + 3) in words, and tell their story in words. The researcher then transcribed these question responses during testing. This ensured that literacy was not impeding students in answering the question, and it was found in observations that most students did not know what was meant and expected by "write in words how you would read 4 + 3". The researcher instead asked, "Can you read me this sum here?" and pointed to the sum, and "Can you tell me a story to match that sum?".

In relation to test administration, the suite of diagnostic tests typically takes approximately 60 min to complete (Booker, 2011); however, students were allowed to continue until they had completed testing. Students were permitted to return for multiple sessions to complete the test, either during a single day or across sequential days, to help prevent fatigue influencing results. This decision was made reflexively during the administration of these tests due to student need; it was found that many students were unable to attempt all five diagnostic tests in a single sitting. Facilitated by the researcher, the students were informed to complete the set questions to the best of their ability, and to show all working.

As well as analysing quantitative achievement by calculating class mean scores, analysis of the types of student errors on each incorrectly answered question was also carried out. Earlier discussion identified that delineations between strands of proficiency are difficult due to their interrelations, therefore understanding students' core difficulties is a complex matter. The intent of the error analysis carried out in this study was not to speculate on students' thinking, but to draw from literature and the researcher's experience to suggest in a reasonable manner what errors may have contributed to the incorrect answers. The small group administration of the test also meant that the researcher could observe the process of students attempts, which also provided insight when conducting error analysis. Once errors were analysed, categories of errors were coded following an emergent design as error categories were not predetermined. Analysis of trends in error patterns helped explain students' achievement on the diagnostic tests.

Newman interviews

Newman's error analysis interviews were conducted individually for Year 3/4 students in the study who answered at least one addition worded problem incorrectly. Students in Year 2/3 were not interviewed as they did not complete any worded questions from the diagnostic tests, indicating that they were still developing conceptual understanding of earlier addition concepts. Newman interviews identified where a student was making errors in a solving a worded question from a list of five steps (reading, comprehension, transformation, process skills, or encoding). These interviews built on standardised mathematics tests (PAT-M tests) and diagnostic tests conducted earlier in the study as they aided in identifying reasons behind students' errors.

The interviews were conducted with individual students by the researcher in a quiet small group working area at the school. The procedure for conducting a Newman interview involved making the student comfortable, then providing them with new copies of the first addition worded question that they answered incorrectly on the diagnostic tests (White, 2005). The student was then asked the Newman

interview questions outlined in Fig. 1 in sequential order. If the student answered an interview question incorrectly, they were then asked one question beyond the breakdown point to ensure their answer was not attributed to difficulties in expressing their knowledge of a particular step clearly (White, 2005). At the completion of the interview the error classification was decided by observing the first point where students answered the Newman interview question incorrectly. The frequency of breakdown points was then analysed allowing for the identification of trends. The comparisons between the pre- and post-program findings determined whether students progressed through the five steps of problem-solving. Detailed description of examples of each type of Newman error is outlined in Reid O'Connor and Norton (2020).

Findings

Standardised tests: Progressive Achievement Test-Mathematics

The standardised test administered in this study, the PAT-M, provided a comparable pre- and post-program measure of students' achievement in mathematics. For example, it was found that the Year 2/3 cohort reported an improvement of over 6 months within the timeframe of the 7-month program (the exact improvement in terms of time is difficult to identify as October of Year 1 is the first administration of the PAT-M to the norming sample). The Year 3/4 cohort reported an improvement of approximately 1 year and 2 months within the 7-month program, twice the expected gain. Whilst this level of analysis did provide important information demonstrating the positive impact of the mathematics proficiency program, students' strengths in relation to the five strands of mathematical proficiency could not be identified from the standardised test alone.

The PAT-M did enable deeper analysis of student's achievement by mathematics strand (i.e. number, algebra, measurement, statistics, probability), and item analysis of student's achievement on specific questions. The test completed by the Year 2/3 cohort at the beginning of the program contained one specific question related to addition. This question was "$?+14=22$". Of the sample of 11 students, it was found that one student was able to correctly answer this question, four answered "6", and six did not answer the question. In the PAT-M test administered to Year 3/4, there were two questions related to addition. These questions were "$25+?=55$" and "27 students from Room Three and 25 students from Room four went to the disco. How many students went altogether?". Four students out of 11 were able to correctly answer the first question, and two were able to answer the second. Whilst this indicates some areas for further attention for the classroom teacher, it does not give detailed insight into students thinking or explain why they might have been struggling with the addition concept. That is, it did not make visible whether students were still developing *procedural fluency* with addition, *conceptual understanding* of the addition concept, or *strategic competence* in selecting the correct operation (for the worded question only). It also could not be concluded whether students' achievement was impacted by the language of the question and their subsequent comprehension, or whether it was an issue of students' productive dispositions towards the

task. The PAT-M did not provide a specific measure of students' adaptive reasoning, or productive dispositions towards mathematics. The Year 2/3 students were also not provided with an opportunity to demonstrate their strategic competence with addition on the PAT-M.

An example of the limitations of the PAT-M in providing an accurate indication of students' mathematical proficiency was the case of a Year 2 student, Daisy, who reported the largest gains in achievement in the Year 2/3 cohort as measured by the PAT-M. Daisy's case demonstrated that achievement on the PAT-M did not equate to the presence of conceptual understanding of core mathematical concepts such as addition. This will be further discussed in the following section as diagnostic tests were able to provide greater depth of insight into students' mathematical proficiency through analysis of students' errors.

Diagnostic tests: Booker's tests

Analysis of diagnostic tests revealed students' thinking by making visible students' strengths or areas for improvement in relation to *conceptual understanding*, *procedural fluency*, and *strategic competence.* Though the diagnostic tests did not provide a specific measure of affect, it was also possible to observe whether students were demonstrating a *productive disposition* as evidenced by their engagement in completing or not completing the test tasks.

Though diagnostic tests enabled the calculation of changes in whole class mean achievement on addition test items, this level of analysis did not give insight into students' strengths in terms of the specific strands of mathematical proficiency. However, two particular questions ("can you read me this sum here", and "can you tell me a story to match that [4 + 3] sum?") did provide further insight into students' *conceptual understanding* of addition. For example, when asked to tell a story for single-digit addition, none of the Year 2/3 cohort were able to do so at the beginning or end of the program, and all indicated that they did not know how to do so. By the end of the program, three students attempted to tell a story (see Fig. 4).

When asked to tell a story for the question, half of the Year 3/4 students ($n = 12$) were able to do so by the end of the program. Incorrect responses are detailed in Fig. 5.

These findings suggested that *conceptual understanding* of addition was still developing for both cohorts, despite a reasonable level of *procedural fluency* with addition facts. This type of insight could not be gained from a written test alone. It was the researcher's one-on-one discussions, enabled by the adaption of the test administration, which allowed for meaningful examination of students' conceptual understanding of the addition concept. Therefore, in examining young students'

Post-initiative, Student 1:	"6 people on the bench and there were 2 people with hats."
Post-initiative, Student 2:	"6 and 2 went to the beach."
Post-initiative, Student 3:	"4 trees and 4 pencils."

Fig. 4 Responses from Year 2/3 students who answered when asked to tell a story for "6 + 2"

Student 1	Pre-initiative: Post-initiative:	4+3: "Jackie* had forty-three marbles. She played with 10. What would that equal?" 6+2: "The icecream man sold 2 icecream cones and only sold 6 icecreams."
Student 2	Pre-initiative: Post-initiative:	No attempt 6+2: "There were 6 bikes and 2 bikes stolen."
Student 3	Pre-initiative: Post-initiative:	Answered with a correct story. 6+2: "There were 6 carrots and 2 were taken away by rabbits."
Student 4	Pre-initiative: Post-initiative	4+3: "Miss Bronwyn had 4 rubbers and gave Miss Diane* 3 rubbers. How much does Bronwyn have left?" Answered with correct story
Student 5	Pre-initiative: Post-initiative	Answered with a correct story. 6+2: "Sarah* had 6 pieces of fudge and gave Sue* 2. How many does she have left?"
Student 6	Pre-initiative: Post-initiative	4+3: "Sue* had 4 dollars and Sarah* had 3 dollars. If they put it together how many iceblocks can they buy?" 6+2: "Sarah* had 2 cats and she had 3 dogs and 3 birds. How many pets did she have?"

Fig. 5 Responses from Year 3/4 students who incorrectly responded when asked to tell a story for "6 + 2" or "4 + 3". *Names have been changed for anonymity

mathematical proficiency, providing space for students' voices in communicating their mathematical thinking provided valuable insight.

Further analysis of students' proficiency was also made possible through error analysis as outlined in Table 2. For the sake of brevity, errors that occurred five or more times only are reported.

One prevalent error was not attempting questions. Not attempting test question occurs for two potential reasons: students do not know the how to attempt the computation (related to *conceptual understanding*), or students are displaying avoidance behaviours in the written test (related to *productive dispositions*). Analysis of responses to worded problems also indicated that non attempts were also prevalent (see Table 3), meaning that students had not made any written marks and had not attempted to record any sum that needed to be solved (regardless of whether they were able to solve it). This indicated that it was unlikely that non-attempts were caused by a lack of *procedural fluency*, particularly as questions where students experienced difficulties included those that were procedurally simple (i.e. only 50% of Year 3/4 students were able to answer "Gabriel's parrot lost 7 black feathers and 6 blue feathers. How many feathers did his parrot lose?" by the end of the program).

Table 2 Frequency of each error type for addition computations

Error	Year 2/3, $n = 12$ Error frequency Pre → post	Year 3/4, $n = 12$ Error frequency Pre → post
No attempt	14 → 0 (improvement)	7 → 13
Number fact error	2 → 9	8 → 11
Answer appears to be random values	21 → 4 (improvement)	5 → 0 (improvement)
Forgot to add on a renamed value	0 → 5	1 → 1
Forgot to record renamed value	2 → 4	8 → 1 (improvement)
Did not rename values, recorded double-digit answers into a single place value column	3 → 15	5 → 7
Started sum from the left	10 → 1 (improvement)	0 → 0

Table 3 Frequency of each error type for the addition worded problems

Error	Year 3/4, $n = 12$ Error frequency Pre → post
No attempt	8 → 4 (improvement)
Carried out question using subtraction	7 → 4 (improvement)
Did not isolate correct values	5 → 1 (improvement)
Unsure of error cause, no working shown	5 → 3 (improvement)

Therefore, the findings indicated that it is reasonable to consider *productive dispositions* towards the task as the root cause of non-attempts. As non-attempts of written test questions were prevalent, the diagnostic test as a data source in examining young Indigenous students' mathematical proficiency had limitations. However, the data provided insight into other important error patterns.

Recording seemingly random values was also a prevalent error for the Year 2/3 cohort, and Daisy demonstrated this error. This finding aligns with the work of Rogers (2014), who notes that renaming with 3-digit numbers is a significant difficulty for students. An example of Daisy's attempts at 3-digit addition involving renaming is outlined in Fig. 6.

Limited *procedural fluency* with addition number facts, or a developing *conceptual understanding* of renaming are potential explanations for Daisy's responses. However, her responses to other computation questions indicated that number facts did not appear to have impeded Daisy's proficiency unless renaming was involved. This suggested that it was the *conceptual understanding* of renaming concepts that needed to be further addressed. This was supported in other error trends for Year 2/3, which indicated that errors associated with renaming (such as failing to rename appropriately) were impacting students' proficiency (see Fig. 7). This depth of detail concerning students' proficiency was not able to be gained from the PAT-M.

In analysis of students' responses to addition worded questions, common errors were associated with *strategic competence*, as indicated by confusing the question with subtraction or difficulties isolating the correct values. This concurs with the confusion between addition and subtraction noted in students' addition stories outlined in Fig. 5. The frequency of worded problem errors is reported in Table 3, with errors with five or more occurrences reported.

It is important to note that error analysis of diagnostic tests did have limitations. For example, analysis indicated that number fact errors were a prevalent error with increases in frequency reported over the course of the program. A possible explanation

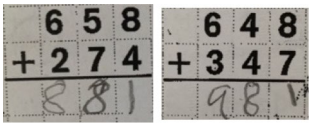


Fig. 6 Example of Daisy's errors (recording random values where renaming was required) in addition computation questions

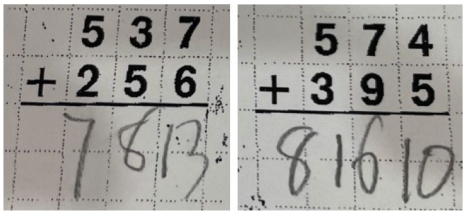


Fig. 7 Example of Year 4 student recording the total value for each place value instead of renaming

for this is carelessness, particularly given that students were largely *procedurally fluent* with addition number facts. However, whilst analysis of the diagnostic test provided insight into the error type, the root cause of many was speculative from this data alone. Newman interviews, discussed in the following section, provided further insight and demonstrated that it was not procedural fluency that was primarily impeding students.

Newman interviews

Newman interviews were able to confirm whether students' challenges identified from the diagnostic test were associated with *conceptual understanding, procedural fluency,* or *strategic competence*. Students' *productive dispositions* were also observable in their willingness to engage with the tasks during the interview. The frequency of Year 3/4 students' stopping point during the Newman problem solving interviews is reported in Table 4 below. The two sample sizes differ pre- and post-program. At the beginning of the program, one student was not administered the addition Newman interview as they correctly answered all addition worded questions on the diagnostic test. At the conclusion of the program, this occurred for three students.

Process skill errors were noted in the findings, and this is related to students' *procedural fluency*. However, comprehension difficulties were the most prevalent error at the beginning of the program. Analysis of students' interview responses revealed that being unable to identify what the question was asking them to do (i.e. comprehension) was often the cause of students choosing the incorrect operation or selecting the incorrect values. That is, comprehension impacted students' *conceptual understanding* of the problem, and their *strategic competence*. An example of this is outlined in the interview script in Fig. 8.

Table 4 Frequency of each error type for Year 3/4 from the addition diagnostic test worded problems

	First error present	
	Pre ($n=11$)	Post ($n=9$)
Answered question correctly	2	4
Reading	1	0
Comprehension	5	0
Transformation	0	2
Process skills	3	3
Encoding	0	0

Question: Jackie had 138 marbles. She played with Cathy and Kylie and won 65 more marbles. How many marbles does she have now?	
Reading: 1. Please read the question to me. If you don't know a word, leave it out.	The student correctly read the question
Comprehension: 2. Tell me, what is the question asking you to do?	"Figure out how much marbles she has. She had 138 she won 65 and they're asking me how much she had left."
Transformation: 3.Tell me how you are going to find the answer.	"Takeaway some stuff."
Process skills: 4. Show me how you get your answer, and "talk aloud" as you do it, so that I can understand how you are thinking.	(Student wrote horizontally 138-65, then put 8 up on his fingers then quickly answered "70" verbally.)
Encoding: 5. Now write down your actual answer.	(Student wrote 70)

Fig. 8 Example of a Year 3 students comprehension error

Knowing that comprehension was the critical barrier for many students in the cohort, the Year 3/4 teacher was then able to target worded problem comprehension throughout the program. This was observed in the teacher implementing Polya's problem-solving heuristics (Polya, 1988) with students (i.e. See, Plan, Do, Check). This method was effective in raising students' proficiency in solving worded problems as evidenced by the elimination of comprehension errors. The reduction in comprehension errors concurs with findings from the worded questions on the diagnostic test, which indicated a reduction in errors associated with selecting incorrect values or operations. Therefore, it was observable that students' *strategic competence* had increased over the course of the program.

Furthermore, it was importantly found that the prevalence of correct answers increased throughout the program. Instances of correct answers indicated that students were able to solve the problem correctly during the interview that they were previously unable to correctly answer during the diagnostic testing. One of the two students who was able to answer the worded question correctly during the Newman interview at the beginning of the program, Greg, would not attempt the written problem on the diagnostic test. During the interview, Greg was able to correctly answer the question verbally without working out on paper, suggesting a well-developed understanding of addition. For Greg, this was a consistent finding. By the end of the program Greg still did not attempt a similar worded problems on the diagnostic test, despite his responses to the addition computation questions indicating proficiency with addition (see Fig. 9). During the post-program Newman interviews, Greg was again able to answer it quickly in his head explaining his thinking verbally during the interview. An example of Greg's interview responses is outlined in Fig. 10.

It was a significant finding that one third of the sample was able to answer the question correctly during the interview despite it being previously answered incorrectly or not attempted on the diagnostic test. During interviews, it was found that

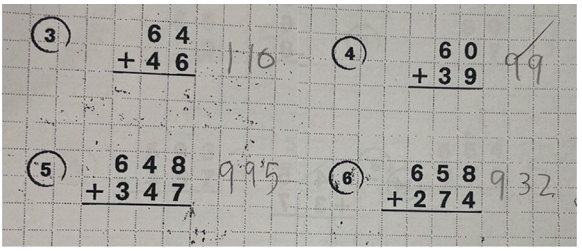


Fig. 9 Greg's (Year 4) correct responses to addition computation questions at the end of the program

students were now attempting worded questions that they previously would not attempt in the written test. In fact, all students attempted all worded questions presented in the Newman interviews. Students had an increased *productive disposition* towards the identical task in an interview setting. This finding is particularly significant as now the classroom teacher had substantial insight into the exact specific problem-solving hurdles for students. This was not possible to know from the diagnostic tests alone due to the high prevalence of students not attempting the questions. Students mathematical thinking that was previously invisible was now made visible via the Newman interview. This is an important finding that has implications for approaches to examining the mathematical thinking of young Indigenous students', and young students in general. In particular, it highlights the clear mismatch between the proficiency students demonstrate on written tests and their actual proficiency.

Discussion

The aim of this paper was to present a discussion of the methodological approaches that made visible the mathematical thinking of young Indigenous students. The combination of a standardised achievement test, diagnostic test, and Newman interview was able to provide a complete and accurate picture of young Indigenous students' mathematical proficiency in relation to their *conceptual understanding*,

Question: Jackie had 138 marbles. She played with Cathy and Kylie and won 65 more marbles. How many marbles does she have now.	
Reading: 1. Please read the question to me. If you don't know a word, leave it out.	The student correctly read the question
Comprehension: 2. Tell me, what is the question asking you to do?	It's plusing.
Transformation: 3.Tell me how you are going to find the answer.	It's plusing.
Process skills: 4. Show me how you get your answer, and "talk aloud" as you do it, so that I can understand how you are thinking.	"203. I was thinking 60 and 30 is 90. So it's 90 and 5 is 95, and 8 more and another hundred is 203"
Encoding: 5. Now write down your actual answer.	(Student wrote 203)

Fig. 10 Greg's (Year 4) correct response to the question during the Newman interview

strategic competence, and *procedural fluency* with addition. These data sources also allowed for inferences to be made on the presence of students' *productive dispositions* towards the tasks. In particular, the analysis of errors on diagnostic tests in conjunction with Newman interviews enabled the categorisation of common student difficulties under the observable strands of mathematical proficiency as defined by Kilpatrick et al. (2001) as outlined in Fig. 11. It was notable that conceptual understanding and strategic competence were more intertwined, contrasting to literature that focuses on the complex and intertwined nature of conceptual understanding and procedural fluency (e.g. Carpenter, 1986; Haapasalo & Kadijevich, 2000; Rittle-Johnson et al., 2015).

A standardised test provided a comparable pre- and post-program measure of students' achievement however students' proficiency, in relation to the five strands, was not discernible. For the PAT-M, findings indicated that majority of students experienced improvements in mathematics achievement. However, PAT-M findings did not provide conclusive data to facilitate a deep understanding of the specific concepts or skills cohorts were proficient in, nor where areas of improvement lied. As supported by Ginsburg (2016), achievement tests served a legitimate purpose in their ability to provide a norm referenced measure of student achievement; however, their ability to provide fruitful information that enables a teacher to understand and subsequently capitalise on a student's mathematical proficiency was limited. Wager et al., (2015, p. 15) suggested that "the temptation to rely on standardised assessment practices may result in misguided understanding about what children actually know about

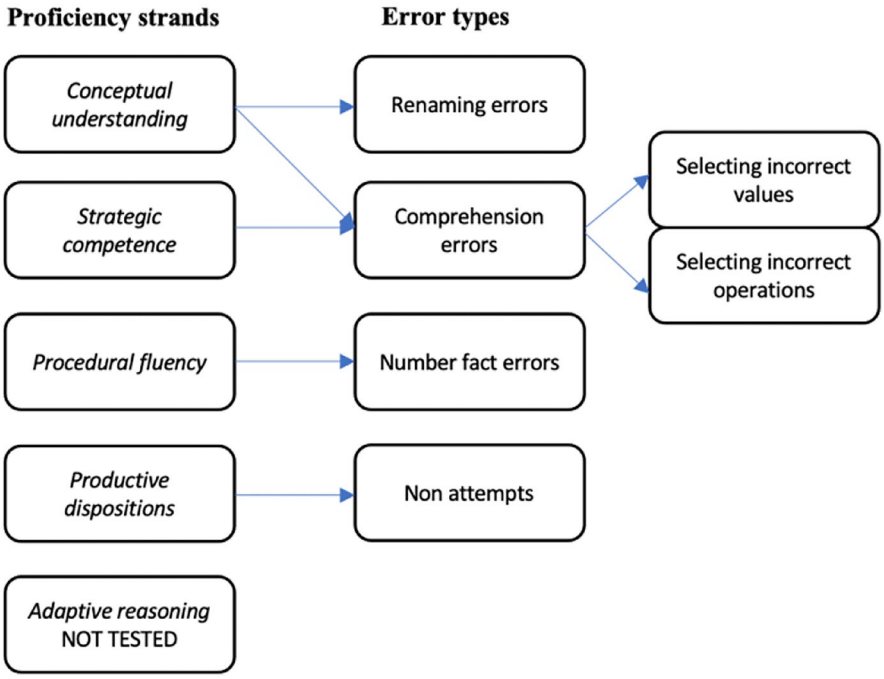


Fig. 11 Classification of error types using the strands of mathematical proficiency

mathematics", a sentiment also reiterated by Schwab (2012), and this was indicated in the findings of this study. This was seen when further analysis of diagnostic tests found that large gains in achievement on PAT-M were not always indicative of students being proficient with core mathematical concepts. Whilst there was a substantial reduction in the gap in achievement comparative to the PAT-M norming sample, deeper analysis of findings from diagnostic tests suggested several critical mathematical concepts in which students were still developing proficiency.

Diagnostic tests allowed for the analysis of student thinking by revealing the strengths of students' mathematical proficiency in relation to their conceptual understanding, procedural fluency, strategic competence, and productive dispositions. Complimenting a diagnostic test with follow-up Newman interviews was insightful as Newman interviews revealed students' thinking in cases of non-attempts on the diagnostic tests. Furthermore, where the root causes of a student's challenges were speculative on the diagnostic test, Newman interviews were able to further reveal and confirm the underlying causes. It is also important to note that diagnostic tests and Newman interviews, in addition to providing critical data on students' mathematical proficiency, also provided providing tangible and actionable direction for the classroom teacher. The ability of error analysis of diagnostic tests to feed forward into teaching has also been noted by Miller (2015) and Papic (2015). For example, in this study Newman interviews demonstrated that students required further support in relation to their strategic competence when solving worded problems, and this was able to be addressed during the program and subsequent gains in proficiency were observed. A written test was not able to reveal this finding as non-attempts were prevalent for this sample of students. For the diagnostic tests, assessing whether students were able to give a contextual story for sums was particularly useful as an assessment tool. These stories indicated that students' conceptual understanding of the initial addition concept required further attention and development. Concurring findings also indicated that diagnostic tests and Newman interviews were complimentary data sources. That is, findings from Newman interviews supported those from the diagnostic tests, with similar trends identified relating to students' conceptual challenges.

The case has been made in this study that the combination of standardised assessment, diagnostic assessment, and interviews were useful in capturing young Indigenous students' mathematical proficiency; however, there was the need to make appropriate adjustments to the administration of diagnostic assessments. Making appropriate adjustments to assessment, such as changing assessment administration to small group delivery with open time limits was important for learners in this context. Imposing the time limit and formal test conditions for the diagnostic test would not have provided an accurate picture of students' proficiency. Modifying questions to allow for students to express their understanding orally was also fruitful in understanding students' thinking.

A limitation of the use of these three assessment tools was that neither the PAT-M, suite of diagnostic tests, nor Newman interviews revealed students' proficiency in relation to *adaptive reasoning*, nor did they specifically test for students' *productive dispositions* beyond behavioural engagement. This is important to note in relation to future directions for methodological approaches in this area. It is suggested that Newman interviews may be extended to include further probing questions

asking students to justify and reflect on their responses during the encoding phase, and to ask them to justify their chosen strategies during the transformation stage. An additional assessment tool that specifically measures students' productive dispositions towards mathematics would also be useful to obtain a complete picture of students' mathematical proficiency.

Conclusion and implications

Young Indigenous students are capable learners of mathematics. Overall, the findings indicated that drawing on methodological approaches that give space to young Indigenous students' voices to make visible their mathematical thinking is important and valid in revealing their proficiency in mathematics. English (2016) purported that alternative assessment "exposes, rather than conceals, young children's mathematical talents" (p. 2), and this was evidenced in the findings of this study. A standardised assessment alone was not sufficient to examine young Indigenous students' mathematical thinking accurately and meaningfully. This has important implications for wider policy and practice, particularly associated with national and international standardised assessment of Indigenous students' mathematical achievement, which is often relied upon for policy and funding purposes. It also raises questions pertaining to the validity of standardised assessment for measuring the mathematical proficiency of any young student in mathematics, though their useful role in providing comparable measures of achievement over time must be noted. Given the discrepancy between what was found regarding students' proficiency on a standardised tests compared to a diagnostic test or interview, whether standardised tests should be drawn upon to primarily drive reform and pedagogical recommendations is clearly questionable. As suggested by English (2016) and Klenowski (2009, 2014), alternative assessments were able to give more accurate insight into the mathematical thinking of Indigenous students. Serow et al. (2016) surmised that "children deserve the right to have mathematics assessment tasks that allow them to demonstrate "what they know", engage them, and consider their interests in an environment that does not cause stress" (p. 251). Diagnostic tests and Newman interviews are one way of making visible the mathematical thinking of young Indigenous students by centring students' voices. Analysis of student errors on diagnostic tests is useful in understanding young Indigenous students mathematical thinking, and it enables the teacher to appropriately tailor future teaching. However, this study revealed that important adjustments were needed in the administration of written tests. Overall, the combination of these data sources is useful for gaining accurate insight detailed insight into young Indigenous students' mathematical proficiency.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Ethics approval Ethics approval gained from the Griffith University Human Research Ethics Committee (GU Ref No: 2017/264).

Informed consent Informed consent was gained from all student participants in line with the ethical approval.

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Archer, A. L., & Hughes, C. A. (2011). Exploring the foundations of explicit instruction. *Explicit instruction: Effective and efficient teaching*, 1–22.

Ashlock, R. B. (1976). *Error patterns in computation* (2d ed.). Merrill.

Aubrey, C., Dahl, S., & Godfrey, R. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal, 18*(1), 27–46.

Australian Curriculum, Assessment and Reporting Authority. (2022). *Mathematics curriculum*. Retrieved August 2, 2022, from Australian Curriculum: http://www.australiancurriculum.edu.au/mathematics/curriculum/f-10?layout=1

Aveling, N. (2013). 'Don't talk about what you don't know': On (not) conducting research with/in Indigenous contexts. *Critical Studies in Education, 54*(2), 203–214.

Black, P., & Wiliam, D. (2005). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan, 80*(2), 139–148.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*(3), 608–645.

Booker, G. (2011). *Building numeracy: Moving from diagnosis to intervention*. Oxford University Press.

Booker, G., Bond, D., Sparrow, L., & Swan, P. (2014). Teaching primary mathematics. Pearson Higher Education AU.

Botha, L. (2011). Mixing methods as a process towards indigenous methodologies. *International Journal of Social Research Methodology, 14*(4), 313–325.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*(2), 155–192.

Carpenter, T. (1986). Conceptual knowledge as a foundation for procedural knowledge: Implications for research on the initial learning of arithmetic. In J. Hiebert (Ed.), *Concepual and Procedural Knowledge: The Case of Mathematics* (pp. 129–148). Hillsdale, NJ., London, Lawrence Erlbaum.

Carstairs, J., Myers, B., Shores, E., & Fogarty, G. (2006). Influence of language background on tests of cognitive abilities: Australian data. *Australian Psychologist, 41*(1), 48–54.

Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Creswell, J. W., & Plano Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Denzin, N. K., Lincoln, Y. S., & Smith, L. T. (Eds.). (2008). *Handbook of critical and indigenous methodologies*. Sage.

Doig, B., McCrae, B., & Rowe, K. (2003). *A good start to numeracy: effective numeracy strategies from research and practice in early childhood*. Melbourne: Australian Council for Educational Research.

Eckermann, A.-K., Dowd, T., Chong, E., Nixon, L., Gray, R., & Johnson, S. (2006). *Binan Goonj: Bridging cultures in Aboriginal health*. Churchill Livingstone Elsevier.

Elwood, J. (2008). Gender issues in testing and assessment. In P. Murphy & K. Hall (Eds.), *Learning and practice, agency and identities* (pp. 87–101). Sage.

English, L. (2016). Revealing and capitalising on young children's mathematical potential. *ZDM - International Journal on Mathematics Education, 48*(7), 1079–1087.

Ginsburg, H. P. (2016). Helping early childhood educators to understand and assess young children's mathematical minds. *ZDM, 48*(7), 941–946.

Grootenboer, P., & Sullivan, P. (2013). Remote indigenous students' understanding of measurement. *International Journal of Science and Mathematics Education, 11*(1), 169–189.

Guba, E. G., & Lincoln, Y. S. (1985). *Naturalistic inquiry* (Vol. 75). Sage.

Haapasalo, L., & Kadijevich, D. (2000). Two types of mathematical knowledge and their relation. *Journal Für Mathematik-Didaktik, 21*(2), 139–157.

Hambleton, R. K., & Rodgers, J. (1995). *Item bias review*. ERIC Clearinghouse on Assessment and Evaluation, the Catholic University of America, Department of Education.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

He, J., and van de Vijver, F. (2012). Bias and Equivalence in Cross-Cultural Research. *Online Readings Psychol. Cult.* 2 (2). https://doi.org/10.9707/2307-0919.1111

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert, *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–23). Hillsdale, NJ: Lawrence Erbaum Associates.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14*(3), 251–283.

Howard, P., & Perry, B. (2011). Aboriginal children as powerful mathematicians. In *Teaching and learning in Aboriginal education* (pp. 130–145). Oxford University Press.

Jorgensen, R. (2018). *Final report: Remote numeracy project.* University of Canberra.

Kanu, Y. (2007). Increasing school success among Aboriginal students: Culturally responsive curriculum or macrostructural variables affecting schooling? *Diaspora, Indigenous, and Minority Education, 1*(1), 21–41.

Kilian, A., Fellows, T. K., Giroux, R., Pennington, J., Kuper, A., Whitehead, C. R., & Richardson, L. (2019). Exploring the approaches of non-Indigenous researchers to Indigenous research: A qualitative study. *Canadian Medical Association Open Access Journal, 7*(3), E504.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics* (pp. 115–135). National research council (Ed.). Washington, DC: National Academy Press.

Klenowski, V. (2009). Australian Indigenous students: Addressing equity issues in assessment. *Teaching Education, 20*(1), 77–93.

Klenowski, V. (2014). Towards fairer assessment. *The Australian Educational Researcher, 41*(4), 445–470.

Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*(2), 265–299.

Leder, G. C., & Forgasz, H. J. (2006). Affect and mathematics education: PME perspectives. In *Handbook of research on the psychology of mathematics education* (pp. 403–427). Brill Sense.

Lee, T. S. (2015). The Significance of Self-Determination in Socially, Culturally, and Linguistically Responsive (SCLR) Education in Indigenous Contexts. *Journal of American Indian Education, 54*(1), 10–32.

Ma, X. (1997). Reciprocal relationships between attitude toward mathematics and achievement in mathematics. *The Journal of Educational Research, 90*(4), 221–229.

Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education, 110*(3), 256–280.

Martin, K. (2002). Ways of knowing, being and doing: A theoretical framework and methods for indigenous and indigenist research. The Australian Public Intellectual Network. Retrieved from http://www.api-network.com/main/index.php?apply=scholars&webpage

Matthews, S., Howard, P., & Perry, B. (2003). Working together to enhance Australian Aboriginal students' mathematics learning. *Mathematics education research: Innovation, networking opportunity*, 9–28.

Miller, J. (2015). Young Indigenous students' engagement with growing pattern tasks: A semiotic perspective. In M. Marshman, V. Geiger, & A. Bennison (Eds.), *Proceedings of the 38th annual conference of the mathematics education research group of Australasia* (pp. 421–428). Sunshine Coast, QLD: MERGA.

Miller, J., & Armour, D. (2021). Supporting successful outcomes in mathematics for Aboriginal and Torres Strait Islander students: A systematic review. *Asia-Pacific Journal of Teacher Education, 49*(1), 61–77.

Mellor, S., & Corrigan, M. (2004). *The Cast for Change: A review of contemporary research on Indigenous education outcomes.* Camberwell, Victoria: Australian Council for Education Research.

Morris, C., & Matthews, C. (2011). *Numeracy, Mathematics and Indigenous Learners: Not the same old thing*. [Paper presentation]. 2011 - Indigenous Education: Pathways to Success. https://research.acer.edu.au/research_conference/RC2011/8august/6

National Health and Medical Research Council. (2018). *Ethical conduct in research with Aboriginal and Torres Strait Isladner People and communities*. Retrieved Nov 3, 2021, from https://www.nhmrc.gov.au/about-us/resources/ethical-conduct-research-aboriginal-and-torres-strait-islander-peoples-and-communities

Nelson-Barber, S., & Trumbull, E. (2007). Making assessment practices valid for Indigenous American students. *Journal of American Indian Education*, 132–147.

Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematical tasks. *Victorian Institute for Educational Research Bulletin, 39*, 31–43.

Newman, M. A. (1983). *Strategies for diagnosis and remediation*. Harcourt, Brace Jovanovich.

Nichol, R., & Robinson, J. (2000). Pedagogical challenges in making mathematics relevant for Indigenous Australians. *International Journal of Mathematical Education in Science and Technology, 31*(4), 495–504.

Niss, M. (2007). Reflections on the state of and trends in research on mathematics teaching and learning. In F. K. J. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1293–1312). Information Age Publishing.

Özerk, K., & Whitehead, D. (2012). The Impact of National Standards Assessment in New Zealand, and National Testing Protocols in Norway on Indigenous Schooling. *International Electronic Journal of Elementary Education*. 4(3), 545–561. Available at: https://eric.ed.gov/?id EJ1070447.

Papic, M. (2015). An early mathematical patterning assessment: Identifying young Australian indigenous children's patterning skills. *Mathematics Education Research Journal, 27*(4), 519–534.

Pegg, J., & Graham, L. (2013). A three-level intervention pedagogy to enhance the academic achievement of Indigenous students: Evidence from QuickSmart. In *Pedagogies to enhance learning for Indigenous students* (pp. 123–138). Springer Singapore.

Perso, T. (2009). Cracking the NAPLAN code: Numeracy in action. *Australian Mathematics Teacher, the, 65*(4), 11–16.

Philipp, R. A., & Siegfried, J. M. (2015). Studying productive disposition: The early development of a construct. *Journal of Mathematics Teacher Education, 18*(5), 489–499.

Philpott, D. (2006). Identifying the Learning Needs of Innu Students: Creating a Model of Culturally Appropriate Assessment. *Canadian Journal of Native Studies*, 26(2), 361–381. Available at: http://www3.brandonu.ca/cjns/26.2/07philpott.pdf.

Polya, G. (1988). *How To Solve It*. Princeton University Press.

Preston, J. P., & Claypool, T. R. (2021). Analyzing assessment practices for Indigenous students. In *Frontiers in Education* (Vol. 6).

Reid O'Connor, B. (2020). Exploring a Primary mathematics initiative in an Indigenous community school. Doctoral dissertation, Griffith University.

Reid O'Connor, B. (2021). A primary education mathematics initiative in an Indigenous community school. In Leong, Y. H., Kaur, B., Choy, B. H., Yeo, J. B. W., & Chin, S. L. (Eds.), Excellence in mathematics education: Foundations and pathways (Proceedings of the 43rd annual conference of the Mathematics Education Research Group of Australasia), pp. 321–328. Singapore: MERGA.

Reid O'Connor, B., & Norton, S. (2020). Supporting Indigenous primary students' success in problem-solving: Learning from Newman interviews. *Mathematics Education Research Journal*. https://doi.org/10.1007/s13394-020-00345-8

Rosenshine, B. (2012). Principles of Instruction: Research-based strategies that all teachers should know. *American Educator, Spring*, 12–39.

Rigney, L., Garrett, R., Curry, M., & MacGill, B. (2020). Culturally Responsive Pedagogy and Mathematics Through Creative and Body-Based Learning: Urban Aboriginal Schooling. *Education and Urban Society, 52*(8), 1159–1180. https://doi.org/10.1177/0013124519896861

Riley, L. (2021). Community-Led Research through an Aboriginal lens. In V. Rawlings, J. Flexner & L. Riley (Eds.), *Community-Led Research: Walking new pathways together*. Sydney: Sydney University Press.

Rittle-Johnson, B., Schneider, M., & Star, J. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Education Psychology Review, 27*, 287–297.

Rogers, A. (2014). *Investigating whole number place value in Years 3–6: creating an evidence-based developmental progression* (Doctoral dissertation, RMIT University).

Sarra, C. (2011). Transforming Indigenous education. In N. Purdie, G. Milgate, & H. Bell, *Two way teaching and learning: Toward culturally reflective and relevant education* (pp. 107–118). Victoria: ACER Press.

Sarra, G., Alexander, K., Carter, M., & Cooper, T. (2016). QUT YuMi deadly maths: Making a difference in mathematics learning for Indigenous and non-Indigenous students. In *Aboriginal and Torres Strait Islander Mathematics Alliance Conference*, 2016–10–30 - 2016–11–02. (Unpublished)

Sarra, G., & Ewing, B. (2014). Indigenous students transitioning to school: Responses to pre-foundational mathematics. *Springerplus, 3*, 685. https://doi.org/10.1186/2193-1801-3-685

Sarra, G., & Ewing, B. (2021). Culturally responsive pedagogies and perspectives in mathematics. In M. Shay & R. Oliver (Eds.), *Indigenous Education in Australia* (pp. 148–161). Routledge.

Sarra, C., Spillman, D., Jackson, C., Davis, J., & Bray, J. (2018). High-expectations relationships: A foundation for enacting high expectations in all Australian schools. *The Australian Journal of Indigenous Education*. https://doi.org/10.1017/jie.2018.10

Schwab, R. G. (2012). Indigenous early school leavers: Failure, risk and high-stakes testing. *Australian Aboriginal Studies, 1*, 3–18.

Serow, P., Callingham, R., & Tout, D. (2016). Assessment of mathematics learning: What are we doing?. In *Research in Mathematics Education in Australasia 2012–2015* (pp. 235–254). Springer, Singapore.

Skille, E. Å. (2021). Doing research into Indigenous issues being non-Indigenous. *Qualitative Research*, 14687941211005947.

Solano-Flores, G., Backhoff, E., Contreras-Niño, L. A., & Vázquez-Muñoz, M. (2015). Language shift and the inclusion of indigenous populations in large-scale assessment programs. *International Journal of Testing, 15*(2), 136–152.

Stephanou, A., & Lindsey, J. (2013). *PATMaths: Progressive Achievement Testes in Mathematics* (4th ed.). ACER.

Stronger Smarter. (2014). *Leadership Program*. Retrieved March 29, 2016, from Queensland Stronger Smarter Institute: http://strongersmarter.com.au/leadership

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., & Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*. Springer Nature.

Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research*. Sage.

Thomson, S., Wernert, N., Rodrigues, S., & O'Grady, E. (2020). *TIMSS 2019 Australia Volume 1: Student performance*. Australian Council for Educational Research

Thornton, S. (2020). (Re) asserting a knowledge-building agenda in school mathematics. *Mathematics Education Research Journal*, 1–17.

Tripcony, P. (2002). Challenges and tensions in implementing current directions in Indigenous education. Paper presented at Australian Association for Research in Education Conference (AARE), Brisbane, Queensland, 2002.

Trumbull, E., & Nelson-Barber, S. (2019). The Ongoing Quest for Culturally-Responsive Assessment for Indigenous Students in the US. In *Frontiers in Education*. https://doi.org/10.3389/feduc.2019.00040

van Tuijl, C., Leseman, P. P., & Rispens, J. (2001). Efficacy of an intensive home-based educational intervention programme for 4- to 6-year-old ethnic minority children in the Netherlands. *International Journal of Behavioral Development, 2*, 148–159.

Woodward, A., Beswick, K., & Oates, G. (2017). The four proficiency strands plus one?: Productive disposition and the Australian Curriculum: Mathematics. In *2017 Mathematical Association of Victoria Annual Conference (MAV17)* (pp. 18–24).

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wager, A. M., Graue, M. E., & Harrigan, K. (2015). Swimming upstream in a torrent of assessment. In B. Perry, A. MacDonald, & Gervasoni, A. (Eds.), *Mathematics and Transition to School: International Perspectives* (pp. 17–34). Dordrecht: Springer.

Warren, E., & Miller, J. (2013). Young Australian Indigenous students' effective engagement in mathematics: The role of language, patterns, and structure. *Mathematics Education Research Journal, 25*(1), 151–171.

White, A. (2005). Active Mathematics In Classrooms: Find out Why Children Make Mistakes - And Then Doing Something To Help Them. *Square One, 15*(4), 15–19.

Warren, E., Young, J., & deVries, E. (2008). The impact of early numeracy engagement on 4 year old Indigenous students. *Australian Journal of Early Childhood Education, 33*(4), 2–8.

Wigfield, A., Eccles, J.S., Simpkins, S., Roeser, R.W., & Schiefele, U. (2015). Development of Achievement Motivation and Engagement. In R.M. Lerner (Eds.), *Handbook of Child Psychology and Developmental Science*, https://doi.org/10.1002/9781118963418.childpsy316

Zevenbergen, R. (2004). Mathematics, social class and linguistic captial: An analysis of mathematics classroom interactions. In B. Allen & S. Johnston-Wilder (Eds.), *Mathematics Education: Exploring the culture of learning* (pp. 119–133). RoutledgeFalmer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.