Assessing conceptual learning in mathematics classrooms

A thesis completed in partial fulfilment of the requirements for the degree of Master of Education at Massey University

By Megan Kanz

February 2020

Abstract

This study explores how building a collective understanding of the big mathematical ideas and learning trajectory within an area of the mathematics curriculum, positions teachers to make consistent, accurate and effective judgements of student's learning. Teachers and schools are shifting their focus from teaching and learning regimes that prioritize procedural mastery towards those that prioritise building conceptual understandings. This is resulting in a growing mismatch between what is taught and what is tested, since existing testing regimes primarily seek to assess procedural skill over conceptual learning. Schools must therefore rely heavily on teachers to judge students developing conceptualisations, until such time that assessment procedures better align with the outcomes sought by the education system (Jones & Inglis, 2015). Additionally, it looks at what factors support teachers to make judgments of student's conceptual understandings.

The paradigm of interpretivism and social constructivism underpins the focus of this research. Relevant literature is drawn on to support the claims made in relation to hypothetical learning trajectories and their positive impacts on teacher knowledge, practice and judgement. The research evidence that supports using free-response tasks is presented and justifies their use for assessing the breadth and depth of student conceptions. Comparative judgement as a tool for assessing free-response tasks is utilised with consistent and reliable results.

The interventions utilised by this design study involved carefully planned, collaborative professional development around the Curriculum Elaborations. Teachers collectively mapped hypothetical learning trajectories, planned appropriate, levelled tasks, assessed student learning through free-response tasks and participated in a comparative judging session for each curriculum area covered. Significant growth was seen in teacher knowledge about the curriculum content and learning progressions. Teachers knew what content to cover, in what order to present it so that it *made sense* and, how learning outcomes planned into the HLT subsequently related to the mathematics curriculum levels. This understanding positioned teachers to made consistent and accurate judgements about their students learning for both teaching and assessment purposes.

The research findings provide insight into the ways teachers can be supported to notice and judge student's conceptual learning through engaging with collaborative professional

development aimed at building their collective knowledge of the curriculum content and progression of learning.

Acknowledgements

Whāia te iti kahurangi kit e tuohu koe me maunga teitei.

Pursue that which is precious and don't be deterred by anything less than a lofty mountain.

This has been one of the most challenging journeys of my life. It is entirely thanks to the support of many that I've managed to complete my masters and this thesis. Wolfgang, you are my rock, thank you for your enduring belief in me, I could never have kept going without you. My girls, Rebecca and Shannon, thank you for the lollies, hugs, smiles and especially for the space. Thank you to my parents and the West Auckland colleagues who became my extended whānau. Fina, Cherie and Maggie, thank you for always listening and most of all for your encouragement. Thank you to my new colleagues in Gisborne for entertaining my pursuits and so willingly giving of your time and energy, without you this study would not have happened.

Bobbie, you have been part of my journey from the very beginning. You are an unwavering inspiration and I feel privileged to have had your support over the years but especially during this last year as you have helped me push through. Thank you for your time, patience, critical input and understanding.

Jodie when I started down this road with you in West Auckland in 2016 the 'end' seemed very far-off. Thank you for giving me the confidence and direction I needed along the way. The robust conversations, constructive feedback and passion have motivated and encouraged me.

To the many beautifully unique tamariki I've met along the way. You fill my days with challenge, passion and lots of smiles. Aroha nui!

Table of Contents

List of Tables

Chapter One: Introduction to the study

1.1 Background and rationale

Both in New Zealand and internationally there has been a shift of focus towards learning for conceptual understanding (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), this means that teachers and schools need reliable means of judging and assessing developing mathematical understanding. Historically, formal assessment practices have prioritized assessing procedural knowledge, schools rely increasingly on teacher observations and judgements of student learning for a broader/more holistic view of achievement and developing conceptual understandings. Teacher knowledge about how students' learning progresses and connections between big mathematical ideas, within various areas of mathematics can be either a scaffold or barrier for student learning. Land and Drake (2014) advocate the benefits of using a trajectory lens when planning learning experiences to build robust, conceptual understanding. Whether referred to as learning progressions or hypothetical learning trajectories, when teachers are able to conceptualise the key understandings that need to be built and how they connect, they are better positioned to make accurate judgements and formative assessment decisions about students' learning. These judgements are however shown to be subject to bias (Smaill, 2013), affected by various contextual influences (Meissel et al., 2017), often inconsistent and variable not only between schools but also between classes within schools. For these reasons, it is important to examine how teachers can be scaffolded to make accurate and consistent judgements of student conceptual learning for both assessment and learning purposes.

1.2 Research objectives

The purpose of this study is to investigate how building a collective understanding of the big mathematical ideas and learning trajectory within an area of the mathematics curriculum, enables teachers to make consistent, accurate and effective judgements of student's learning. The use of free-response tasks and a comparative judgement (CJ) assessment tool are also explored as a way to expose student's conceptual understanding and further support teacher judgements based on these understandings. The findings of this study further extend the knowledge base reflected in the literature review by offering a contextual exemplar of factors that support New Zealand primary teachers to judge and assess students' conceptual learning.

The following research questions are addressed:

1. How does building a collective understanding of the big mathematical ideas and learning trajectory enable teachers to make judgements on student's learning?
2. What factors support teachers to assess student's conceptual understanding?

1.3 Thesis structure

The thesis begins with a review of literature relevant to the objectives of this research. Chapter three presents the research questions and a justification for the design research approach used. Data gathering methods and analysis techniques are detailed and a description of the sample, context and schedule given. Research rigor and ethical considerations are also discussed. Chapter four presents the findings, beginning with the themes that emerged from the baseline data, moving on to those that transpired following each intervention. Chapter five discusses these findings and answers the research questions posed. The final chapter (Chapter six) summarises the research, and takes note of implications and next steps.

Chapter Two: Literature Review

2.1 Overview

This chapter provides an overview of the literature that relates to the aims of the research. The objective is to situate the study within the current context by showing how the questions posed relate to existing research within the field and by highlighting the gaps it addresses. Section 2.2 describes common assessment practices in New Zealand and introduces the notion of assessing conceptual rather than procedural learning. Section 2.3 then gives a description of conceptual knowledge and includes detail on what mathematics teaching and practices promote conceptual understanding. Section 2.4 then discusses using hypothetical learning trajectories and learning trajectory-based instruction, along with a problem-solving approach, to build conceptual understanding. It also highlights subsequent positive impacts on teacher knowledge and practice. Section 2.5 focusses in on the New Zealand context and introduces the New Zealand Curriculum Elaborations. Section 2.6 moves on to address the complexities of teacher judgement. It details how judgements are made, draws attention to issues such as reliability, teacher bias and ways of mitigating these biases. It ends by discussing how shared frameworks can be used to build consistency and reliability. Section 2.7 introduces the concept of Comparative Judgement as an alternative means of assessing. The final two sections address the areas of mathematics covered by the current study, shows how these areas are generally assessed, and offers a rationale for this to change.

2.2 Assessment in New Zealand

Historically, assessments both internationally and within New Zealand have focused on students' ability to calculate and follow procedures rather than their conceptual learning. The nature of procedural fluency – learning reflected by students ability to follow learned processes and procedures – means how it "is measured has become relatively standardized: participants solve a set of problems, and a score is calculated based on how many correct answers they obtain or based on the specific procedures they use to arrive at those answers" (Crooks & Alibali, 2014, p. 345). This in turn allows for accurate and efficient marking on mass where papers can be reliably scored by following regular mark schemes or rubrics.

Within New Zealand primary schools, mathematics is commonly assessed with the following assessments: Global Strategy Stage Assessment (GLoSS) (Ministry of Education, 2013), Junior Assessment of Mathematics (JAM) (Ministry of Education, 2014) and Progressive Achievement Tests (PATs) (The New Zealand Council for Educational Research, 2019). JAM is designed to assess students working within levels 1 and 2 of the New Zealand Curriculum. The assessment is broken up into modules that aim to assess number strategies, number knowledge and the geometry, algebra and measurement strands (Ministry of Education, 2014). Both JAM and GLoSS are administered as a 1:1 interview with students where the assessor follows a clear "script" that prompts what to ask at each stage, how to proceed and subsequently "assess" the student (Ministry of Education, 2013, 2014). While JAM seeks to assess mathematical knowledge across separate domains and strategy use, GLoSS seeks only to assess student's use of number strategies. GLoSS may be administered to students working from level 1 to 5 of the New Zealand Curriculum (Ministry of Education, 2013). In both instances, student's mental strategies are assessed rather than their written thinking. PATs are standardised, computer generated, multiple-choice item tests designed for students in years 3-10. They claim high reliability where scores can be used to track progress, compared within school and with national groups across year levels (The New Zealand Council for Educational Research, 2019)

Despite the shift in focus towards promoting conceptual learning, these assessment practices remain rooted in assessing the procedural with none of these assessments offering space for students to show their conceptual understanding. This is partially because multiple-choice standardised tests "really only test knowledge recall" (Berube, 2004, p. 264) and, "risk promoting a narrow and arguably distorted view of students' mathematical thinking" (NCETM, 2009 as cited in Hunter & Jones, 2018, p. 400). These ideas are further backed by Jones and Karadeniz (2016) who explain how testing the "recall and application of facts and algorithms … privilege[es] procedural knowledge" over conceptual knowledge (p. 1). As noted by Jones and Inglis (2015) the challenge then is for examination to better align with the outcomes sought by our education system. Or put another way, "assessment processes should match the objectives of curricula … and as such assessments should capture conceptual understanding" (Jones & Karadeniz, 2016, pp. 6-7). Where an assessment accurately assesses what it sets out to, it is said to have high construct validity.

The question then is how to assess conceptual knowledge – what do students need to do to display their conceptual understanding within a subject. One way in which conceptual understanding can be displayed is through the use of open-ended tasks, also called free-response tasks. Sullivan et al. (2006) concluded that these tasks offer invaluable insight into students' thinking by allowing freedom to explore a range of ideas, providing opportunities to extend thinking and draw generalisations. They also noted that open tasks are more accessible than closed tasks since students can approach them in their own ways. These open tasks are regularly used with writing assessments where students are not only assessed on specifics like grammar and spelling but also prompted to produce a piece of writing in response to a short prompt (Hunter & Jones, 2018). Assessment of mathematics learning through the use of free-response tasks is not yet common practice.

Until very recently in New Zealand, schools were required to report to the Ministry of Education on students' achievement according to the National Standards for reading, writing and mathematics. Achievement was determined by teachers making an OTJ (Overall Teacher Judgement) of students' learning in relation to the New Zealand Curriculum (Ministry of Education, 2019). These OTJs generally drew on a range of evidence including teacher's anecdotal observations and more formal summative assessments. However, Bonne (2017) explains that there was doubt about whether National Standard's data provided "a reliable picture of student performance, either within one school or across all local schools" (p. 18). Furthermore, both principals and teachers perceived National Standards to represent only a "narrow slice of what students know and can do, rather than their overall performance". As with any high-stakes type testing there is the inclination to teach to the test which in turn allows "test content to define curriculum" (Abrams and Madaus, 2003 as cited in Berube, 2004, p. 266). With the end of National Standards and associated reporting, educators in New Zealand have the opportunity to explore alternative means of assessing student learning and, better align the assessment process with the curriculum.

This section has described assessment practices within New Zealand and introduced the notion of assessing conceptual rather than procedural understandings. Next, a description of conceptual knowledge is given with reference to the mathematics teaching that promotes conceptual understanding.

2.3 Procedural versus conceptual knowledge

Procedural knowledge has been defined as "knowledge of sequences of steps or actions that can be used to solve problems" (Rittle-Johnson & Seigler, 1998, as cited in Crooks & Alibali, 2014, p. 345). In contrast, conceptual understanding involves the comprehension of mathematical concepts, operations and relationships (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2019) or as "mental connections among mathematical facts, procedures and ideas" (Hiebert & Grouws, 2007, p. 380). Crooks and Alibali (2014) unpack how conceptual knowledge proves useful in multiple ways. Likewise, the National Governors Association Center for Best Practices and Council of Chief State School Officers (2010) state "students who lack understanding of a topic may rely on procedures too heavily… In short, a lack of understanding effectively prevents a student from engaging in the mathematical practices" (p. 8). With the general consensus that having conceptual knowledge confers benefits above and beyond those associated with having procedural skills (Crooks & Alibali, 2014) the move towards teaching for conceptual understanding is clearly justified.

Mathematics teaching that holds building conceptual understanding as a valued outcome, requires students be suitably challenged. Research literature (Henningsen & Stein, 1997; Hunter, 2008; Simon & Tzur, 2004; Smith & Stein, 1998) highlights the important role that challenging, rich mathematical tasks play in setting students up to engage cognitively, reason deeply and build conceptual understanding. These validations have prompted educators to foster a sustained problem-solving approach to teaching and learning.

The next section looks at using hypothetical learning trajectories and learning trajectory-based instruction to support problem based learning and the development of conceptual understanding in students and highlights the subsequent positive impacts reported on teacher practice.

2.4 Hypothetical learning trajectories and learning trajectory-based instruction

Before the instructional tasks can be selected, teachers need to know where the learning should be heading. In other words they must consider a hypothetical learning trajectory (HLT). The current study draws on the definition of an HLT proposed by Simon and Tzur (2004). The HLT starts from students existing mathematical conceptions and includes conjectures about how learning will progress towards the big mathematical ideas (Small, 2010) or goals to be developed. Included are purposefully planned mathematical tasks that will promote the learning at each step. These tasks also need to be "fine-tuned" as teaching and learning progresses to "develop the level of thinking that a particular student needs" (Clements, 2011, p. 369). Sztajn et al. (2012) define Learning Trajectory Based Instruction (LTBI) as a way of teaching that bases instructional decisions on students' use of learning trajectories.

Teacher content knowledge is critical for these processes to be executed well. Where teacher's content knowledge and their teaching practices are grounded around the big mathematical ideas they are positioned to effectively select and sequence instructional tasks that progress students thinking while explicitly making connections with these ideas (Carmel, 2005). Sarama et al. (2017) reported on how teachers within their study on Learning Trajectories-Based Professional Development and Learning Trajectories-Based Instruction (LTBI), became familiar with the changing levels of mathematical thinking pre-schoolers displayed along developmental progressions, through the use of planned instructional activities and end goals.

Holt Wilson (2014) conducted a design study that tracked changes in a group of teachers' practice through engaging in professional development that built an understanding of learning trajectories to inform planning, instruction and assessment. He found that initially teachers concept knowledge lacked precise models of thinking which meant their descriptions of student's thinking were vague and contained "general or irrelevant observations" (p. 234). After engaging in structured professional development experiences around learning trajectories, teachers' ability to notice and anticipate students' mathematical learning improved. This supported teachers to move students thinking along a "continuum of conceptual development" (p. 237). In addition, this learning gave teachers specific language to use when discussing students' mathematical conceptions with colleagues.

This section addressed HLTs generally and the impact of LTBI on teacher practise. The next section looks more closely at the New Zealand context by introducing the New Zealand Curriculum Elaborations.

2.5 The New Zealand Curriculum Elaborations

The New Zealand Curriculum Elaborations are a set of documents that unpack how learning progresses in the various mathematical domains, through levels 1-8 of the New Zealand Curriculum. For instance, under Level One of patterns and relationships, the Learning Objective (LO) *create and continue sequential patterns* is unpacked into the following progressions: reproduce; continue with justification; then invent and communicate the 'rule' to others. As a student's learning moves into level Two, this knowledge progresses to being able to use the repeating element to make predictions at given ordinal positions. This builds on to using additive strategies to find further terms and using numeric tables of values. When teachers know the specifics of these progressions or where to reference them, they are able to logically sequence learning experiences, notice students conceptions, misconceptions or gaps and, to make judgements of these conceptions accordingly.

Moving on from learning trajectories, the next section addresses the complexities of teacher judgement. It begins by describing the process and, then looks at issues related to reliability, teacher bias and ways of mitigating such biases. Finally, it describes in detail the use of HLTs or shared frameworks to build consistency and reliability.

2.6 Teacher judgement

Teacher judgement as a form of assessment has been an ongoing part of a teacher's daily work in mathematics. Teachers make formative (and in some instances summative) assessments of how students are tracking through day-to-day noticings and anecdotal observations of their learning. This construct of noticing is defined by Jacobs et al. (2010) as an interrelated set of skills that enable teachers' to attend to children's strategies, interpret their understandings and make decisions about how to respond in relation to these understandings. These noticings inform the judgements teachers make and subsequently impact on decisions about ongoing instruction including "instructional pace, level of support, and level of task difficulty" (Alvidrez & Weinstein, 1999; Clark & Peterson, 1986; Hoge & Coldarci, 1989 as cited in Meissel et al., 2017, p. 48). Teachers are able to make judgements about not only how well students are following set procedures, but also on their use of mathematical practices holistically and their conceptual learning. Campbell (2014) explains that "this approach to assessment, with its reliance on an understanding of each child build over time rather than based simply on a one-off performance in a set test, has several arguable advantages" (Campbell, 2014, p. 517).

There are however reports that highlight the inconsistencies between judgements and standardised tests. The inconsistencies are largely due to the overall subjective nature of individuals and subsequently assessments that result from teacher judgements are less reliable that those that result from standardised testing programs (Heldsinger & Humphry, 2010). Assessments that rely on teachers making subjective judgements on the quality of student responses are reported to have low reliability (Adie et al., 2012). Smaill (2013) report that judgements are subject to either positive or negative bias and, can be a mix of "students' attainment of intended learning outcomes with judgements of students' effort, work habits and other 'academic enabler' traits" (Brookhart, 2013, p. 84). Blank et al. (2016) as cited in Meissel et al. (2017, p. 58), explain bias as "deep cognitive and emotional responses that people have". On one hand, these biases may be related to student characteristics such as "gender, ethnicity, socioeconomic background, and students' special needs or ESOL status" (Meissel et al., 2017, p. 50). On the other hand, they may be influenced by classroom or school contexts and varying interpretive approaches.

Meissel et al. (2017) compared research conducted on teacher judgement with that done on teacher expectations. They explain that although these notions are fundamentally different, with teacher expectations a predictive measure of future achievement versus teacher judgement, a measure of current achievement, that teacher judgement is likely subject to the same biases. As biases are found to negatively impact on student learning opportunities and life chances, these researchers claim that "the ramifications of any bias in these judgements are particularly serious" (p. 49).

One way in which teacher judgements can be made more valid as well as reliable is by having teachers appraise "against some background, or reference framework" and, by making an "explicit response" (Sadler, 1998, p. 80) according to a pre-determined criteria for instance. Furthermore, there must be a consistent "recognition of performance" that corresponds with a point on the said framework (Adie et al., 2012, p. 224). The idea of using a framework to reference student learning aligns with the idea of learning progressions where a learning progression addresses individual concepts and increasingly more sophisticated ways of thinking about that concept (Land & Drake, 2014). This thinking aligns with that of LTBI detailed earlier within this chapter. In this way, teachers are better equipped to judge how well students learning is progressing by referencing students work back to the HLT with which they are working.

Land and Drake (2014) examined ways that expert teachers made use of curriculum and research-based progressions to inform their instructional designs. They noticed that teachers were supported in the development and use of progressions when big mathematical ideas were provided in a concise, unpacked and bulleted manner. Additionally, reference to the order in which these ideas should be explored was detailed i.e. the trajectory of conceptual understandings clearly described.

Sztajn et al. (2012) attend to four "highly used frameworks" (p. 147) of instructional design when considering LTBI namely mathematical knowledge for teaching (MKT), task analysis, discourse facilitation, and formative assessment. The MKT framework referenced draws on the model proposed by Ball et al. (2008). Their six categories under MKT are used by Sztajn et al. (2012) as the basis to show how each knowledge category informs a different aspects of LTBI. Furthermore, they propose that by unifying these four frameworks, LTBI can offer a theory of teaching that is "organised around and grounded in research on student learning" (p. 147).

Teachers need to have shared understandings of HLTs and LTBI in order to effectively notice, and to support their subsequent judgements. Where teachers are involved in collaboratively unpacking curriculum they construct collective understandings of what is expected of children to achieve at each curriculum level (Education Review Office, 2018). These shared understandings of syllabus and standards build reliability and consistency as they are independent of "the individual teacher, student, location or time"(NSW Department of Education, 2019).

This section identified the many factors that affect teacher judgement. It discussed influences, biases and ways judgement decisions can be supported. Finally, the links were made to LTBI and use of HLTs to build consistency and reliability. Next, comparative judging is described and it's use as an alternative means of assessment discussed.

2.7 Comparative Judgement

Comparative judgement (CJ) is a means of assessing based on the notion that people are better at making direct comparisons between two objects, rather than making a comparison against specified criteria (Thurstone, 1927). In CJ, student responses are presented in pairs with the group of judges individually required to select the "better" response each time. This process establishes a measurement scale where responses are assigned percentile scores and ranked accordingly (Jones et al., 2015).

Subsequently, CJ assists with combating the multitude of factors, described in the previous section, that affect teacher judgement. In addition, the notion that current testing regimes are unable to effectively measure conceptual understanding was raised in Section 2.1 and, the alternative of using an open or free-response task to capture this understanding, proposed. In summary, free-response tasks are specifically designed to allow for a wide range of responses that cannot always be anticipated, giving students opportunity to display their conceptualisations in ways that are meaningful to them. This feature however is what makes assessment of these tasks, using traditional rubric or mark scheme methods, difficult (Jones et al., 2015).

Jones and Inglis (2015) explain a key difference when looking at how traditional mark schemes differ from CJ. "Mark schemes attempt to capture the construct of interest using explicit, precise and detailed assessment criteria. CJ instead relies on the collective understanding of the construct by a relevant community of experts" (p. 341). So in terms of being able to measure conceptual understanding in mathematics, the collective understanding of this construct will inform how judgements are cast. This notion is supported by Hunter and Jones (2018) who, in their study using Free-Response tasks and a CJ assessment method, found that they were able to gain consistent insights into primary-aged students mathematical thinking. For these reasons, CJ will be used as the assessment tool within the current study.

Next the two areas of mathematics covered by the current study are detailed with attention brought to how these areas are currently assessed. The importance of supporting students to build conceptual understanding in these areas is stated and, subsequently the need for assessing these understandings established.

2.8 Algebra

Algebra has an underlying role to play in all areas of mathematics, making it a linchpin to future mathematics success (NCTM, 2000; National Mathematics Advisory Panel, 2008 and RAND Mathematics Study Panel, 2003 as cited in Knuth et al., 2016). Kaput (2000) describes algebra as the gateway to future educational and employment prospects, especially since it is often a requirement for graduation. These claims have supported the integration of early algebra into the primary curriculum with the understanding that "early algebra education can potentially eliminate some of the difficulties students have with algebra in the secondary grades" (Knuth et al., 2016, p. 68). Algebra however is an area that teachers commonly struggle to teach and in which students commonly underperform (Brown & Quinn, 2007).

Mathematics learning must build from conceptual understanding before skill proficiency (Hiebert & Grouws, 2007). This is particularly true for algebra where students have been shown to believe algebra is about memorizing disconnected rules and procedures (Kieran, 1992 as cited in Woodbury, 2000). It is critical that teachers and curriculum "help students build internal representations of procedures that become part of larger conceptual networks before encouraging the repeated practice of procedures" (Hiebert & Grouws, 2007, p. 830). With the increasing pressure for teachers to provide opportunities for students to build conceptual understandings in algebra, so to grows the need to assess these understandings. Currently, algebra is assessed through a few multiple-choice questions in PATs (The New Zealand Council for Educational Research, 2019) and through module 9 in the JAM (Ministry of Education, 2014). These assessments offer marginal insight into student's conceptual understandings in algebra. As explained by Varygiannes (2013/2014) if we intend to assess student understanding of a concept it is not about asking more or closed questions that call on traditional algorithms or formulas but, rather it is about asking questions that are open-ended enough to enable learners to demonstrate greater understanding.

2.9 Fractions

A comprehensive understanding of fractions and rational numbers is critical since, along with algebra, is a predictor of future success with mathematics (Brown & Quinn, 2007). Furthermore, fractions, decimals and percentages are used by 68% with blue-collar jobs (Handel 2016, as cited in Braithwaite et al., 2018). The challenge to make sense of mathematics is very real for both teachers and students. Lamon (2007) reports that fractions, rations, and proportions, of all topics in the school curriculum, are potentially the most cognitively challenging, mathematically complex, "most protracted in terms of development" (p. 629), most difficult to teach, and yet most critical in terms of success in mathematics and science at higher levels.

Teachers have been shown to make a significant impact on student learning (Hattie, 2008) where teacher knowledge and pedagogy contributes significantly to students performance. Given the complex nature of this area of mathematics, it is not surprising that teachers often grapple with not only having limited MCK of fractions (both conceptual and procedural), they also struggle with limited pedagogical content knowledge. This means they are left unable to interpret student's misconceptions, and have insufficient knowledge of appropriate "instructional representations and strategies" (Depaepe et al., 2015, pp. 84-85). Subsequently, an improved teacher conceptual understanding of fractions would positively improve student outcomes. For students to develop concrete conceptual understanding of fractions they must be taught without being confined to applying “preconstructed rules” since “creating a dependence on rules in early learning experiences can inhibit students’ ability to construct meaning, and mask many of the underlying interconnections important to conceptual understanding” (Skemp, 2006 as cited in Anderson & Pritchard, 2010, p. 52).

2.10 Summary

This review of literature has set the scene and provided a rationale for the current study. It began by reviewing current assessment practices in New Zealand and highlighted the gaps that exist within this regime. Next, the distinction between procedural and conceptual learning was provided and a problem-based learning approach, that promotes conceptual understanding introduced. The relevant theories on the use of HLTs and LTBI was then discussed and their positive impact on students’ conceptual learning, as well as teacher practice included. Next, this discussion was made relevant to the New Zealand context by focussing in on the New Zealand Curriculum Elaborations. Teacher judgement and the complexities surrounding this notion was discussed next, paving the way for a description of CJ. Finally the two areas of mathematics addressed within the study were described and reference made to how conceptual learning is or, is not assessed under each area. The following chapter details the methodological design of the current study.

Chapter Three: Methodology

3.1 Introduction

This chapter details the methodological design used in the study. Section 3.2 re-states the research questions. Section 3.3 addresses the epistemological, ontological and methodological considerations that underpin the research and advocates for the qualitative approach taken. Section 3.4 outlines the more specifically the design research approach chosen. Section 3.5 addresses the role played by the researcher. Section 3.6 covers how data was gathered through interviews, professional development and planning activities, and assessment responses. Section 3.7 covers data analysis, interpretation and grounded theory. Section 3.8 details the sample, context and study schedule with reference to the various phases of the study. Section 3.9 explores research rigor covering aspects of trustworthiness, reliability and validity. Section 3.10 covers ethical considerations and finally 3.11 summarises the chapter.

3.2 Research questions

To meet the aims of this study, the following research questions were posed:

3. How does building a collective understanding of the big mathematical ideas and learning trajectory enable teachers to make judgements on student's learning?
4. What factors support teachers to assess student's conceptual understanding?

3.3 Epistemological, Ontological and Methodological Considerations

In order to understand the methodological approach utilised in this study, it is helpful to briefly describe the research paradigm underpinning design research and the "world view" held by the researcher since, "every worldview within which the researcher becomes immersed holds the key to knowing" (Bishop, 2005, p. 124).

Broadly speaking, paradigms are ways of thinking that encapsulate how researchers and research approaches make certain assumptions about how the learning will take place, as well as what will be learnt through the study (Creswell, 2002). These assumptions can be about "what is knowledge (ontology), how we know it (epistemology), what values go into it (axology), how we write about it (rhetoric), and the process of studying it (methodology)" (Creswell, 2002, p. 6). Methodological justifications must therefore consider all of these elements when making research claims (Yanow & Schwartz-Shea, 2015). The researcher acknowledges that the current study aligns predominantly with interpretivism and the ideas of social constructivism.

The DMIC teaching and learning model enacted by the teachers in this study, reflects constructivist theory – where understandings are socially constructed by the learning community. As described by Kretchmar (2013) "knowledge is constructed by individuals through their experience, and is not necessarily representative of 'the real world'" (p. 1). This notion is further developed by Wilson (2008) who describes reality as "made up of socially constructed concepts that are shared" (p. 37). These notions reflect the ontological and epistemological views held by the researcher and influence how she sees herself "positioned" within the research. Her world-view mirrors social constructionism, believing that our concepts and world-views are continually being re-constructed, believing that "there can be as many different views of the world as there are people viewing it" (Lincoln and Gubba, 2000 as cited in Humphrey, 2013, p. 8).

A qualitative research methodology is advocated by researchers wishing to gather descriptive data of various kinds, such as people's individual words, actions or behaviours (Taylor et al., 2016). This study therefore sits under the "umbrella" (Punch & Oancea, 2014, p. 144) of qualitative research. In addition, in the current study the researcher draws on an interpretivist view, starting from Weber's notion of *verstehen* – "understanding on a personal level the motives and beliefs behind peoples actions" (Taylor et al., 2016, p. 15) within a socially constructed reality. Merriam (1998) explains how this view assumes that meaning is embedded within people's unique experiences, where the researcher attempts to unpack how they come together to build in-depth understandings. This key understanding – "that meaning is socially constructed by individuals interacting in their world" justifies the qualitative research methodology advocated within this study (Merriam & Grenier, 2019, p. 3).

3.4 Design research

Design research is more of an "approach" to research or "methodological framework" rather than a particular strategy, or seen as a "genre of flexibly using existing research approaches" to gain design-based insights (Bakker, 2018, p. 7). Design research in education embraces the constructivist outlook, relying heavily on the collaborative efforts of researchers and teachers where teaching and learning is planned and analysed within a context (Hathaway & Norton, 2018). Design researchers consider these constructed contexts when formulating hypothesis about teaching and learning. As succinctly put by The Design-Based Research Collective (2003), "practitioners and researchers work together to produce meaningful change in contexts of practice" (p. 6).

Design research is important in helping to understand how, when and why innovative educational practices work. (The Design-Based Research Collective, 2003). More specifically, design research aims to develop, test, implement and diffuse these innovative practices in order to shift teaching and learning along a scale from malfunction towards excellence. (Kelly, 2003 as cited in Kelly et al., 2008).

As with an inquiry process, design research is iterative in nature. That is, "development and research take place through continuous cycles of design, enactment, analysis and redesign"(The Design-Based Research Collective, 2003, p. 5) as was the case within the current study. These interactions helped to refine the understandings that emerged, and connect the defined "processes of enactment to outcomes of interest" (p. 5). A further characteristic of design-research is that research should "lead to sharable theories that help communicate relevant implications to practitioners and other educational designers" (p. 5). These ideas align with the aims of this study and justify utilising a design research approach.

3.5 Researcher Role

The researcher begins by firstly aligning research questions and design so that data collection and analysis supports the research aims. In qualitative studies, the researcher is seen as the "primary instrument for gathering and analysing data" (Merriam, 1998, p. 20) so that as full an understanding of the case as possible may be made (Punch & Oancea, 2014). Furthermore the researcher is responsible for gathering information about what made any changes happen (Kelly et al., 2008).

Although on study leave during the intervention phase of the current study, the researcher is employed as a teacher within the school where the research took place. She had had some advisory capacity within the school, but was primarily seen as a classroom teacher. The established "insider role" needed to be reframed as the researcher repositioned herself outside the classroom in a new role as mentor/facilitator. This required the building of rapport, especially with those teachers in the year 3 and 4 learning neighbourhoods with whom she had not worked previously. The researcher also needed to critically reflect on potential power relations and build an "atmosphere of trust" by being empathetic, listening intently and by being a good communicator (Merriam, 1998, p. 23).

As described in section 3.3, in the current study the researcher acknowledges that her world-view, perspectives and own subjectivities position her unconsciously within the research. Rather than claiming objectivity, she is aware that these positions may infer personal biases since "all observations are filtered through the researcher's selective lenses (Taylor et al., 2016, p. 184). These factors thus influence how the she responds to, interprets and frames the research. Neuman (2014) describes the need for careful and sensitive interpretation at each stage of the research. Merriam (1998) further describes this notion of "sensitivity" as being "highly intuitive", explaining that the researcher must be "sensitive to the context and all the variables within it" (p. 21). Furthermore, sensitivity includes how the researcher shapes her biases and subjectivities that influence the investigation and its findings (Merriam, 1998).

3.6 Data gathering methods

Data collection during the current study was iterative, reflective of the design research approach. The initial design was conjectured then, through the "iterative design process featuring cycles of invention and revision" more specialized conjectures were "framed and tested" (Cobb et al., 2003, p. 10) thus maintaining reflexivity throughout the project.

As is characteristic with design experiments, multiple sources of data were collected in order to gain an in-depth and holistic understanding of "a complex, interacting system" or "*learning ecology"* (Cobb et al., 2003, p. 9). Data was collected in response to "both learning and the means by which that learning was generated and supported" (p. 12). This included facilitator led, collaborative professional development and planning sessions, assessment responses, interview responses, and artefacts. As outlined by Mathison (1988) this multifaceted process supports triangulation by providing a "rich and complex picture" (p. 15) of the case and in turn enhances internal validity (Merriam, 1998). This section concludes with a schedule for the research outlining the phases it followed.

3.6.1 Interviews

Interviews were used as the main form of data gathering in the current study with the intention of gathering "descriptive data in the subjects' own words" so that the researcher may "develop insights on how subjects interpret" what is being studied (Bogdan & Biklen, 2006, p. 103). Thirteen teachers and two senior management members participated in the planning and professional development phase of this study. Eleven of the teachers agreed to being interviewed before, during and after the project.

Initial teacher interviews were done individually and exploratory in nature to gather in-depth information about current practices and understandings. Pre-set questions were used but a semi-structured interview process was followed which allowed the researcher to prompt for clarification or extend the questioning and "adapt to particular respondents and situations," (Punch & Oancea, 2014, p. 184) gaining more detailed insights where necessary.

Intermediate interviews and post interviews were conducted in small focus-type groups using a semi-structured approach and pre-set questions and one teacher was interviewed individually. The change from solely individual interviews was made in response to participant feedback after the first session. Furthermore some of the questions that required more forethought were shared with participants ahead of the interview. This put participants more at ease, and not 'on the spot'. It also helped deepen individual's responses and "stimulate … making explicit their views, perceptions, motives and reasons" (Punch & Oancea, 2014, p. 186). To avoid individuals keeping quiet about their individual experiences or being "too embarrassed to share them in a group" (Bogdan & Biklen, 2006, p. 109), these groups were kept small (2 to 4 in a group) and included individuals who chose to be interviewed together, generally teachers from within the same learning neighbourhood.

3.6.2 Professional development and collaborative planning activity intervention

Data was collected through an interview process, prior to the professional development (PD) and planning activity intervention, which partially informed the focus of the sessions. During these sessions, teachers worked collectively to identify learning outcomes (LOs) and big ideas associated with the two focus areas in order to plan a hypothetical learning trajectory (HLT) for each area that covered Levels 1 to 3 of the Curriculum. The collaborative approach was intentionally utilised to support teachers in constructing shared understandings across the year levels. The New Zealand Curriculum Elaborations were used to inform this process. Next, with the support of exemplars, the teachers prepared the learning activities and tasks that would be presented to students. During the second PD and planning session and in response to analysis of the first session, teachers also anticipated solution strategies for problems planned. Resultant documentation was collected and discussions audio recorded for analysis. The first round of analysis informed changes and modifications to the second round of PD and planning, as per the nature of design research.

3.6.3 Assessment responses

As part of the study, open or free response assessment tasks (See Appendices A, B, C and D) were used to assess student learning and evaluate the level of conceptual understanding gained. These written, photographed or documented responses represented artefacts used to support and complement the research findings.

3.6.4 Schedule

The study consisted of 8 phases conducted over a 15 week period, commencing in week 3 of the first term and concluding in week 6 of the second term. A summary timeline of data collection is detailed in table 3.1

Table 3.1: Summary timeline of data collection

Study weeks	School term/ week	Action taken
Week 1	T1W3	Phase 1: Baseline teacher interviews/PD planning
week 2	T1W4	Phase 2: Patterns and relationships PD & collaborative planning session
Week 3	T1W5	Teaching and learning of patterns and relationships
Week 4	T1W6	
Week 5	T1W7	
Week 6	T1W8	
Week 7	T1W9	Phase 3: Testing of Patterns and relationships
Week 8	T1W10	Phase 3: Comparative Judging done
Week 9	T1W11	Phase 4: Intermediate interviews
Week 10	T2W1	Phase 5: Fractions PD & collaborative planning session
Week 11	T2W2	Teaching and learning of fractions
Week 12	T2W3	
Week 13	T2W4	
Week 14	T2W5	Phase 6: Testing of fractions
Week 15	T2W6	Phase 7: Comparative judging of fractions and, Phase 8: Final interviews

Phase 1

This phase included all baseline data gathering. In addition to exploring previous whole school data and discussions with school management and the DMIC team, it was decided that the focus areas would be the patterns and relationships part of algebra, and fractions.

The 11 teachers who agreed to be full participants in the study were individually interviewed. Interview questions (Appendix E) were pre-planned and designed to capture a baseline of understanding as follows:

- Current assessment practices used
- How teacher's judge student understanding

- Against what these judgements are made
- Teacher knowledge of how students learning progresses through levels 1-3 of the curriculum in each area

The interviews were audio recorded and wholly transcribed. Transcriptions were analysed by the researcher using grounded theory to identify the spread of practice and understanding amongst teachers. This phase informed the 2nd phase.

Phase 2

This phase involved the researcher preparing and then conducting an afternoon of professional development and collaborative planning on the patterns and relationships strand of algebra, using the New Zealand Curriculum Elaborations. During this, the Curriculum Elaborations were unpacked and used as a basis to co-construct a long-term plan that identified learning outcomes along a HLT. Learning neighbourhoods then planned instructional activities in a shared document that evolved during the teaching and learning that followed.

Phase 3

After a 5 week teaching and learning period, students were given one of two open assessment tasks on patterns and relationships (see Appendix A and B). The researcher then uploaded these response sheets into a CJ computer program. All participants were then involved in a comparative judging session that ranked the response sheets.

Phase 4

During phase 4 teachers were interviewed (see Appendix F) to explore any wonderings about, and responses to the 2nd and 3rd phases of the study. Interviews were audio recorded and wholly transcribed. Transcriptions were analysed by the researcher using grounded theory to identify themes that would inform phase 5 of the study.

Phase 5

The 5th phase involved a similar process of PD and collaborative planning at the start of the second term. This time about fractions, and with some modifications reflective of the outcomes of phase 4. Teaching and learning ran for four weeks.

Phase 6

This phase involved students answering one or two open assessment tasks on Fractions (see Appendix C and D). Again, the researcher uploaded these response sheets into a CJ computer program.

Phase 7

During this phase participants were again involved in a comparative judging session that ranked student response sheets.

Phase 8

This final phase involved the researcher interviewing teachers in focus groups of varying size (see Appendix G for interview questions), to explore further wonderings and responses, as well as suggestions about future implications.

3.7 Data analysis and interpretation

As described by Bogdan and Biklen (2006), "[a]nalysis involves working with the data, organizing them, breaking them into manageable units, coding them, synthesising them and searching for patterns" (p. 159). Through a "dynamic and creative" process of analysis, the researcher attempted to "gain a deeper understanding" of what was being studied, continually refining her interpretations (Taylor et al., 2016, p. 160).

3.7.1 Grounded theory

Grounded theory finds its origins in the work of Glasser and Strauss (1967). It is a method of data analysis that maximises on the potential for "discovering theories, concepts, hypotheses, and propositions" that emerge directly from the data (Taylor et al., 2016, p. 156). Grounded theorists strive to thoroughly understand people's experiences by meticulously attending to the detail (Ryan & Bernard, 2000). A grounded theory approach to data works well with a design research methodology in that grounded theory too is iterative in nature. The researcher seeks to become more "grounded" in the data, developing "increasingly richer concepts and models of how the phenomenon being studied really works" (Ryan & Bernard, 2000, p. 783). Thus, a key component is that data collection and analysis happens simultaneously so that theoretical ideas that begin emerging may be continuously refined or investigated (Charmaz, 2006). Charmaz (2014) says "grounded theory coding generates the bones of your analysis" from which connections are built and meaning is constructed. She goes on to explain how as the coding becomes more focussed it "shapes an analytic frame from which you build the analysis" (p. 113). The researcher is therefore able to advance theory development throughout the process of data collection and analysis, all the while remaining as "faithful to the data" as possible (Taylor et al., 2016, p. 159). Krueger and Casey (2000) even go so far as to claim that doing analysis as you go enhances data collection by alerting the researcher to gaps, missed opportunities or questions that could be changed to extract more detailed responses.

A grounded theory approach to analysing the data was applied and *Nvivo* coding was used to develop insights and build theoretical understandings from the data. Working the data in this way allowed for emergent themes, concepts and propositions to be developed. For example, baseline data revealed three themes based around teacher knowledge. These included knowledge about: curriculum content and learning progressions; curriculum content and delivery; and seeking and using resources to support planning and assessment decisions. The changes to these emergent understandings, following each intervention, are explored within the findings of the study.

3.8 Sample and context

This section details the participants and the context of the study.

3.8.1 Sample

The current study was set at one New Zealand primary school in the regional city of Gisborne. Taiawhiti School (pseudonym) students come from a mix of socio-economic backgrounds and the school is reflective of the bicultural nature of the region.

The study involved all Year 1 to 4 teachers, 13 in total and 2 members of the senior management team, not in classrooms. All 15 participants were part of the PD and planning sessions, and the assessment sessions. Eleven of the 13 teachers were interviewed three times over the course of the study.

3.8.2 Context

Developing Mathematical Inquiry Communities (DMIC) approach to mathematics teaching and learning was described in the introduction chapter. This approach is currently being implemented by teachers in Year 1-4 at Tairawhiti School with the support of DMIC mentors through an ongoing PD program. As part of this program, a communication and participation framework (CPF) has been adopted as a structure that teachers use to support students development of mathematical practices. Teachers are required to notice student's conceptual mathematical thinking and adoption of mathematical practices build, change and be expressed in new and different ways. Consequently, these areas are the ones teachers wish to assess. This CPF however, offers subjective and anecdotal-type opportunities for teacher noticings rather than formal, reliable and objective ones that can be accurately measured, monitored and reported on. As noted in Section 2.1 of the Literature Review, the current testing regime does not accurately assess students' use of mathematics but rather it largely assesses student's procedural learning. Hence the need to establish a reliable and objective means of assessing students conceptual understanding.

The current study looks at how teachers can be scaffolded to notice, make judgements on, and assess student's conceptual learning. It used the New Zealand Curriculum Elaborations as a framework to support long term planning by identifying learning outcomes and mapping these with teachers along a hypothetical learning trajectory (HLT). Tasks were planned and anticipated as part of this collaborative process. Open assessment tasks, a CJ assessment tool, and the planned HLT informed assessment judgements so that formative assessment could be made through a trajectory lens.

As described by Cobb et al. (2003) it is important to "distinguish in the specification of the design between elements that are the target of the investigation and those that may be ancillary, accidental or assumed as background conditions" (p. 10). This study assumes that within all classrooms the ambitious mathematics integrated within some elements of complex instruction advocated by DMIC is the method of instruction. The aspects of building pedagogical practices for orchestrating discourse, referred to by Land and Drake (2014) is assumed as already in place within classrooms. DMIC also focusses on building students' conceptual knowledge and use of mathematical practices above procedural knowledge in an effort to teach mathematics in a meaningful, authentic and connected way and, to build deeper, long-lasting understanding. Strong conceptual understanding means that students are more likely able to "apply their mathematical knowledge to problem solving in varied and unfamiliar contexts" (Jones & Inglis, 2015; Jones et al., 2015, p. 151). This current study is not looking at student/teacher interactions within the classroom, lesson structure, mathematical talk, social norms or any other classroom based pedagogies, all of which are being considered through DMIC within the school.

Previous whole school data highlighted two areas of mathematics where students were underperforming namely patterns and relationships in algebra, and fractions. Interestingly enough, there seems to be a global consensus that these areas are particularly troublesome for students and teachers.

The design research approach has been justified as most suitable to this study within the preceding sections of this chapter. It considered, through an iterative process, the planning and supports needed to scaffold teachers to make judgements consistently and to consider the HTL students are on. Various tools including professional development (PD) on use of the Curriculum Elaborations, collaborative planning documents, open problems and a CJ testing approach were utilised. As is common with design research, a detailed description of the context, setting and participants is given so that the range of data gathering methods is detailed with a description of the data analysis process.

3.9 Research rigor

Trustworthiness is a term engineered by Lincoln and Guba (1985) to describe the way the qualitative researcher ensures the overall quality of the research. Other concepts, including the notions of validity and reliability, relate generally to the trustworthiness of the research findings and the extent to which they rest upon the data (Merriam, 1998). The multitude of approaches described in the following paragraphs collectively add to the trustworthiness of the findings drawn from the data.

3.9.1 Reliability and validity

Internal validity, truth value, and credibility are all terms used to describe whether the study measures what it aims to measure, or how well the study captures "what is really there" (Merriam, 1998, p. 201). As stated by Miles et al. (2014), "do the findings of the study make sense?" (p. 312). Lincoln and Guba (1985) describe credibility as internal validity resulting from the researcher's ability to represent "multiple realities revealed by informants as adequately as possible" (p. 215).

In the current study, data was gathered in an iterative manner, over a period of time, from multiple participants to formulate deep and rich understandings and, in turn build robust and authentic credibility. Furthermore, in a process similar to member checks, the researcher checked back in with participants during the intermediate and post interview phases and utilised their reflective feedback to ensure that emerging understandings were plausible.

External validity and applicability refer to the ability to generalise results from a study to the wider population and are associated with quantitative research (Krefting, 1991). Generalisation is not the goal of this qualitative study. Instead, the researcher strives for dependability – built by being concerned with "accuracy and comprehensiveness" of the data (Bogdan & Biklen, 2006, p. 40). They go further to explain that reliability in qualitative research is the fit between what is being recorded as data and what is actually happening in the setting under study. In order to maintain dependability the researcher's role and position have been explicitly identified, as well as the paradigms and analytic constructs underpinning the research (Miles et al., 2014).

A rich description of the study's context is provided with the understanding that the portrayed reality is "holistic, multidimensional, and ever-changing" (Merriam, 1998, p. 205). This information and the use of multiple data collection methods triangulated the research, strengthening reliability and internal validity (Mathison, 1988). A comprehensive audit trail has been maintained with data collection and analysis described in detail. Subsequently, those wishing to reference this research may make informed decisions about the potential for transferability within their own unique contexts.

3.10 Ethical considerations

This study has been designed to adhere to the Massey University Code of Ethical Conduct of Research, Teaching and Evaluations involving Human Participants (Massey University, 2015) as well as the REVISED CODE (Massey University, 2017). The purpose of these codes is to "provide protection for all participants in research, and certain teaching and evaluation programmes, as well as to protect researchers and institutions" (Massey University, 2015, p. 3). The Code is underpinned by key principles that cover the areas of respect for persons, privacy and confidentiality, minimisation of risk of harm, consent, avoidance of unnecessary deception, social and cultural sensitivity, and justice. As per requirements, a low risk ethics application was made and approval gained prior to the commencement of data collection. All participants involved with the study were given appropriate information from which to give their informed consent. Permission was also gained from the school principal and the board of trustees (see Appendix I).

Ethical considerations particularly relevant to this study were anticipated under two broad headings. Firstly, those that emerged from the researcher's position as a teacher on study leave and secondly, those that were related to the setting.

As eluded to earlier within section 3.5, the researcher needed to re-position herself from a classroom teacher to a researcher with a role more reflective of a facilitator and mentor. It was critical for the researcher to build a rapport of trust and openness with participants in an attempt to anticipate and manage any ethical dilemmas that surfaced. Throughout the project, participant's perspectives were heard and valued, minimising any risk of harm, and allowing for robust but safe discussions. PD and planning meetings were held at the time that staff hui (meeting) would normally proceed, avoiding the pressure of additional time constrains. Interviews were scheduled at times and places that worked for individual participants and only those that fully consented were interviewed. The researcher made herself available in-between meeting and interview times to answer questions and generally give support as needed.

Secondly, it was anticipated that ethical dilemmas would arise associated with contextual aspects of the study – related to its setting, pre-existing planning/assessment processes and individual participant's perspectives. As stated by Smith (2012, pp. 35-36) "research is not an innocent or distant academic exercise but an activity that has something at stake and that occurs in a set of political and social conditions". The support of senior management, an openness to the research and upholding a culture of trust, prevented dilemmas evolving. Furthermore, pseudonyms were used, and individually identifiable data avoided so that confidentiality and anonymity was upheld.

3.11 Summary

This chapter began by re-stating the research questions and describing the underlying epistemological, ontological and methodological considerations that influenced the research. It justified the qualitative, design research approach adopted, which facilitated the iterative interventions of the study and this led on to a description of the role played by the researcher. All data gathering methods were discussed next and details of the research schedule provided. The grounded theory approach used was described within the data analysis and interpretation section and examples of the themes generated by this approach provided. Specific detail was included to describe the sample and context in an effort to distinguish what elements were being targeted by this research. Issues of reliability and validity were addressed and finally the ethical considerations, relevant to this study were discussed.

Chapter Four: Findings

4.1 Overview

Section 4.2 describes the beginning point of the current study. It identified that all teachers, from the outset, used their judgement decisions as well as traditional assessment practices to assess students learning, yet did not feel well equipped to make these judgements. After looking more closely at their responses in relation to both algebra and fractions, sections 4.3 to 4.5 describe the three key themes that emerged from this data regarding teacher curriculum knowledge. Section 4.6 describes the changes that occurred following intervention one. 4.6.1 and 2 cover the use of comparative judgement and it's results, 4.6.3 describes the growth of teacher knowledge attributed to using the HLT, 4.6.4 describes the power of the collective and impact of collaboration and 4.6.5 looks as how using free-response, open assessment tasks enabled teachers to assess students developing conceptualisations. Section 4.7 details intervention 2, offering a discussion on the comparative judging process and results. References are made to reliability thereof, various technical considerations and time constraints. Section 4.8 includes additional implications for teacher knowledge and confidence.

4.2 Baseline data and beginning point

The teachers made assessment judgements of students' learning using a mix of traditional, formal assessments and anecdotal teacher observations at the initial stage of the research. They reported that noticing, observing, responding to, and judging were part of the fundamental practices they engaged in on a daily basis. All teachers (n=11) described observing students in some shape or form and, that these observations informed the judgements they made about student learning, teaching steps, and assessment decisions.

Further questioning also probed how well equipped the teachers considered they were to make judgements of students' mathematical reasoning. This produced mixed and in some instances, contradicting results. While all teachers reported relying on observations to inform their judgements, only three felt well equipped to make these judgements. One stated being somewhat confident. One stated they had room to improve and six stated they were not equipped to make them. In addition, teachers found it difficult to articulate what helped them make good judgements at this point in time.

More specifically, when describing how assessment judgements relating to the patterns and relationships strand of algebra, and fractions learning, short and very general responses were provided. Clearly, for these teachers, although these assessment processes were used, the understandings they took from them were limited and unspecific.

4.2.1 Looking specifically at algebra

Six teachers (n=11) reported having limited, unclear or little knowledge about how students' learning in algebra progressed through levels 1 to 3 of the Mathematics Curriculum for Patterns and Relationships (Ministry of Education, 2015).

The remaining five reported having satisfactory knowledge of how student learning progressed.

4.2.2 Looking specifically at fractions

All but one teacher (n=11) reported having some or little knowledge of how student learning progressed in this area through levels 1 to 3 of the Mathematics Curriculum (Ministry of Education, 2015).

The themes related to the teachers' curriculum knowledge that emerged from the baseline data will be described and discussed in the next section.

4.3 Three keys themes regarding teacher curriculum knowledge

Three key themes emerged from the initial interview data which related broadly to teacher curriculum knowledge. These included knowledge about: curriculum content and learning progressions; curriculum content and delivery; and seeking and using resources to support planning and assessment decisions.

4.3.1 Knowledge about algebra curriculum content and learning progressions

When asked what they knew about how student learning progresses in relation to patterns and relationships, and about the activities with which students would be involved, teachers (ten of the 11 interviewed) typically responded by giving vague statements. For example:

Teacher 1: Um not much, um… what is it just the patterns?

Teacher 2: Their ability to be able to form a pattern

Teacher 3: This level it starts with the physical, making the patterns with the equipment, the counters, the bears, drawing patterns.

Only one teacher (n=11) described repeating and sequential patterns explicitly, stating how at Level 2 students should be able to identify them and continue them and find what the… pattern is and how to make the pattern grow. This statement illustrates that the teacher has a fundamental understanding of appropriate terminology, and the progression within Level 2 of the curriculum. However, for most of the teachers their descriptions lacked clarity and included general responses which illustrated a lack of key understandings of the progressions and content.

Closer analysis showed that nine of the 11 teachers were unable to offer insight into levels beyond the one they primarily found their students to be working in. One teacher could vaguely describe the Big Ideas at Level 1 but not at Level 2, the level her students should be working towards achieving at year-end. Some examples:

Their ability to be able to form a pattern… the more complicated the pattern the higher level they'd be right?

Or

I guess as it furthers up the school … it would be number patterns wouldn't it, you know counting, skip counting 2s and 5s and those number patterns.

Or

What is it, just the patterns?

These vague statements suggest a superficial understanding of the big ideas and progression of learning for algebra.

4.3.2 Knowledge about fraction curriculum content and learning progressions

The above findings were mirrored for fractions with again only one teacher (n=11) able to offer specifics about concepts to be developed at Levels One and Two of the Mathematics Curriculum. She stated:

So obviously level one is more kind of halves and quarters leading up as you go through. Level one is more kind of materials based and being able to find fractions of shapes and stuff, moving into that whole idea of being able to find a half and quarter of a group and then we do lots of like … if five is a quarter what's the whole.

This statement illustrates understanding of a trajectory of learning and how to build conceptual understanding of fractional thinking. Ten (n=11) teachers made unspecific or vague statements about how fractional thinking progressed through levels one to three of the Mathematics Curriculum. For example:

I think it's like halves and quarters and then it builds to like thirds and fifths

Or

Kind of hard to know cause I've been level 1 for so long… like a whole is a whole thing and if I cut it in half and exposing them to what a half looks like and gradually getting higher and higher and then quarters and eighths and da, da, da.

Or

I guess it's the number, the numeral, relating it to the pieces, is it normally like quarter and half… not clear on that.

These statements indicate a limited understanding of how fractional thinking progresses.

This section has explored how teacher response to initial questions on how learning progressed in each of the two focus areas took the form of general and vague statements that offered little insight into how understandings built along a learning trajectory.

4.4 Knowledge related to content and delivery

Teachers appeared to lack knowledge about the learning trajectories related to mathematical content within each curriculum level. This included the big mathematical ideas to be developed and appropriate learning activities to achieve this. For example, the following teacher statements illustrate a lack of specificity of appropriate tasks:

All I know at level one, it's more like hands on stuff so like cutting… food is a good one

Or

So starting off at the very basic, once again the visual, you know the half, the pizza's and those strip things.

Or

I'm one of those teachers who uses a pizza, cuts it up and matches it with the fractions.

Teacher knowledge of the learning trajectories was limited, as characterised by these vague descriptions that offered limited detail. The teachers did not appear to be able to detail specifically how to support students to develop different concepts and were unable to link big mathematical ideas with activities.

Only one teacher referred to finding fractions of sets separately to fractions of shapes (regions) and then described working back from the part to find the whole.

Three teachers (n=11) spontaneously expressed apprehension about teaching these content areas. One teacher stated*:*

I struggle with fractions personally.

Another described how she avoided teaching fractions altogether:

I didn't really touch it that much because it was a bit hard.

These statements represent a lack of confidence both in relation to understanding fractional concepts and teaching them.

4.5 Seeking and using resources to develop teacher knowledge

Seeking appropriate resources is one possible way of countering a lack of understandings of the key ideas which underpin many mathematical concepts. In the initial interviews, three of the 11 teachers named some of the curriculum supporting documents or frameworks they used. In an effort to gain further insight into what documents were referenced, the researcher prompted a further four of the remaining seven teachers about resources they used when planning or looking at how learning would progress. They were either unable to say where they sourced this information, or they gave unclear, confused and differing suggestions about where they looked. As one teacher explained: *I probably look at, um what are they called, they're not ICan's, they're not the learning progressions, like the old stages… the number framework stages, that gives me an idea.*

No reference was made by teachers to curriculum materials including the achievement objectives or curriculum elaborations.

Four teachers described their colleagues as resources and said they would ask them for advice on what to teach. This was illustrated by one teacher describing how she would seek input from a peer: *Say this is what majority of my class are doing, this is what I'm thinking of doing next, am I on the right track, should I be thinking of doing this, or do I need to take a step back?* This can be seen as an example of a teacher reflecting on their planning and seeking peer analysis of their planning and their students' next learning steps.

4.6 Intervention one

4.6.1 The use of comparative judgement

Comparative Judgement (CJ) as an assessment tool was described in Section 2.7 of the Literature Review. The current study used a CJ method as a means of collectively assessing student learning with open tasks. CJ as a tool for testing had not previously been utilised by the school in the current study. Teachers were given a brief outline of how the process worked and the researcher worked through a few judgements with the teachers collectively before they began casting their own decisions.

4.6.2 Comparative Judgement results

Overall reliability of 0.9% was achieved for both of the tasks judged (Level 1 problem Patterns and Relationships and Level 2 to 3 problem Patterns and Relationships). This reliability level is extremely high, indicating overall consistency between judges was maintained. Further detail on this CJ reliability and task results is included in Appendix J: Comparative Judging reliability and results for patterns and relationships. Judging results were shared with teachers.

Teachers' feedback in response to interviewing is reported on in the next section which has been organised around four broad headings. These headings are: the hypothetical learning trajectory and teacher knowledge; collaboration; assessing conceptual understanding and open tasks and; comparative judging as an assessment process.

4.6.3 The Hypothetical Learning Trajectory (HLT) and growth of teacher knowledge.

Following their engagement in the professional development and with using comparative judgement, all teachers reported that having the shared planning document which included the co-constructed hypothetical learning trajectory (HLT) [See appendix K], as a reference point was extremely useful. This document and the HLT was a direct result of the collaborative planning session, informed by the Curriculum Elaborations. All teachers specifically described referring back to this shared document either during planning and teaching or when making judgement decisions about students understandings. As one teacher reflected:

> *I like it because it allows you to see the whole journey a child is going on and sitting particularly within our classes we have learners who are sitting within level 1 so it's really nice to see that progression.*

She continued by describing how after completing the comparative judgements, her understanding of the trajectory within each level had further developed, *so you're really looking at say, were they able to continue that pattern? How did they represent whatever data they* *collected from it? So it makes it really, really clear what is within each level.* This illustrated her growth in understanding of the trajectory of learning which she clearly attributed to her use of the criteria for each level in the curriculum.

Four teachers (n=11) reported using the HLT *like steps*, which gave them a *better understanding* of where the learning needed to go and how they could help it to get there. This new understanding positioned these teachers to scaffold students thinking and make logical, sequenced connections to key understandings described in the HLT.

In another instance, a teacher explained how she had noticed how some students displayed a multiplicative understanding of the growing pattern in the Level 2 and 3 Patterns and Relationships Assessment Task[1] by recording the functional rule. She added on how this extended their thinking into Level 3 of the curriculum. This teacher had built her knowledge of the curriculum levels through her use of the HLT.

Six teachers (n=11) stated how knowing the HLT enabled them to keep pushing students as they knew where the learning could go next. For example, three teachers reflected together:

Teacher 1: *So the progression, that's really helped me. It's gone beyond oh let's just make a pattern, making patterns, it's not just about that*

Teacher 2: *It clarifies the where to next*

Teacher 3: *And being able to extend them in that lesson as well, knowing that where to go, so you're doing your lesson and they've got it, you know where you can go with it to extend them. Taking that learning further.*

These teachers had attributed their ability to extend students' thinking to their understanding of the HLT. They also recognised that they now knew how the big ideas built on from each other. This provides clear evidence of their growing conceptual knowledge.

Eight teachers (n=11) stated how knowing the sequence of learning and having a reference supported them to notice gaps in student understanding. As reflected by one teacher, *it was good seeing the gaps in their learning, like they could continue a pattern and they could tell you the pattern but when they made their own pattern it could have been AWOL* (Absent without leave). She also referred to the specific sequence of big ideas covered by the HLT within Level 1

[1] Appendix B : Patterns and Algebra Level 2 and 3 Assessment Task

of the mathematics curriculum for patterns and relationships. This shows how she was now able to offer specific and detailed insight about her students' learning.

A teacher reflected at length on how the planning session had helped her develop an understanding of ordinal numbers and the role they played in building conceptual understanding in algebra. She described how this was a new approach for her: *In all my years of doing patterns, I've never thought to say show me the 8th, show me the 15th, you know, linking that ordinal position to the pattern. Just never thought.* Another teacher then added on, a*nd I wish I'd spent more time on that this time round, I might go back to it later on and re-do it.* These reflective statements demonstrate a clearer understanding of what big ideas and relationships need to be made explicit to students. The first teacher had connected to how explicitly associating ordinal numbers with positions in a pattern and valuing this practice made these taken as shared understandings visible to all. The second teacher, by reflecting with her peers, had noticed the benefit of these understandings and acknowledged the need to revisit this big idea in the future, thus ensuring her students can internalise these shared understanding.

Both these teachers were able to discuss specifics of their students learning with their colleagues. Using the HLT and having an understanding of the big ideas therein has given them precise vocabulary, not previously used, to describe their students' learning with added detail. The next section moves on to look at findings associated with the collaborative nature of the intervention.

4.6.4 Collaboration and the power of the collective

As described in the Methodology Section, 3.6.2 Professional development and collaborative planning activity intervention, this study intentionally utilised collaborative approaches in order to construct shared understandings amongst teachers and across year levels. All teachers reported the collaborative nature of the planning session useful. As stated by one teacher, *planning as a Learning Neighbourhood also allows for key discussions that help us to have a shared understanding of learning outcomes.* In this comment, she highlighted a useful outcome of the collaborative planning session that assisted with school-wide moderation. In another instance, a teacher stated *I have really enjoyed planning together, and the sharing of ideas and resources*. This statement has highlighted the importance of teachers having space to plan together and build collective understandings.

The first iteration of planning and teaching/learning focussed primarily on planning the problems and little time was allocated to anticipating solutions. Four teachers fed back that they would appreciate more time spent together, anticipating how students may solve the problems and unpacking possible misconceptions in a collaborative format. They stated that they had benefitted from the collaborative planning of problems and wanted this extended into the anticipation aspect of the planning too.

This became a more specific focus during the second collaborative PD session in phase 5 outlined later in this chapter (see 4.7 for more detail).

4.6.5 Assessing conceptual understanding and open tasks.

When teachers were asked what they noticed about students' responses in relation to patterns and relationships, the *openness* of the assessment task was commonly seen as a strength. These strengths have been separated into two categories, those relating to differentiation within the assessment problems and, those relating to student interest. A thread throughout this section is how these tasks supported teachers in assessing students' developing conceptual understandings.

4.6.5.1 Differentiation

Nine teachers of the 11 described how the assessment problems allowed all learners access to the tasks. As one reported they were not afraid to give the problems a go and similarly, a second teacher stated *they weren't afraid to show their working like some of them had scribbles that they'd tried.* These teachers had all acknowledged how the openness of the problem gave their students a way to access the task. One of the nine teachers extended on this idea stating, *it showed me their go-to strategies and what they tended to do by themselves*. Through her statement, we can see her analysis of how the problem offered her students space to show their learning in a way that made sense to them.

Three teachers specifically reported that their students were *giving more than expected.* Similarly, an additional four teachers described how their students tried multiple methods of recording their thinking. For example, one stated *it was awesome to see our kids making those relationships… giving more than one way of working it (*out*) and like the tables but also… dabbling into multiplication and awesome to see some of them using graphs and things like that.* In this example, the teacher identified the value of making connections between different solution paths used by her students, and how the openness of the problem facilitated them to try multiple methods.

During discussions two groups of teachers (six teachers in total) discussed how during assessment tasks their students had attempted multiple solution strategies. They noticed that often students recorded initial attempts, crossed them out and then tried different approaches. These teachers highlighted how their students *weren't afraid to make those mistakes,* that they were undeterred and rather than give up *they tried something different.* Overall, the teachers valued the positive mathematical disposition the students were displaying. The design of the assessment tasks provided students with an opportunity to recognise that problems can have more than one solution and that mistakes are an important part of learning mathematics. It also demonstrated to the teachers that their students were able to demonstrate deeper, richer conceptual knowledge.

Two teachers, when discussing how some of their students performed with the Level 1 Patterns and Relationships Task[2] specifically described students' gaps or misconceptions. Together they reflected on how some students were unable to create their own patterns. The first teacher described how *they could continue the pattern, they could describe a pattern, I think I even had one little fella who found the 8th and the 15th but he could not make his own pattern.* The second teacher added on to this, agreeing with her observations and describing how *a lot of my kids could read it and explain it and do it and carry it on but when I said create their own pattern they just got five different colours and didn't*. From these statements, we can observe that the teachers were beginning to carefully consider both their students' conceptual understandings as well as specific areas that needed to be addressed and explicitly unpacked.

4.6.4.2 Student interest.

Student interest or engagement was another theme which emerged. Six teachers of the 11 described student interest or engagement as a strength when discussing what they noticed about their student's responses to the assessment tasks. For example, specific statements included that students were *eager to show everything they knew*; *keen to draw* and, *focussed on their papers*. These six teachers highlighted how motivation-related factors such as interest generated by the openness of the problem, had a positive effect on how their students engaged with the assessment tasks.

4.7 Intervention two

Phase five of the study involved the second intervention. A similar collaborative planning session was facilitated where the curriculum elaborations that related specifically to the fractions part of number were unpacked as learning outcomes into a shared document. As a direct result of the feedback on the first planning session, this time teachers focussed not only on planning the instructional tasks but also on anticipating possible misconceptions or ways students may

[2] See Appendix A: Level 1 Patterns and Relationships Task

attempt to solve them. This additional level of planning was added to the collective document as a point of future reference. Teaching and learning continued for four weeks after which students completed one or two open assessment tasks. As with the first CJ session a reliability of 0.9 was achieved, once more proving the method to be highly reliable with consistency between judges maintained. [See Appendix L: Comparative judging results and reliability for fractions tasks, for details].

This reliability was checked by analysing representative samples at different percentile ranks. Again the respective percentile ranks matched with differing levels of sophistication, also detailed in Appendix L.

This section discusses the feedback received that relates directly to the CJ process. Please note that these are overall findings and in response to both CJ sessions.

4.7.1 Feedback on the comparative judging process and results

The process of comparative judging (CJ) was received with mixed reactions from teachers. These will be discussed under the headings of reliability and implications on teacher knowledge, technical details and time constraints.

4.7.1.1 Reliability

Nine teachers (n=11) appreciated the reliability the CJ process brought to the assessment. Some compared it with moderation and discussed how the process allowed all their students work to be viewed by others, rather than just one or two as per traditional moderation procedures. For example, *I like the reliability of it, I think it's more reliable because there's more people contributing to the end result.* This statement shows acknowledgement that when teachers are judging with a shared understanding of learning outcomes, the results are more reliable.

The idea of anonymity was specifically raised as a positive benefit by three of these nine teachers. As stated by one teacher, i*t's also quite good as well when you don't know whose work it is… and so if we knew say the person, we'd be like oh right, you don't know if you'd lean towards that person because you know them.* As this teacher described, the anonymised nature of CJ eliminated unconscious bias.

One more senior teacher expressed concerns with how a couple of individual transcripts were ranked, particularly those that showed higher order strategies. When referring to the top ranked transcripts she stated:

> There were some learners whose pages looked the same/similar, but they were awarded different levels/judgements… so as with other assessment tasks you need to consider how it reflects what you already know about the learner/other assessments etc. rather than basing it on the one sample.

Within the percentile rank 80-100 there are 11 out of the total 283 student candidates. This represented 3.9% of the total number of candidates. When looking closely at these transcripts the student ranked at the 82nd percentile arguably had a clearer, more structured representation of equivalent fractions that suggested a more embedded conceptual understanding than the student ranked at the 100th percentile. These transcripts are included in Appendix M.

The comments made by this teacher imply a sophisticated understanding of the big ideas developed at Level Three of the mathematics curriculum for fractions. Furthermore, she has referred to the idea of basing assessment decisions on more than one assessment task, as per the guidelines for making overall teacher judgements (OTJs) which build from the understanding that no single information source can accurately summarise a student's achievement (Ministry of Education, 2019).

It is worth noting that these results, although a very small percentage of the overall judgements, are interesting. Four teachers expressed their lack of confidence with judging the Level 2 to 3 Fractions Assessment Task[3] and perhaps this resulted in some discrepancies with how the more sophisticated student responses were judged and subsequently ranked. As one teacher reflected when asked what steps she took to choose the best response when judging the fractions transcripts. *To be honest, the Level 2 I didn't really, my pedagogy wasn't there so … that's when I asked Genny (pseudonym). Only because I didn't know that level, even though I had the goals there.* Similarly, a teacher described how she *lack*[ed] *confidence in regards to the early level 2.* This suggests a working knowledge of the strategies and conceptual understandings that develop further along the HLT was not accessible to all teachers.

The way the comparative judging was set up allowed teachers to converse (if desired) with adjacent individuals. Eight teachers (n=11) described this collaborative system as a positive. As stated by one teacher, it meant that there was someone to check those *trickier ones* with. This statement highlights how the teacher sought peer support when faced with more complex judgements.

[3] See Appendix D: Level 2 to 3 problem Fractions

This section has discussed findings related to the reliability of the CJ process. Next findings related to the more technical considerations are addressed.

4.7.1.2 Technical considerations

Recording of student and teacher voice on scripts, particularly those from the first testing iteration, varied considerably. It was sometimes unclear if what was written on response sheets was the teacher's interpretation or what the student had said. As a direct reaction to this initial feedback, it was jointly decided that student voice would be recorded in quotation marks after the letter S. It was also noted that where representations (particularly those of more junior classes) were obscure or more difficult to decipher, that student voice was very useful. For instance, one New Entrant child drew a line down the middle of his page for the Level 1 to 2 Fractions Task[4]. When asked by his teacher to *tell me about your drawing* he said he had *drawn it in half*. Without this student voice scribed on his response sheet a random line down the page would not be automatically viewed as the student dividing the page into half.

In both instances, assessment tasks were launched by teachers and students then set about solving them by recording their strategies and thinking on their answer sheets. How these problems were launched as well as how much or, what teacher prompting followed was not planned or mapped out and this was raised as a concern by four of the 11 teachers. These four teachers raised an important point. The idea of planning 'script' type prompts that could be used to support teachers in this process was suggested by one teacher during the second judging session as a way of ensuring problems be launched without pushing students towards a certain way of thinking or solution route.

Two teachers from different cohorts described how their students were unable to start the level 1 fractions problem saying *they just sat there*, and *they just looked at me blank*. These findings were atypical and could be due to a number of unaccounted for factors. Nonetheless, their students inability to represent their understanding of a half without significant prompting, suggests the way they have tackled tasks to date has potentially been over-scaffolded. Additionally, they showed that the students needed exposure to open tasks within their class programs if they were to be expected to tackle them during assessments.

> *The assessment question was very different to what we would normally do. It was very open. For me, I'd done lots of halves or groups, so half of ten, whatever. I was trying to prompt them… "Remember we had the packet of jellybeans and they were sharing*

[4] See Appendix C: Level 1 to 2 problem Fractions

them". So, I was trying to give them those things. They're like, "oh, yeah, that's right, there was eight, so I'm going to make half". So, giving them those prompts to help them show half of a packet or whatever because I know that they could do it rather than just cutting a circle".

This extract suggests that the teachers had realised that these students needed further opportunities to generalise their evolving conceptualisations. They recognised that the students had not built sufficiently robust understandings to justify, explain or apply them in their own ways, and are overly dependent on set procedures and the teacher's support.

4.7.1.3 Time constraints

All teachers made some mention of the time it took to make the actual assessment judgements. For the second CJ session teachers took between 8.1 and 21 seconds per judgement with the mean time being 12.2 seconds. This equates to a mean time of 2440 seconds or 40.7 minutes. The assessment regime prescribed to at the school in question remained unchanged for the duration of the current study with these additional assessments and CJ process "added on" to those already in place. The double up of time was a factor seen as a negative for three teachers. The 40 minutes it took to make these judgement decisions was part of an afternoon assessment session that took place when staff meetings would normally occur and not set as an additional time commitment.

Two teachers stated that administering the first assessment was time consuming and *hard*. Further questioning by the researcher revealed that these teachers tested students either individually or in small groups. The first teacher stated:

> *I ended up doing it like two at a time which took a whole bloody week* and, the second stated *So I broke them up into two groups of five and then some people were away… even then that was tricky then to actually manage and capture what they were doing.*

These comments were atypical and not reflective of the other nine teachers interviewed. The researcher took the opportunity to model administering the second round of assessment tasks and also explained during the second PD and planning session how best to administer the assessment.

Even though the majority (eight) of the teachers described benefits from the CJ process, it is clear that time constraints related to assessment practices were a concern for some teachers.

4.8 Final data

This final section addresses findings that emerged as a result of the third and final interview session, following both interventions and CJ sessions. Where findings mirror those described within the previous sections of this chapter, following the first intervention, they have not been repeated. Teacher responses that add depth or are supplementary have been included. Additional data was collected in response to the final set of interview questions and is reported under two headings; additional implications for teacher knowledge and confidence, and collaboration.

4.8.1 Additional implications for teacher knowledge and confidence

The CJ process meant all teachers saw and judged responses from Year 1 through to Year 4. Three teachers reported this time round that they benefited from seeing assessment responses presented *in another way*. This included reflections that being exposed to how students from other classes represented their understandings provided them with other ways they might "teach" fractions in the future. As one teacher stated *I was the classic circle so all my kids just drew circles in their assessment where it would have been good if I had done rectangles, other shapes, just to see if they could get it in another way or another shape.* The response from this teacher indicated that she was receptive to new ideas, and noticed these.

At this point, all teachers gave added support for the collaborative planning process. More specific mention was given to planning of the actual problems and a shared document that was collectively added to. As stated by one teacher:

> *When we came together as a staff and we were trying to think up problems, because that's sometimes the hardest thing trying to think up a problem that's going to target that particular learning outcome. That's what, four heads or five heads is better than doing it on your own.*

The collaborative approach had helped this teacher with her day-to-day planning (See Appendix N: Co-constructed plan for fractions with two problem examples). This had removed a previous barrier and been the scaffold she needed to write appropriate problems.

In another instance, one teacher stated while describing the process of planning and anticipating, that it was *really helpful especially* (anticipating) *those misconceptions, and the way that kids might work them out, because even like some things that Sally (pseudonym) put, I just didn't even think about them.* This teacher had built an understanding of the breadth of big ideas covered by the learning outcomes in the joint plan. She had also noticed the benefits of anticipating not only successful solution strategies but also potential misconceptions as points from which to build understanding. Though the collaborative planning process, she was being exposed to new ways of thinking.

4.9 Summary

This chapter has detailed the findings of the study. It began by outlining the three key themes that emerged from of baseline data gathering. These included knowledge about: curriculum content and learning progressions; curriculum content and delivery; and seeking and using resources to support planning and assessment decisions. Following the first intervention, the use of comparative judgement as an assessment tool was discussed and the reliability of the results presented. The impact of the HLT was discussed next and how its use supported the growth of teacher knowledge identified. Then the collaborative nature of the intervention and power of the collective understandings built were detailed. The opportunities for assessing conceptual understanding using free-response tasks was addressed next. Following the second intervention, feedback on both comparative judging sessions was detailed. Lastly, results following the third and final interview session that gave supplementary or additional information relating to teacher knowledge and confidence were reported.

Chapter Five: Discussion

5.1 Overview

This chapter discusses what was reported in the Findings. Section 5.2 begins by looking at how teacher knowledge improved, characterised by the teachers' growing ability to describe in detail their students' learning in relation to the collective HLT used. Section 5.3 unpacks further how teachers used the HLTs and the impact this had on their teaching and student learning. Section 5.4 then looks at how using open assessment, free-response tasks supported teachers to assess student conceptual learning, touches on the unaccounted for benefit of student interest, and addresses the atypical results reported in Chapter 4.7.1.2 Technical considerations. Finally Section 5.5 discusses Comparative Judging as an assessment tool.

5.2 The impact on teacher knowledge

Shown in the data is a clear picture of how the teachers shifted from making general and vague comments about student learning in specific mathematical strands to making specific and detailed descriptions of the curriculum content. This included developing both depth of knowledge but also breadth across Levels One to Three of the Mathematics Curriculum with support of the Hypothetical Learning Trajectory (HLT) developed from the Curriculum Elaborations.

Unpacking the Curriculum Elaborations and building a collective understanding of ideas meant all teachers had been supported to develop a shared understanding of the trajectory of learning. The HLT had become clearly defined and available to them. This supported teachers to make connections with the ideas and develop their own understandings. For assessment, planning, and teaching purposes teachers must be able to articulate and describe what they notice about their students' learning. In contrast to the vague and general descriptions given prior to the intervention, teachers had begun to give precise descriptions of what students were doing in relation to the HLT. These findings are similar to those reported by Holt Wilson (2014) who found that the *equipartitioning learning trajectory* (ELT) actioned in his professional development intervention study, positioned teachers to give richer descriptions of students' thinking through daily observations of performance.

In their multi-year project titled the BLINDED Project, when interviewing teachers about LTs, Superfine and Wenjuan Li (2017) found that teachers described the "*sequences* of student learning and *obstacles* students typically encounter[ed]" (p. 1259). Their research found a large variation in the detail of these descriptions, especially when describing the specifics of students thinking. They introduce the term grain size, where the larger the grain size, the less specific the detail. Likewise, in the current study teachers' descriptions reflected those of a finer grain size following the various interventions. Specifically, the language teachers used to describe their students mathematical learning, use of mathematical practices, and conceptual understandings shifted from being short and general to being detailed and specific. It shifted to include mathematically appropriate terminology that linked back to the big ideas and learning outcomes detailed within the HLT for each learning area. Furthermore, teachers engaged with each other in sophisticated discussions about the specifics of their students' learning using this newly acquired language. These findings mirror those of Holt Wilson (2014) who concluded that the PD on the ELT in his study enriched teachers' ability to discuss specifics about students' progress in equipartitioning with their colleagues.

5.3 Referencing collective Hypothetical Learning Trajectories

Improved teacher knowledge, characterised by the shift towards using specific, mathematically rich language was attributed to the collaborative professional development and collectively produced HLT. The teachers described using the HLT they built *like steps* and, that this gave them a *better understanding* of what they needed to teach. These claims are similar to those made by Sarama et al. (2017) who found when questioning teachers following their LTBPD intervention that teachers explicitly referred back to the individual components of the learning trajectories, describing how their understandings of these had improved. The teachers in their study described “the process of becoming familiar” (p. 69) with the developmental progressions, appropriate instructional activities and set goals.

A teacher’s ability to notice students mathematical thinking so that they can make both in the moment and planned decisions about which ideas to follow requires having in-depth and cohesive knowledge of how student’s conceptions interplay. If building student conceptual understanding depends on teacher knowledge of HLTs and teachers do not have this knowledge then making accurate observations and subsequent judgments of student learning will be near impossible. Well-informed teachers support students by helping them build an understanding of the intermediate steps necessary for conceptual understanding of the big ideas or learning goals. Sullivan et al. (2009) explain that it is critical teachers know how concepts and ideas relate in order to make these ideas accessible to students. This supports the claim made by Sarama et al. (2017) that professional development in LTs is critical if teachers are to be skilled at connecting these ideas in a way that students can learn them. The interventions in the current study positioned teachers to make these connections. This shift meant teachers knew what content to cover, in what order to present it so that it *made sense* and, how learning outcomes that were planned into the HLT subsequently related to the mathematics curriculum levels.

A number of benefits emerged as a result of utilising the shared HLT. These benefits are discussed under the heading differentiation.

5.3.1 Differentiation

From the findings, it can be seen that once teachers had access to the sequence of learning and ideas to be built they were better able to notice *gaps* and misconceptions students held. As part of their BLINDED Project, Superfine and Wenjuan Li (2017) identified the concept of “gap” in teacher reflections and defined this as “a lack of prior math knowledge that students should have become proficient in previously” (p. 1259). In the current study, noticing these meant that teachers were better positioned to effectively target these gaps or misconceptions. Furthermore, the idea of making the implicit explicit is a key instructional practice that especially supports struggling learners (Selling, 2016; Warshauer, 2015). A teacher's ability to enact this practice becomes more accessible when they are familiar with the HLT and key understandings underpinning the mathematics they're presenting to their students. In other words, teachers in the current study were supported to ensure any taken-as-shared or implicit understandings could be made explicit and available to all students.

Having an understanding of the HLT not only enables teachers to support struggling learners make connections, it also supports them to extend learners working at higher achievement levels. These teachers were positioned to notice and predict students' mathematical activity and judge their developing conceptual understandings. They noted ways they could push their students towards more sophisticated levels of understanding and attributed this to having a clear map of how the ideas built on each other. This finding is similar to that of Holt Wilson (2014) who found having the "conceptual development" mapped out in his ELT enabled teachers to "anticipate the ways students might engage with subsequent instruction" (p. 235).

Similar to the findings of Sarama et al. (2017), teachers in the current study expressed an appreciation of the HLT, especially the planned instructional tasks as "tools for teaching" (p. 69). The idea of structuring tasks along an HLT is supported by Sullivan et al. (2016) who describe applying variation theory to the order in which tasks are presented to students. That some elements of the initial task remain invariant and others change in order to show students that "their new knowledge is flexible and does not just apply to problems of the original type" (p. 162).

5.4 Using open assessment tasks

One of the key factors that supported teachers in the current study to assess student's conceptual understanding was the use of free-response, open assessment tasks. This section discusses these findings and reports on an unexpected benefit that emerged from the data – that using open assessment tasks also supports student interest and engagement. Lastly atypical results from two teachers are discussed to ensure a rounded view of all data is presented.

5.4.1 Assessing conceptual understanding with free-response, open tasks

As shown by the analysis, teachers found that using free-response, open assessment tasks provided deep insights into students' conceptualisations. When presented with open assessment problems, students would display their *go-to* strategies, which offered useful insight about their conceptual understandings. Teachers found that they could assess how students applied their thinking to new situations; if they had generalised concepts taught, if they had misconceptions or gaps in their knowledge. A study by Sole (2018) supports this finding when she discusses how open-ended assessment questions provide opportunities for teachers to learn what strategies their students choose to employ as well as what strategies they “reject or do not even consider” (p. 463). The current study found that by using open problems, students’ conceptual understandings were clearly presented, providing invaluable feedback to the teacher.

A benefit of using open-ended problems is that they offer multiple solutions and pathways to get there. The openness of the assessment problems used in the current study meant there were multiple entry points so students with varying levels of understanding and prior knowledge were able to successfully engage with the tasks in some way (Hodge & Walther, 2017). The notion put forward by Varygiannes (2013/2014), that “asking less may indeed enable a learner to demonstrate more understanding” (p. 278) was found to be true for the current study with reports of students attempting multiple solution strategies that often surprised their teachers, again providing further feedback on their conceptualisations. Kabiri and Smith (2003) describe how open-ended problems give students the opportunity to go beyond the expectations of the teacher and the current study supports this notion.

5.4.2 Open tasks and student interest.

Student interest and engagement are key to learning. In their experimental study into whether multiple solutions matter, Schukajlow and Krug (2014) found that using open ended problems with vague conditions that allowed for multiple solutions, had a positive effect on students’ autonomy and competence which in turn improved motivation-related factors such as interest. Similarly, in their report from two studies that looked at factors supporting positive dispositions, interest and engagement in mathematics. Mueller et al. (2011) developed a framework that begins with posing of “an open-ended, engaging, and challenging task that the students have the ability to solve” (p. 40). They concluded that from this starting point, and with the support of various contextual factors, students build positive dispositions towards mathematics, intrinsic motivation, self-efficacy and autonomy, all resulting in mathematical reasoning and subsequent conceptual understanding.

Nine teachers in the current study reported their students were enthusiastic about tackling the assessment problems, that they tried multiple solution pathways and often had a number of attempts at individual problems. These teachers had instilled in their students a shared understanding that in mathematics there can be more than one correct answer to a problem or multiple ways to solve it. These understandings are supported by Varygiannes (2013/2014) when he describes the benefits of using open problems. The students in the current study have had to make decisions about which strategies to employ and how to proceed, a complex process that builds persistence, independence and initiative. These notions are supported by Sole (2018) who explains how using open-ended problems pushes students to go "beyond procedural proficiency to make and justify decisions" (p. 462).

5.4.3 Atypical results related to open tasks

Two teachers from different cohorts reported contrasting results for the Level 1 Fractions problem, stating that their students were unable to start the task or represent their understanding of a half without significant prompting. Sole (2018) states that implementing open assessment problems bring additional classroom management considerations including time allocation, and the need for group work. She explains how students need time to work collaboratively in class at these types of problems before they can be expected to confidently approach them on their own. This thinking is supported by Sullivan et al. (2016) who, in their study that examined posing challenging tasks to prompt for problem solving and reasoning while developing persistence in students, found that having suitable tasks and lessons "is necessary" but it does not automatically guarantee learning (p. 169). Clearly, the students in the current study needed further opportunities to grow their confidence to tackle open problems by building sufficiently robust understandings in class that are less dependent on set procedures and teacher support or prompting. As stated in the findings, these students needed further opportunities to generalise their evolving conceptualisations. The author acknowledges that other factors such as teacher perceptions/beliefs, pedagogical content knowledge and lesson structure may be at play here (Sullivan et al., 2016) but these notions are outside of the scope of the current study.

5.5 Comparative Judging as an assessment tool

Comparative Judgement (CJ) was found to be a highly reliable assessment tool that supported teachers to make consistent judgements of students' conceptual learning. As described in the previous section, the use of free-response questions allowed widespread opportunity for student thinking to be made visible to these teachers. This approach, along with the CJ process provided space for teachers to accurately appreciate and assess the breadth of student understanding in each area since the CJ results were a direct consequence of teachers' collective understanding of each co-constructed HLT and related big ideas. These findings mirror those of Hunter and Jones (2018) who concluded that CJ offers teachers a "window onto children's mathematical thinking" (p. 406). As explained by Jones and Inglis (2015), this approach means teachers are not imposing, with set mark schemes, narrow ways of interpreting the tasks. Instead, the CJ process "assimilates the varied ways" (p. 342) all teachers in this community interpret the construct. Subsequently, CJ offers a means of reducing bias in teacher judgement and, with the justified desire to assess conceptual in addition to procedural learning, offers an alternative to historically one-sided approaches.

Through the CJ process, teachers were exposed to examples of student responses from a range of year levels and classes within each level. This process supported teachers to build new conceptualisations themselves, develop their own understandings and in turn fine-tune their pedagogical approaches. This new knowledge positioned these teachers to progress their students conceptualisations by making connections to the shared HLT and big ideas in new and different ways.

5.6 Summary

This chapter began by discussing the impact the interventions had on teacher knowledge. It elaborated on how teachers' comments of student learning shifted from being vague and general descriptions towards becoming specific, detailed and referenced back to the HLT. How the HLT supported teachers to unpack their students developing conceptualisations was addressed and this led to a discussion of how the HLT positioned teachers to effectively differentiate for all students. The extensive benefits that came from using free-response tasks were discussed next, and how these tasks enabled teachers to make judgements on students' conceptual learning detailed. Finally the appropriateness of the CJ as a tool for assessing students' conceptual learning was discussed.

Chapter Six: Conclusion, and Implications

6.1 Conclusion and implications

Teachers and schools are shifting their focus from teaching and learning regimes that prioritized procedural mastery towards those that prioritise building conceptual understandings. With this shift, there is a growing mismatch between what is taught and what is tested, since testing regimes primarily seek to assess procedural skill over conceptual learning. Schools must therefore rely heavily on teachers to judge students developing conceptualisations, until such time that assessment procedures better align with the outcomes sought by the education system (Jones & Inglis, 2015). Since these judgements are shown to be subject to bias (Smaill, 2013), affected by various contextual influences (Meissel et al., 2017), and often inconsistent between classes within schools, it is important to examine how teachers can be scaffolded to make accurate and consistent judgements of student conceptual learning for both assessment and learning purposes. The current study has investigated how building a collective understanding of the big mathematical ideas and learning trajectories within two areas of the mathematics curriculum, enabled teachers to make consistent judgements on student's learning. It also considered what factors supported teachers when assessing students conceptual understanding.

Teacher knowledge of how student learning progresses and connects the big mathematical ideas can either support or hinder student learning. Baseline data identified that teacher knowledge of curriculum content, delivery and resources was limited. Consequently, in the current study this was addressed and teacher knowledge was supported through carefully planned, collaborative professional development that involved using the Curriculum Elaborations to collectively map hypothetical learning trajectories (HLTs) for each of the curriculum strand areas covered. These learning and planning sessions produced shared documents that included learning objectives along an HLT, learning tasks at each curriculum strand level and, in some instances, anticipated solution pathways. Clearly, this resulted in enhanced teacher knowledge with the possibility of positive outcomes for student learning. Schools need to be considering teacher knowledge and how to grow it and this current study suggests one successful way.

If teachers have limited knowledge of curriculum content and progressions their ability to teach and assess the curriculum strand area will be extremely limited. The current study found that professional development was critical to position teachers to make accurate, referenced observations and assessment judgements. At the start, teachers lacked understanding of how student learning progressed through levels one to three for each area, and lacked knowledge about the mathematical content that needed to be covered within each level. This was characterised by their short and vague descriptions that offered limited detail about learning or progress, prior to the interventions, a finding mirrored by others that have researched HLTs and learning trajectory based instruction (LTBI) (Holt Wilson, 2014; Sarama et al., 2017; Sztajn et al., 2012). How schools are to manage the professional development of their teachers needs careful consideration and even an individual response may be needed for the different teachers. This also has implications for policy makers. They need to recognise the importance of professional development as a way to enhance student learning.Significant changes were found in how teachers discussed students learning following the interventions. Teachers used specific and detailed language that often included sophisticated mathematical terminology, referenced back to aspects of the HLT and ideas built within the collaborative planning documents. The shared documentation and collaborative nature of the intervention supported teachers to build collective and detailed understandings related to curriculum content, delivery and learning progressions for each area. These clear and sophisticated understandings supported teachers with their curriculum delivery by positioning them to notice student's conceptions and link these back to the HLT. This practice facilitated differentiated instruction where teachers supported struggling students by making explicit connections, noticed gaps and misconceptions, effectively planned next learning steps, and extended higher achieving students towards the next level of conceptual understanding. A clear implication here is that teachers grow and learn when working together in a safe collaborative space where they can explore together their current understandings and misunderstandings. However, it would seem that the facilitation needs to be a knowledgeable 'other' to lead the building of a collaborative HLT.

As referenced within section 2.2 of the literature review, the use of free-response assessment questions is not yet common practice but acknowledged as a way of gaining invaluable information about students conceptual thinking (Sullivan et al., 2006). This was true for the current study where the free-response assessment tasks offered deep insight into students learning. The openness of the problems meant all students could engage with the tasks, offering invaluable feedback to teachers about their gaps, misconceptions, developing conceptualisations and preferred strategy choices. The nature of the tasks offered students the opportunity to attempt multiple solution strategies, which exposed the breadth of their understanding. Furthermore, these students were motivated and engaged, ready to persist with the challenge that each problem presented. What became clear during this investigation was that students need opportunity in class to engage collaboratively with open problems to build their independence from the teacher and learned procedures. In this way, teachers will reap the full benefits of using the free-response tasks when assessing their students and making judgements on their learning.

The Comparative Judgement (CJ) assessment tool supported teachers to be consistent and free from bias when judging the breadth and depth of student learning, since the CJ results were a consequence of their collective understanding and interpretation. Nonetheless, these teachers found the additional time commitments onerous, over and above the pre-existing testing regime. This factor has implications for teachers and suggests that schools need to re-examine their current assessment regimes and decide on the benefits of what they do. The use of the Comparative Judgement tool should not be an add on but rather a replacement for out-dated and out-moded practices.

6.2 Limitations

The current study explored how a collective understanding of the big mathematical ideas and learning trajectory, in two areas of mathematics, supported teachers to make judgements on their student's learning in one New Zealand primary school. It also identified factors that supported teachers in assessing their student's conceptual understandings. All teachers participated in the professional development, teaching and learning, and assessment interventions and, all had access to the collective planning documentation developed. The extent to which these plans were implemented was down to the individual teacher. Furthermore, the classroom culture, curriculum delivery and behaviour management practices were not considered when looking at the impact of the interventions and this was detailed within section 3.8 of the Methodology. The current study chose an open, free-response assessment task approach to assess depth and breadth of student learning. If choosing this type of assessment task teachers must ensure their students are familiar with how to tackle them by providing opportunity and time to do so in class. The fact that two teachers reported their students could not access the open assessment tasks without significant prompting, suggests that how these plans were implemented, and the un-accounted for classroom-based factors, played a role for these students.

A further limitation was that the current study was in one primary school in New Zealand and involved only eleven teachers. The results may have been different with a larger group of teachers or in a different school context.

6.3 Final comments

There is a changing focus and global shift towards building student's conceptual understandings in addition to their knowledge of procedures (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). However, the complex task of assessing these conceptualisations lies with classroom teachers who currently rely on assessment practices that prioritize assessing the procedural. The current study offers a way to support teachers, through collaborative professional development on learning trajectories and the big mathematical ideas, to make consistent, accurate and effective judgements of students learning and specifically, of their conceptual learning for both teaching and assessment purposes. It presents significant benefits to using free-response tasks and considers a Comparative Judgement assessment tool to combat teacher bias and support consistency.

Appendices

A: Level 1 problem Patterns and Relationships

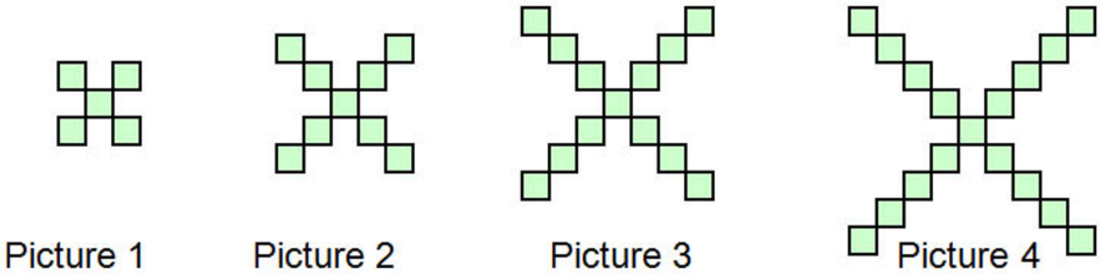
Nevaeh is eating jellybeans and she likes to eat them in this order:

Can you describe and continue the pattern?

If you were eating jellybeans what pattern would you eat them in?

What colour would the 8th jellybean be? What about the 15th? What about the 31st?

B: Level 2 to 3 problem Patterns and Relationships



Write and draw everything you know about this pattern

– this could be continuing the pattern, describing the pattern in words or numbers, drawing a table or ordered pairs, showing a graph of the pattern, writing the rule for the pattern in words or symbols.

C: Level 1 to 2 problem Fractions

Write and draw everything you know about halves and quarters.

Are there any other fractions you know that you can write and draw about?

D: Level 2 to 3 problem Fractions

Heather says she can prove $\frac{3}{5}$ is the same as $\frac{6}{10}$

Do you agree or disagree with her? Can you explain why you agree or disagree?

Can you write some other fractions you know that are the same? Can you prove they are the same in different ways?

E: Interview 1: Baseline and initial interview questions

Assessment and teacher judgements
Baseline data questions:

1. What are your current mathematics assessment practices and why?
2. How do you use the information that you gather?
3. What helps you make a good assessment judgement of students' mathematical understanding/thinking?
4. What evidence do you use to base these assessment judgements on?
5. Do you feel well equipped to make assessment judgements on students' mathematical thinking and their next steps, why/why not?
6. What do you know about how children's knowledge of patterns and relationships progresses through levels 1 to 3 of the curriculum?
7. What sort of activities would children be engaging with at Level 1, 2 and 3 for patterns and relationships?
8. What do you know about how children's knowledge of fractions progresses through levels 1 to 3 of the curriculum?
9. What sort of activities would children be engaging with at Level 1, 2 and 3 for fractions?

F: Interview 2: Intermediate questions

Post judgement questions for teachers:

1. Has planning the Learning Outcomes along a trajectory helped you to notice students developing levels of conceptual understanding of Patterns and Relationships? *Please explain why/why not, and give examples if you can.*
2. How does comparative judgement compare to forms of assessment that you have used in the past?
3. What benefits do you see from comparative judgement?
4. Are there any challenges which you see from comparative judgement?
5. After completing the comparative judgements, do you feel you have a better understanding of the students' mathematical thinking? Why or why not?
6. What steps did you take to choose the better answer?
7. What difficulties did you encounter when comparing the students' mathematical responses?
8. What did you notice about the student responses in relation to patterns and relationships?
    a. What were the strengths?
    b. Were there any common weaknesses or misconceptions?
    c. How will you use this assessment to inform your teaching? (Next steps from here)
9. Are there any ways you would change the implementation of the comparative judgement tasks at [redacted] School to better reflect student understanding? Is there anything we could change before we tackle fractions?

G: Interview 3: Final questions

FINAL QUESTIONS/REFLECTIONS

After the fractions planning/teaching/testing, please consider the following questions:

1. What steps did you take to choose the best response?

2. What did you notice about the student responses in relation to fractions and the tasks presented?
    a) What were the strengths?
    b) Were there any common weaknesses or misconceptions?
    c) How will you use this assessment to inform your teaching/practice? (Next steps from here)

Overall, after both iterations of planning/teaching/testing…

3. Do you feel more able to make consistent and accurate judgements about students learning in these areas? *(Why/why not, please explain/give examples of what was most useful/alternative suggestions).*

4. What are your thoughts on the use of open tasks as a means to assessing student's conceptual mathematical understandings? *(Any benefits/challenges, examples would be great).*

5. What are your thoughts on comparative judging as a means of assessing? *(any benefits/challenges/noticings)*

H: Assessment and teacher judgements – Teacher information sheet and Primary participant consent form – Individual.

Assessment and teacher judgements

TEACHER INFORMATION SHEET

As part of my Masters Research project, I am working with Dr. Jodie Hunter and Prof. Roberta Hunter of the Institute of Education at Massey University. As a school, we are interested to know what factors enable teachers to make accurate assessment judgements in mathematics. We will be focusing on the pattern and relationships part of algebra and on fractions.

I am interested in working with you to investigate the enabling factors that support you to make judgements on students understanding of patterns and relationships, and fractions.

This letter is to formally invite you to participate in the project either fully or partially.

Full participation in the project will involve up to three short interviews (5-10 minutes each) during term one and again in term 2. These will be audio-recorded. No additional time commitments will be required outside of normally arranged staff PD days and syndicate meetings. I would also like to video-tape and observe one or two classroom teaching sessions where the focus is on noticing student thinking.

Partial participation will not involve any interviews, only your attendance and participation in the normally arranged staff PD days and syndicate meetings where the focus will be on assessment and teacher judgements. These meetings will be voice and video recorded to inform my research.

All data (electronic files and copies of children's work) will be stored in a secure location, with no public access and used only for this research. In order to maintain anonymity the school name and names of all children/teachers will be assigned pseudonyms in any publications arising from this research. At the end of the year, a summary of the study will be provided to the school and made available for you to read.

Please note that you have the following rights in response to the request to participate in this study:

• decline to participate;
• in any lesson have the right to ask for the video tape to be turned off at any time;
• withdraw from the study at any point;
• ask any questions about the study at any time during participation;
• provide information on the understanding that your name will not be used unless you give permission to the researcher;
• be given access to a summary of the project findings when it is concluded.

If you have further questions about this project you are welcome to discuss them with me personally:
Megan Kanz: Email: [redacted] Phone: [redacted]

If you have concerns that you feel should be taken to my supervisor you may contact her as follows:
Jodie Hunter: Massey University, Institute of Education. Phone: (09) 4140800 Extension 43518. Email: j.hunter1@massey.ac.nz

This project has been evaluated by peer review and judged to be low risk. Consequently it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research. If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director (Research Ethics), email humanethics@massey.ac.nz. .

Assessment and teacher judgements

CONSENT FORM - INDIVIDUAL

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I agree/do not agree to **participate fully** in this study under the conditions set out in the Information Sheet.

As a full participant, I agree/do not agree to being video-taped during mathematics lessons.

I agree/do not agree to **participate partially** in this study under the conditions set out in the Information Sheet.

Signature: ______________________ **Date:** __________

Full Name - printed ______________________

I: Assessment and teacher judgements – School consent form

Assessment and teacher judgements

SCHOOL CONSENT FORM

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I agree/do not agree to the study taking place at Taiawhiti School (pseudonym)

Signature: .... **Date:** ....

Full Name - printed ....

J: Comparative Judging reliability and results for patterns and relationships

As explained in the methodology, patterns and relationships was assessed by two open type questions [See Section 3.6.3]. Year 1 and 2 children were given the Level 1 to 2 problem and Year 3 and 4 children were given the Level 2 to 3 problem. These problems were judged separately meaning separate percentile levels were obtained for each.

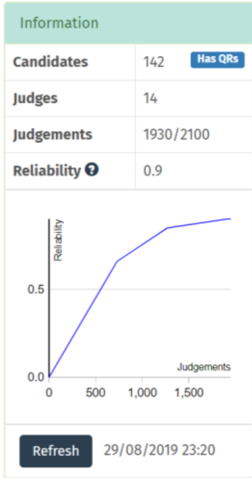
The Level 1 to 2 problem was judged by 12 teachers, 2 senior management staff who had sat in on the professional development day and the researcher making 15 judges in total. All but one judge made between 101 to 150 judgements. One teacher made 18 judges. This teacher had made some technical errors during her first judging round (selecting “left” continuously while trying to get her system to match that of her neighbour’s). This resulted in a number of incorrect judgements that the researcher chose to delete. She was then re-added but did not need to complete all her judgements as the system reported that “enough” judgements had been cast. The infit score for this judge was 1.71 meaning judgements were inconsistent (see table below)

Infit	Local	Mod	Assigned	Median Time	% Left Click	Exclude ...	Created ...	Time Chart ...	Remove ...
1.71	18	0	150	23.4s	55.6%	N	04/04/19		

The researcher therefore chose to run two sets of analytics on the task, one with this judge included and one with this judge excluded. Excluding this judge made no difference to the overall reliability between judges which remained at 0.9 under both conditions. The decision was therefore made to include all judges in the overall analytics of the task.

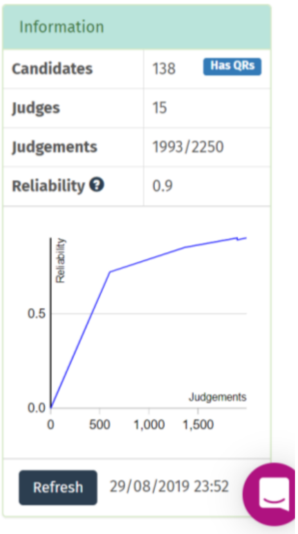
a. 15 Judges (all included)

b. 14 Judges (1 excluded)



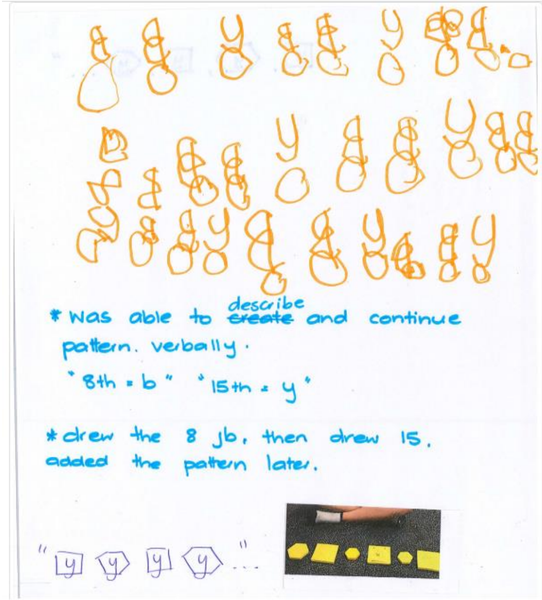
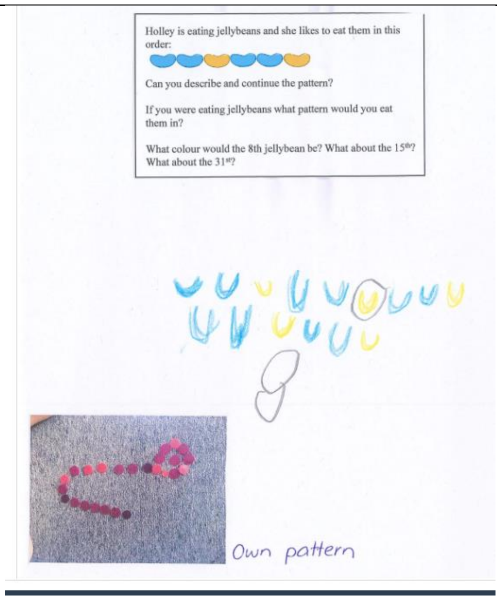
The 0.9% reliability is extremely high showing that even when there is some inconsistency with individual judges, the overall consistency between judges in this instance was good.

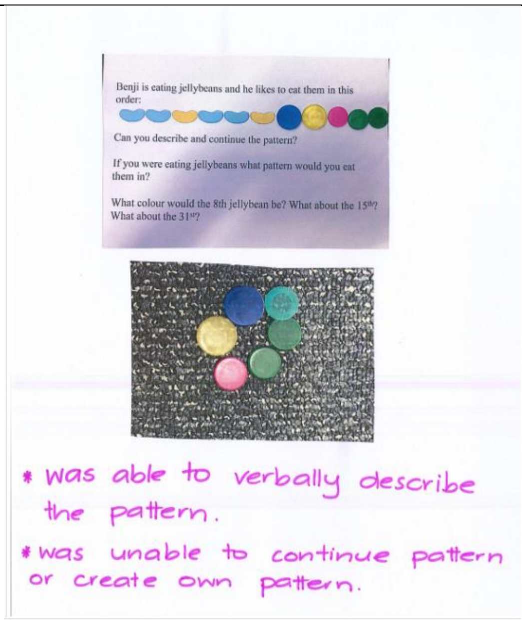
The Level 2 to 3 problem was also judged by 12 teachers, 2 senior management staff who had sat in on the PD day and the researcher making 15 judges in total. All but one judge made between 101 to 150 judgements. The teacher who had made 18 judgements in the first task made 62 in this one. She improved on her consistency achieving an infit score of 0.98. Overall reliability was recorded at 0.9 showing again that the overall reliability between judges was good.



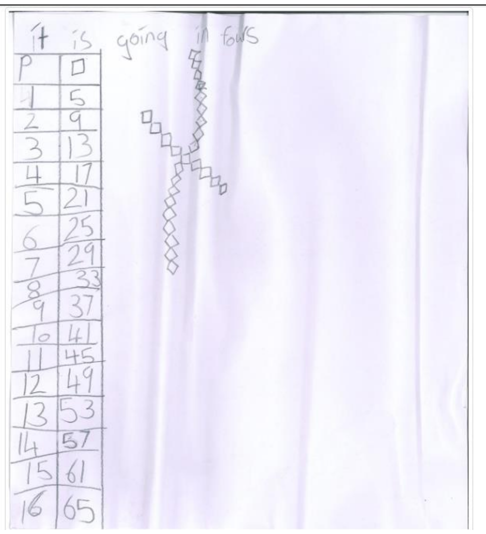
In addition to the high overall reliability achieved during the judging process, task responses at representative percentiles were examined. These are tabulated below.

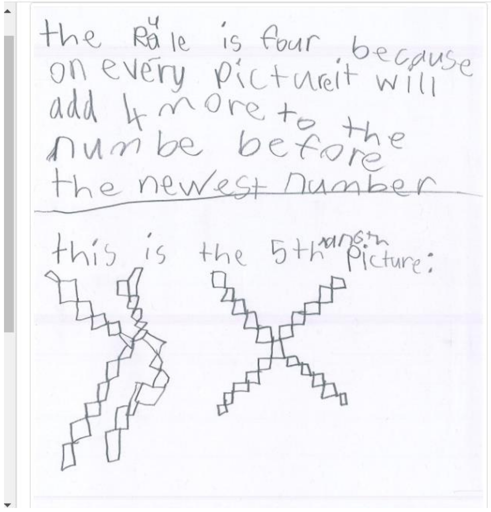
Task 1 – Level 1 to 2 problem

Scaled score (percentile rank between 0-100)	Year group	Response sheet
81	2	
50	1	

24	2	

Task 2, Level 2 to 3 problem.

Scaled score (percentile rank between 0-100)	Year group	Response sheet
75	4	

54	3	the Rule is four because on every picturit will add 4 more to the numbe before the newest number this is the 5th picture:
25	3	The next growing pattern will be 5, because it goes 1, 2, 3, 4, 5

K: Co-constructed plan for patterns and relationships with an example of one problem

Mathematics Number and Algebra **Level One to Three**	Date/Term: Term 1, 2019 Weeks 5-9 (assess W9) Unit Topic: Equations and expressions/ Patterns and relationships

Achievement objectives:

Patterns and relationships

Level One: Create and continue sequential patterns.

Level Two: Find rules for the next member in a sequential pattern.

Level Three: Connect members of sequential patterns with their ordinal position and use tables, graphs, and diagrams to find relationships between successive elements of number and spatial patterns.

Learning Outcomes (children will be able to:)

Level One – Middle goals

- reproduce a pattern using objects, drawings or symbols
- continue patterns and explain the pattern
- invent their own patterns
- communicate the rule for their pattern
- predict a point in a pattern

Level two – Middle goals

- identify the repeating element
- predict the pattern in a given position
- identify an additive rule for a pattern
- use the additive rule to find further terms
- represent a pattern using a table of data

Level three – Middle goals

- identify the repeating element
- use multiplicative thinking to predict a pattern
- recognise either an additive or multiplicative rule
- describe rules in their own words
- use rules to find further terms
- use tables, graphs, diagrams so describe relationships

Jai and Malachi are building snakes with the unifix cubes.

Jai goes red/green/red/green 

Malachi goes Yellow/blue/yellow/blue 

Their snakes are in patterns.

Can you copy their patterns?

Can you make them longer?

What colour will the 6th cube be?

What about the 10th cube?

Can you make your own pattern snake and tell us the pattern?

Megan Kanz
Feb 21, 2019
Resolve

I think this problem would be good to set up on those 20 unit number lines we have. Then you can bring in the ordinal position idea...

Megan Kanz
Feb 21, 2019
Resolve

An extension might be if they notice anything about the colours that go with each number on the number line - all even no's for instance will be the same colour.

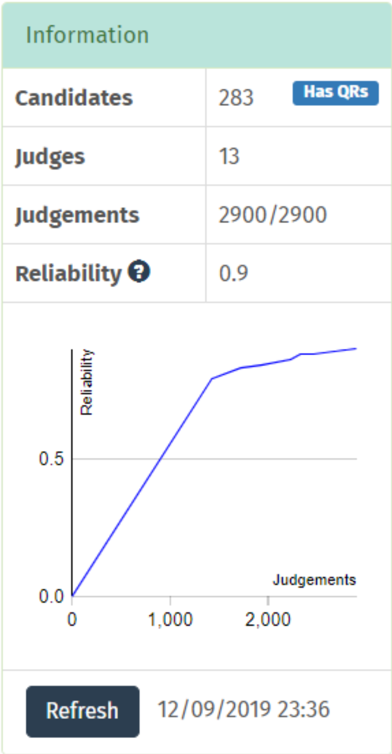
L: Comparative judging results and reliability for fractions tasks

Rather than uploading student answer sheets under separate tasks, all answer sheets were uploaded together. This meant all student responses from year 1 to 4 were judged at the same time. The CJ system was set up to allow for more than one page to be scanned in per participant and judges could scroll down to view all pages associated with a specific student.

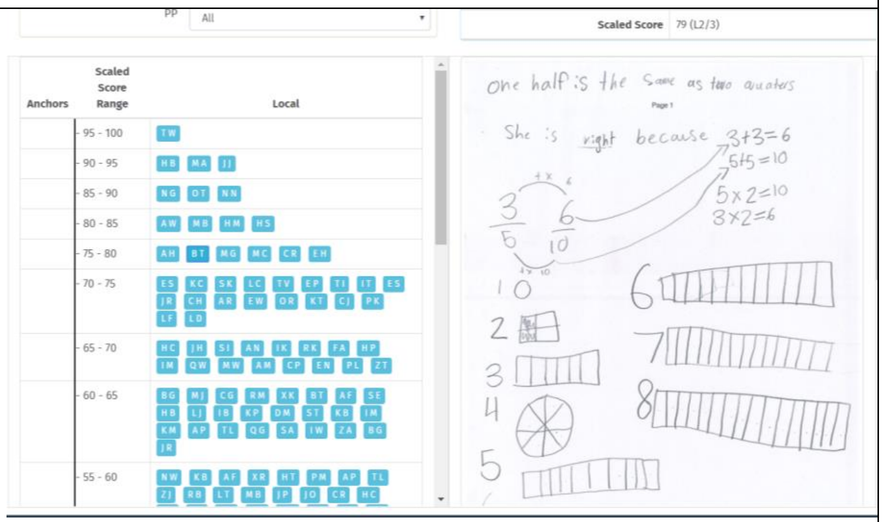
The fractions problem was judged by 10 teachers, 2 senior management staff and the researcher making a total of 12 judges. Two teachers from the first session were unable to attend due to health reasons. This meant that the total number of judgements to be made by teachers was higher than expected. Rather than expect staff to make additional judgements, their required totals were kept at 200 and the researcher made 500 judgements.

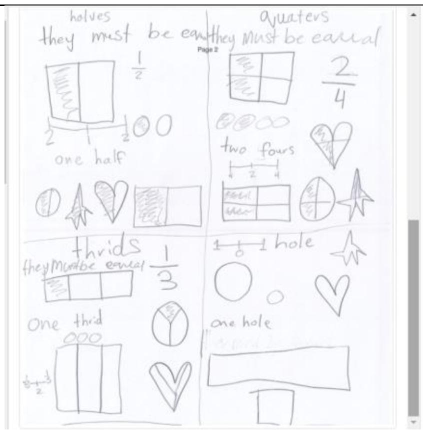
Infit	Local	Mod	Assigned ...	Median Time	% Left Click ...	Exclude ...	Created ...	Time Chart ...	Remove ...
1.17	200	0	200	14.0s	49.0%	N	04/06/19		
1.16	200	0	200	9.0s	44.5%	N	04/06/19		
1.04	200	0	200	21.5s	51.0%	N	04/06/19		
1.03	200	0	200	8.5s	52.0%	N	04/06/19		
1.01	200	0	200	13.8s	44.0%	N	04/06/19		
0.92	200	0	200	12.5s	55.5%	N	04/06/19		
0.75	200	0	200	14.3s	42.5%	N	04/06/19		
0.70	200	0	200	10.2s	46.5%	N	04/06/19		
0.70	200	0	200	10.0s	53.5%	N	04/06/19		
0.67	200	0	200	8.1s	49.5%	N	04/06/19		
0.66	200	0	200	14.7s	44.5%	N	04/06/19		
0.59	500	0	500	11.9s	45.8%	N	03/06/19		
0.52	200	0	200	9.8s	50.0%	N	04/06/19		

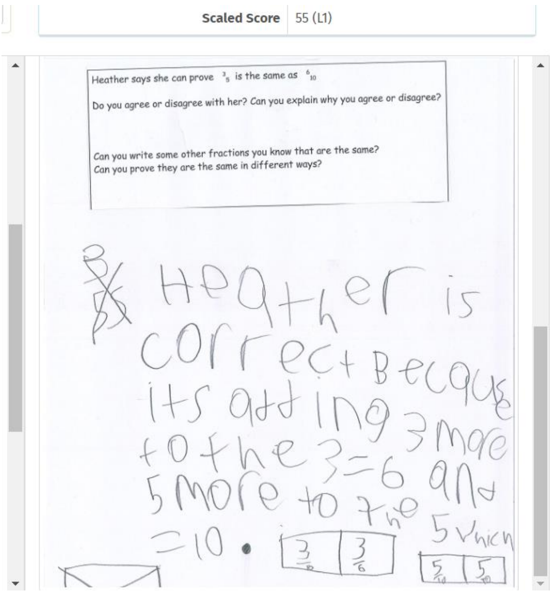
A total of 2900 judgements were made and this produced a reliability of 0.9 showing again that the method was incredibly reliable. This was even with some inconsistency noted with 5 judges receiving individual infit values of between 1.00 and 1.3 (although only 2 were over 1.1).

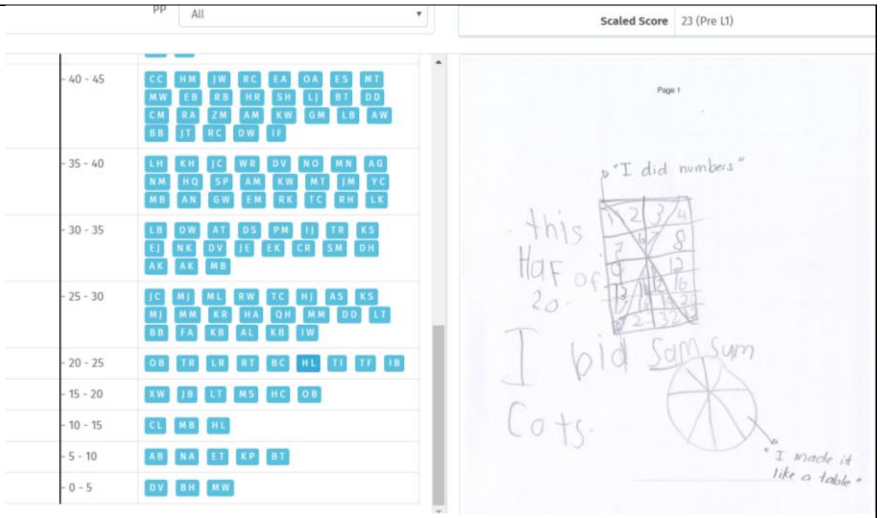


Task examples at representative percentiles were again examined and tabulated below.

Scaled score (percentile rank between 0-100)	Year group	Response sheet
79	4	

<table>
<tr><td></td><td></td><td>
</td></tr>
<tr><td>55</td><td>3</td><td>

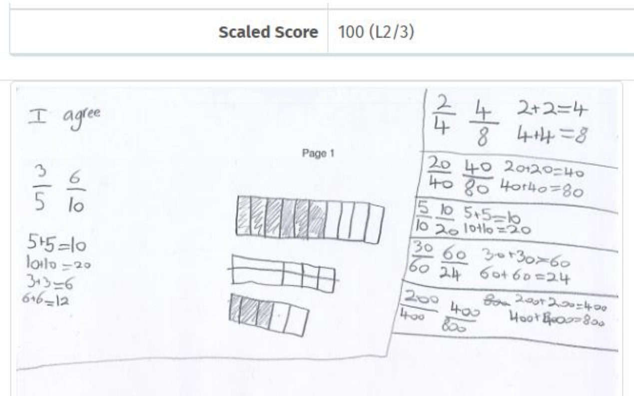

</td></tr>
</table>

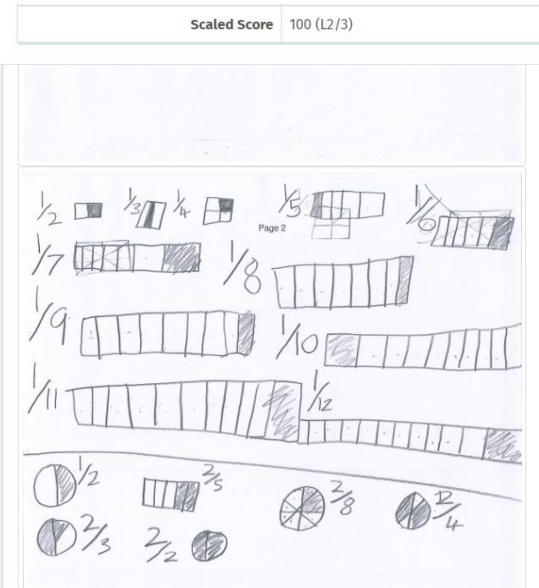
23	Y2	

PP
All
Scaled Score
23 (Pre L1)
40 - 45
CC HM JW RC EA OA ES MT MW EB RB HR SH LJ BT DD CM RA ZM AM KW GM LB AW BB JT RC DW IF
35 - 40
LH KH JC WR DV NO MN AG NM HQ SP AM KW MT JM YC MB AN GW EM RK TC RH LK
30 - 35
LB DW AT DS PM IJ TR KS EJ NK DV JE EK CR SM DH AK AK MB
25 - 30
JC MJ ML RW TC HJ AS KS MJ MM KR HA QH MM DD LT BB FA KB AL KB IW
20 - 25
OB TR LR RT BC HL TI TF IB
15 - 20
XW JB LT MS HC OB
10 - 15
CL MB HL
5 - 10
AB NA ET KP BT
0 - 5
DV BH MW
Page 1
"I did numbers"
this Haf of 20.
I bid Som Sum Cots.
"I made it like a table"

M: Transcripts from the 100th and 82nd percentile

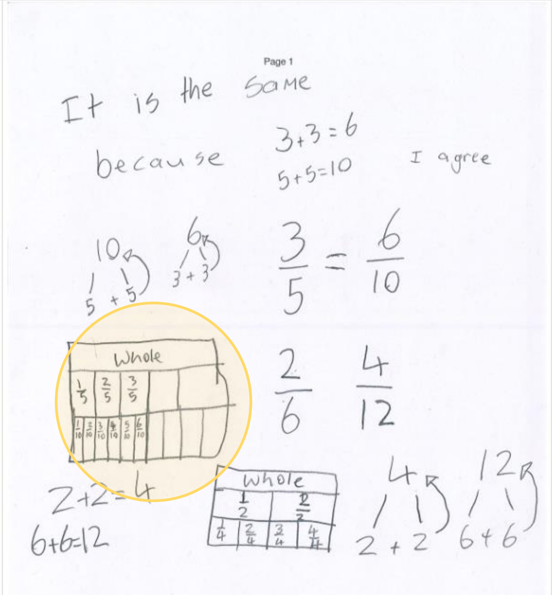
Transcript from the 100th percentile





Transcript from the 82nd percentile highlighting how the student has shown the equivalent fractions relative to the same sized whole and generalised this understanding to another situation.

Scaled Score	82 (L2/3)



N: Co-constructed plan for fractions with two problem examples

Mathematics Fractions **Level One to Three**	Date/Term: Term 2 2019 Weeks 1-5 Unit Topic: Fractions

Achievement objectives:

Level One: Use a range of counting, grouping, and equal-sharing strategies with whole numbers and fractions.

Level Two: Know simple fractions in everyday use. Use simple additive strategies with whole numbers and fractions.

Level Three: Know fractions and percentages in everyday use. Use a range of additive and simple multiplicative strategies with whole numbers, fractions, decimals, and percentages

Big ideas to link back to: (From NZ Maths)

- The denominator tells the number of equal parts into which a whole is divided. The numerator specifies the number of these parts being counted.
- The size of the fractional amount depends on the size of the whole.
- When working with fractions, the whole needs to be clearly identified.
- Equivalent fractions have the same value.
- When adding fractions, the units need to be the same because the answer can only have one denominator.
- The more pieces a whole is divided into, the smaller each piece will be.
- The key to proportional thinking is being able to see combinations of factors within numbers.
- A fraction can represent more than one whole.
- Division is the opposite of multiplication.
- The relationship between multiplication and division can be used to help simplify the solution to problems involving the division of fractions.

Learning Outcomes (children will be able to…)

Level One – Middle goals *(start with halves and quarters)*

- Use equal sharing to solve division and fractions of sets problems without counting every object
- Show that fractions are equal parts of a whole
- Find halves and quarters (of a measure/continuous model e.g. length/region of a shape and, discrete model e.g. set of a group)
- Relate symbols and words with models of fractions
- Order unit fractions (1/4, 1/3, ½, etc) or fractions with the same denominator (1/4, 2/4, ¾)

Level two – Middle goals *(halves, thirds, quarters, fifths, sixths, eighths, tenths)*

- Use additive thinking to find fractions of sets
- Understand how fractions can be said and written in numerals and words
- Understand the order and size of fractions with common denominators including improper fractions (1/4, 2/4, ¾, 5/4)
- Order unit fractions (1/4, 1/3, ½, etc)

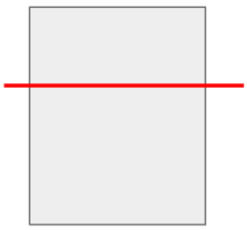
- Locate fractions on a number line
- Find equivalent fractions and justify why they are equivalent with materials/drawings/number.

Level three – Middle goals

- Use multiplicative thinking to find fractions of quantities e.g. 2/3 of 12
- Use multiplicative thinking to find simple equivalent fractions e.g. doubling or halving
- Add and subtract fractions with the same denominators
- Understand the meaning of the digits in a fraction
- Understand how fractions can be said and written in numerals and words
- Understand the order and size of fractions with common denominators or common numerators (e.g. 2/7<3/9<3/5)
- Find equivalent fractions and justify why they are equivalent with materials/drawings/numbers

First problem example:

James said he will give Bill half of his sandwich. He cuts his sandwich like this:



Has he cut the sandwich in half?
Can you draw how he could cut it properly in half so it is fair?

Second problem example:

Harmonee had a birthday party to go to. The party started at 3:30pm. At the moment it is 1:15pm.

How long does Harmonee have to get ready for the party?

If she was a 1/4 of an hour late to the party, what time did she get there?

What if she was 3/4 of an hour late to the party? What time would that have been?

Hazel had a birthday party to go to. The party started at 3pm. At the moment it is 1pm.

How long does Harmonee have to get ready for the party?

If she was a 1/4 of an hour late to the party, what time did she get there?

What if she was 3/4 of an hour late to the party? What time would that have been?

References

Adie, L., Klenowski, V., & Wyatt-Smith, C. (2012, 05//). Towards an understanding of teacher judgement in the context of social moderation [Article]. *Educational Review, 64*(2), 223-240. https://doi.org/10.1080/00131911.2011.598919

Anderson, D., & Pritchard, R. (2010). Bridging the gap: Challenging conceptions of fractions and decimals. In R. Averill & R. Harvey (Eds.), *Teaching primary school mathematics and statistics: Evidence-based practice*. NZCER Press.

Bakker, A. (2018). *Design research in education : a practical guide for early career researchers* (First edition ed.). Routledge, an imprint of Taylor and Francis.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407.

Berube, C. T. (2004). Are standards preventing good teaching? *Clearing House, 77*(6), 264.

Bishop, R. (2005). Freeing ourselves from neocolonial domination in research: a kaupapa Maori approach to creating knoledge. In N. Denzin & Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3 ed., pp. 109-138). SAGE Publications.

Bogdan, R., & Biklen, S. K. (2006). *Qualitative research for education : an introduction to theory and methods* (5th ed ed.). Pearson A & B.

Bonne, L. (2017). National standards : how are they playing out in schools now? *New Zealand principal, 32*(1), 17-18.

Braithwaite, D. W., Tian, J., & Siegler, R. S. (2018). Do children understand fraction addition? [Article]. *Developmental Science, 21*(4), 1-1. https://doi.org/10.1111/desc.12601

Brookhart, S. M. (2013). The Use of Teacher Judgement for Summative Assessment in the USA. *Assessment in Education: Principles, Policy & Practice, 20*(1), 69-90. http://dx.doi.org/10.1080/0969594X.2012.703170

Brown, G., & Quinn, R. (2007). Fraction proficiency and success in algebra. What does research say? *Australian Mathematics Teacher, 63*(3), 23-30.

Campbell, T. (2014). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment [Review]. *Journal of Social Policy, 44*(3), 517-547. https://doi.org/10.1017/S0047279415000227

Carmel, R. C. (2005). Big ideas and understandings as a foundation for elementary and middle school mathematics. *Journal of Mathematics Education Leadership, 7*(3), 9-24.

Charmaz, K. (2006). *Constructing grounded theory : a practical guide through qualitative analysis* [Non-fiction]. SAGE.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE.

Clements, D. H. (2011). Learning trajectories: Foundations for effective, research-based education. In L. R. Weist & T. Lamberg (Eds.), *Proceedings of the 33rd annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education* (pp. 365-372). University of Nevada.

Cobb, P., Confrey, J., Disessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research [Article]. *Educational Researcher, 32*(1), 9-13. https://doi.org/10.3102/0013189X032001009

Creswell, J. W. (2002). *Research design: Qualitative, quantitative and mixed methods approaches*. Sage Publications.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics [Review Article]. *Developmental Review, 34*, 344-377. https://doi.org/10.1016/j.dr.2014.10.001

Depaepe, F., Torbeyns, J., Vermeersch, N., Janssens, D., Janssen, R., Kelchtermans, G., Verschaffel, L., & Van Dooren, W. (2015). Teachers' content and pedagogical content knowledge on rational numbers: A comparison of prospective elementary and lower secondary school teachers [Article]. *Teaching and Teacher Education, 47*, 82-92. https://doi.org/10.1016/j.tate.2014.12.009

[Record #242 is using a reference type undefined in this output style.]

Glasser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Hathaway, D., & Norton, P. (2018). *Understanding problems of practice : a case study in design research*. Springer.

Hattie, J. (2008). *Visible learning : a synthesis of over 800 meta-analyses relating to achievement*. London ; New York : Routledge, 2009.

Heldsinger, S., & Humphry, S. (2010). Using the Method of Pairwise Comparison to Obtain Reliable Teacher Assessments. *Australian Educational Researcher, 37*(2), 1-19.

Henningsen, M., & Stein, M. K. (1997). Mathematical Tasks and Student Cognition: Classroom-Based Factors that upport and Inhibit High-Level Mathematical Thinking and Reasoning. *Journal for Research in Mathematics Education, 28*(5), 524-549.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning : A project of the national council of teachers of mathematics.* (pp. 371-404). Information Age.

Hodge, L. L., & Walther, A. (2017). Building a discourse community: Initial practices. *Mathematics Teaching in the Middle School, 22*(7), 430-437.

Holt Wilson, P. (2014). Learning trajectories and professional development. In K. H. Nguyen, J. Confrey, & A. P. Maloney (Eds.), *Learning over time : Learning trajectories in mathematics education* (pp. 227-241). Information Age Publishing.

Humphrey, C. (2013). A Paradigmatic Map of Professional Education Research [Article]. *Social work education*(1), 3.

Hunter, J., & Jones, I. (2018). Free-response tasks in primary mathematics: A window on students' thinking.*Making waves, opening spaces* 41st annual conference of the Mathematics education Research Group of Australasia, Auckland.

Hunter, R. (2008). Facilitating communities of mathematical inquiry. *New Zealand Mathematics Magazine, 45*(2), 1-13.

Jacobs, V. R., Lamb, L. L. C., & Phillip, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal of Research in Mathematics Education, 41*(2), 169-202.

Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: can comparative judgement help? *Educational Studies in Mathematics, 89*(3), 337-355. https://doi.org/10.1007/s10649-015-9607-1

Jones, I., & Karadeniz, I. (2016). An alternative approach to assessing achievement. Proceedings of the 2016 40th Conference of the International Group for the Psychology of Mathematics Education, Szeged, Hungary.

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing mathematical problem solving using comparative judgement. *International Journal of Science & Mathematics Education, 13*(1), 151-177. https://doi.org/10.1007/s10763-013-9497-6

Kabiri, M. S., & Smith, N. L. (2003). Turning traditional textbook problems into open-ended problems [Article]. *Mathematics Teaching in the Middle School, 9*(3), 186.

Kaput, J. J. (2000). *Transforming algebra from an engine of inequity to an engine of mathematical power by "algebrafying" the K-12 curriculum [microform] / James J. Kaput*. [Washington, DC] : U.S. Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Center, 2000.

Kelly, A. E., Baek, J. Y., Lesh, R. A., & Bannan-Ritland, B. (2008). Enabling innovations in education and systematizing their impact. In A. E. Kelly, R. A. Lesh, & J. Baek (Eds.), *Handbook of design research methods in innovation: Innovations in science, technology, engineering and mathematics learning and teaching* (pp. 3-19). Routledge.

Knuth, E., Stephens, A., Blanton, M., & Gardiner, A. (2016). Build an early foundation for algebra success [Article]. *Phi Delta Kappan, 97*(6), 65. https://doi.org/10.1177/0031721716636877

Krefting, L. (1991). Rigor in qualitative research: the assessment of trustworthiness. *American Journal of Occupational Therapy, 45*(3), 214-222.

Kretchmar, J. (2013). *Constructivism*. Salem Press Encyclopedia,. http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ers&AN=89164134&site=eds-live&scope=site

Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). SAGE.

Lamon, S. (2007). Rational numbers and proportional reasoning : Toward a theoretical framework for research. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning [electronic resource] : a project of the National Council of Teachers of Mathematics* (pp. 629-667). Information Age.

Land, T. J., & Drake, C. (2014). Enhancing and enacting curricular progressions in elementary mathematics. *Mathematical Thinking and Learning: An International Journal, 16*(2), 109-134. https://doi.org/http://dx.doi.org/10.1080/10986065.2014.889502

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage Publications.

Massey University. (2015). *Code of ethical conduct for research, teaching and evaluations involving human participants* http://www.massey.ac.nz/massey/fms/Human%20Ethics/Documents/MUHEC%20Code%202015.pdf?CA6CFFFAA5150271FD193A0B0C56C5CF

Massey University. (2017). *Code of ethical conduct for research, teaching and evaluations involving human participants, REVISED CODE* http://www.massey.ac.nz/massey/fms/Human%20Ethics/Documents/Revised%20Code.pdf?BBFDDEDDA9C2DBF2BAAC9C6951C58ED4

Mathison, S. (1988). Why Triangulate? *Educational Researcher, 17*(2), 13-17. https://doi.org/10.3102/0013189X017002013

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education, 65*, 48-60. https://doi.org/10.1016/j.tate.2017.02.021

Merriam, S. B. (1998). *Qualitative research and case study applications in education* (Rev. and expanded ed.). Jossey-Bass.

Merriam, S. B., & Grenier, R. S. (2019). *Qualitative research in practice : examples for discussion and analysis* (Second edition ed.). Jossey-Bass, a Wiley Brand.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis : a methods sourcebook* (Third edition ed.). SAGE Publications, Inc.

Ministry of Education. (2013). Global strategy stage: GLoSS assessments. Interview forms 1-4. Additional Information. Retrieved 4/7/2019, from https://nzmaths.co.nz/sites/default/files/Numeracy/GloSS/GloSS_Additional_Information.pdf

Ministry of Education. (2014). Junior Assessment of Mathematics (JAM) teachers guide. Retrieved 4/7/2019, from https://nzmaths.co.nz/sites/default/files/jam/JAM-Teacher-Guide.pdf

Ministry of Education. (2015). *The New Zealand Curriculum, for English-medium teaching and learning in years 1-13*

Learning Media Limited.

Ministry of Education. (2019). *Making an overall teacher judgment*. Retrieved 2/7/2019 from https://assessment.tki.org.nz/Overall-teacher-judgment/Making-an-overall-teacher-judgment

Mueller, M., Yankelewitz, D., & Maher, C. (2011). Sense making as motivation in doing mathematics: Results from two studies. *The Mathematics Educator, 20*(2), 33-43.

[Record #186 is using a reference type undefined in this output style.]

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2019). Standards for mathematical practice. In *Common core state standards,* . National Governors Association Center for Best Practices, Council of Chief State School Officers,. http://www.corestandards.org/about-the-standards/contact/

Neuman, D. (2014). Qualitative Research in Educational Communications and Technology: A Brief Introduction to Principles and Procedures. *Journal of Computing in Higher Education, 26*(1), 69-86. https://doi.org/http://dx.doi.org/10.1007/s12528-014-9078-x

NSW Department of Education. (2019). *Consistent teacher judgement*. Retrieved 20/03/2019 from https://education.nsw.gov.au/teaching-and-learning/curriculum/key-learning-areas/pdhpe/stages-4-and-5/programming#Consistent6

Punch, K., & Oancea, A. (2014). *Introduction to research methods in education* (2nd ed.). London ; Thousand Oaks, California : SAGE Publications, 2014.

Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 769-802). Sage Publications.

Sadler, R. D. (1998). Formative assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice, 5*(1), 77-84. https://sigmapubs.onlinelibrary.wiley.com/doi/abs/10.1111/wvn.12239

Sarama, J., Clements, D. H., & Spitler, M. E. (2017). Evidence of teacher change after participating in TRIAD's learning trajectories-based professional development and implementing learning trajectory-based mathematics instruction. *Mathematics Teacher Education and Development, 19*(3), 58-75.

Schukajlow, S., & Krug, A. (2014). Do multiple solutions matter? Prompting multiple solutions, interest, competence and autonomy. *Journal for Research in Mathematics Education, 45*(4), 497-533.

Selling, S. K. (2016). Making mathematical practices explicit in urban middle and high school mathematics classrooms [research-article]. *Journal for Research in Mathematics Education*(5), 505. https://doi.org/10.5951/jresematheduc.47.5.0505

Simon, M., & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning: An elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning, 6*(2), 91-104.

Smaill, E. (2013). Moderating New Zealand's National Standards: teacher learning and assessment outcomes. *Assessment in Education: Principles, Policy & Practice, 20*(3), 250.

Small, M. (2010). *Big ideas from Dr. Small. Creating a comfort zone for teaching mathematics. Grades K-3*. Nelson Educaion Ltd.

Smith, L. T. (2012). *Decolonizing methodologies. [electronic resource] : research and indigenous peoples* (2nd ed.). Zed Books

Smith, M. S., & Stein, M. K. (1998). Reflections on Practise: Selecting and Creating Mathematical Tasks: From Research to Practise. *Mathematics Teaching in the Middle School, 3*(5), 344-350.

Sole, M. A. (2018). Open-ended questions: A critical class component. *Mathematics Teacher 111*(6), 462-465.

Sullivan, P., Borcek, C., Walker, N., & Rennie, M. (2016). Exploring a structure for mathematics lessons that initiate learning by activating cognition on challenging tasks. *Journal of Mathematical Behavior, 41*, 159-170. https://doi.org/10.1016/j.jmathb.2015.12.002

Sullivan, P., Clarke, D., & Clarke, B. (2009). Converting mathematics tasks to learning opportunities: An important aspect of knowledge for mathematics teaching. *Mathematics Education Research Journal, 21*(1), 85-105.

Sullivan, P., Mousley, J., & Zevenbergen, R. (2006). Teacher actions to maximize mathematics learning opportunities in heterogeneous classrooms. *International Journal of Science & Mathematics Education, 4*, 117-143.

Superfine, A. C., & Wenjuan Li. (2017). Characterizing teachers' informal conceptions of learning trajectories in mathematics [Article]. *Conference Papers -- Psychology of Mathematics & Education of North America*, 1259-1259.

Sztajn, P., Confrey, J., Holt Wilson, P., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher, 41*(5), 147-156.

Taylor, S. J., Bogdan, R., & DeVault, M. L. (2016). *Introduction to qualitative research methods : a guidebook and resource* (4th edition ed.). Wiley.

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

The New Zealand Council for Educational Research. (2019). *Progressive Achievement Tests (PATs)*. NZCER. Retrieved 4/7/2019 from https://www.nzcer.org.nz/tests/pats

Thurstone, L. (1927). A law of Comparative Judgement. *Psychology Review, 34*, 273-286.

Varygiannes, D. (2013/2014). The impact of open-ended tasks. *Teaching Children Mathematics, 20*(5), 277-280.

Warshauer, H. K. (2015). Strategies to support productive struggle. *Mathematics Teaching in the Middle School, 20*(7), 390.

Wilson, S. (2008). *Research is ceremony: indigenous research methods*. Fernwood.

Woodbury, S. (2000). Teaching toward the Big Ideas of Algebra. *Mathematics Teaching in the Middle School, 6*(4), 226-231.

Yanow, D., & Schwartz-Shea, P. (2015). *Interpretation and method : empirical research methods and the interpretive turn* (2nd ed.). Routledge.