Addressing inequity and underachievement: Intervening to improve middle leaders' problem-solving conversations

Jacqueline Margaret Patuawa¹ · Claire Sinnema¹ · Viviane Robinson¹ · Tong Zhu¹

Accepted: 10 January 2022
© The Author(s) 2022

Abstract

Reducing inequity is the moral imperative confronting today's educational leaders. Central to reducing inequity is leaders' ability to solve the school-based problems that contribute to it, while building the positive and trusting professional relationships required for teachers to commit to the hard work of improvement. A theory of collaborative problem-solving informed our intervention designed to improve the effectiveness of leaders' behaviour as they worked with their teachers to accelerate the reading achievement of students yet to reach age-related standards. A concurrent mixed methods design was used to evaluate the impact of the intervention by analysing transcripts of interviews and leader-teacher conversations and student reading achievement data. Leaders' effectiveness in their conversations improved significantly as did the reading outcomes of their target students. Our findings suggest that even short interventions grounded in strong theory with appropriate learning opportunities can affect a positive change in leadership behaviour and student outcomes.

Keywords Collaborative problem solving · School improvement · Middle leaders · Reading achievement · Inequity in student outcomes

Introduction

Despite governments significantly investing in improving literacy achievement, many jurisdictions struggle to produce equitable outcomes for all learners (Schleicher, 2019). School leaders, and increasingly middle leaders, are tasked with raising achievement and reducing disparity. Middle leaders, in particular, are increasingly responsible for the implementation of their school's improvement agenda because they, more than any leader, work closer with teachers to solve the problems that perpetuate this disparity (Bassett, 2018; De Nobile, 2018, 2019; Harris & Jones,

✉ Jacqueline Margaret Patuawa
j.patuawa@auckland.ac.nz

¹ Faculty of Education and Social Work, The University of Auckland, Auckland, New Zealand

2017; Harris et al., 2019; Patuawa, Robinson, Sinnema, & Zhu 2021). Problem solving is largely exercised socially, and requires skilled and confident middle leaders to engage all stakeholders in respectful conversations about issues affecting the quality of teaching and learning (Argyris & Schön, 1974; Robinson & Le Fevre, 2011; Robinson, Meyer, Le Fevre, & Sinnema 2020; Timperley, 2015). There is, however, currently little empirical research investigating middle leaders' capabilities for collaborative problem solving (CPS), and even less research that describes interventions and high-quality tools to improve and measure leaders' problem-solving capabilities (Brenninkmeyer & Spillane, 2008; Graesser et al., 2018).

This article reports phase 2 of a 3-phase intervention research programme investigating the CPS behaviour of middle leaders. Phase 1 researched the effectiveness of middle leaders' CPS behaviour pre-intervention (Patuawa et al., 2021). In this article, we report the impact of two interventions on those same leaders' problem-solving behaviours, as they worked with their teachers to accelerate achievement for students yet to meet age-related reading standards. We investigated the following questions:

1. What is the reported impact of the intervention on leaders' behaviour in problem-solving conversations about reading achievement problems?
2. What is the reported impact of the intervention on problem outcomes?
3. What is the assessed impact of the intervention for the reading achievement of the target students?

The importance of problem solving

Problem solving is increasingly recognised as central to the work of leaders (Brenninkmeyer & Spillane, 2008; Graesser et al., 2018; Kin & Kareem, 2019; Mintrop & Zumpe, 2019; Mumford et al., 2000; Robinson, et al., 2020). Problem-solving skills are linked to efficiency, effectiveness, and improvement; however, much time, energy, and resources are poorly spent in education systems and schools because of poor problem solving (Argyris, 1991; Graesser et al., 2018; Mintrop & Zumpe, 2019; Robinson, et al., 2020; Weber & Khademian, 2008). For example, Hattie (2015) suggests that many policymakers struggle to have the uncomfortable conversations required to address problems with the effectiveness of teaching, but instead choose to focus reform efforts on introducing politically attractive policies (i.e., structural changes such as reducing class sizes, delivering more technology, and new building designs) that have minimal impact on improving student learning.

Leaders' skills in problem solving

Two theories of problem solving (deliberative and intuitive) are most salient in the literature. The deliberative theory describes leaders progressing through a sequence of stages including problem identification (precise description of the problem, its seriousness, and whether there is sufficient demand for it to be solved); causal inquiry (surfacing and testing assumptions about cause); and

designing solutions (ensuring that the proposed solution strategies explicitly address the identified causes) (Leithwood & Stager, 1989; Mintrop & Zumpe, 2019; Robinson et al., 2020). The intuitive theory describes leaders taking swift action to iteratively trial a range of solutions until they positively impact the problem. In pursuit of the right solution, leaders rely on their intuition, knowledge of what has and has not worked previously, and their prior learning and experience without rigorously testing the quality of their original thinking (Mintrop & Zumpe, 2019; Patuawa et al., 2021). Research suggests that leaders typically default to an intuitive approach in their problem-solving efforts (Mintrop & Zumpe, 2019; Patuawa et al., 2021; Robinson et al., 2020). While this approach works, there are a range of problematic patterns (especially when solving more complex problems) identified in descriptive studies. These problematic patterns include leaders commonly assuming agreement about the existence and seriousness of the problem; using general and imprecise language when raising the issue (Cardno, 1998, 2007; Mintrop & Zumpe, 2019; Patuawa et al., 2021; Sinnema, Le Fevre, & Robinson, 2013; Sinnema, Ludlow, & Robinson 2016; Timperley, 2015); and rarely engaging in causal inquiry, precluding them from critically evaluating the quality and alignment of proposed solutions in relation to the problem and its causes (Leithwood & Stager, 1989; Marcy & Mumford, 2010; Mintrop & Zumpe, 2019; Mumford et al., 2000; Patuawa et al., 2021; Robinson et al., 2020).

While intervention studies investigating the improvement of leaders' problem-solving behaviours are rare, there is evidence that interventions can be effective (Leithwood & Steinbach, 1992; Mintrop & Zumpe, 2019; Robinson et al., 2020). For example, drawing on their research into the problem-solving behaviour of expert and typical leaders, Leithwood and Steinbach (1992) used an experimental design to test whether problem solving could be taught. Leithwood and Steinbach administered a pre-test to a control and an experimental group of school leaders, prior to executing a 4-day intervention that explicitly taught the experimental group a multi-component deliberative model of problem solving. The experimental group's post-test results showed significantly improved problem-solving skills compared to the control group. Similarly, a study by Marcy and Mumford (2010) found that explicit teaching of causal analysis significantly improved the performance of undergraduates who assumed the role of a university president in charge of improving academic results.

A more recent study reporting on the thinking of three leaders (from a group of nine) as they attempted to solve complex problems in their schools over two years, showed the difficulty of helping leaders switch from more intuitive to more deliberative modes of problem solving. Data drawn from observations, interviews, coaching, and document analysis revealed that despite being encouraged and prompted by their coaches to test assumptions about their theories for improvement and think more deeply about alternatives, the leaders could not transfer their theoretical thinking into practice and reverted to their original framing of problems as preferred solutions. In not testing their existing theories they acted intuitively, missed attention to causal inquiry, and largely bypassed others' interests, instead treating them as "vessels to be filled" (Mintrop & Zumpe, 2019).

Many leaders need support with the social dimension of CPS, in part because they experience a task relationship dilemma in many problem-solving conversations (Argyris, 1991; Argyris & Schön, 1974, 1996; Patuawa et al., 2021; Robinson & Le Fevre, 2011; Sinnema, Le Fevre et al., 2013; Schwarz, 2017; Sinnema et al., 2015). This dilemma arises when leaders are faced with having to discuss issues of teaching practice; for example, they feel conflicted between their need to address the practice and their desire to avoid upsetting the teacher. Faced with this dilemma, leaders often unconsciously prioritise either the task or the relationship, neither of which is conducive to effective problem solving (Argyris & Schön, 1974; Schwarz, 2017; Robinson, 2011; Robinson et al., 2020). The dilemma is explained by three control-focused motivations that most people default to when faced with disagreement, threat, or embarrassment (Argyris & Schön, 1974; Robinson, 2011; Schwarz, 2017). These three motivations are: (1) a belief that I am right, and others are wrong; (2) a need to unilaterally control the process for getting what I want; and (3) a need to unilaterally protect myself and others from embarrassment or hurt. These control-focused motivations explain the dilemma between progressing the task and maintaining relationships, because the “I am right and you’re wrong” stance erodes the trust that leaders need to work productively with teachers to resolve teaching and learning problems (Argyris & Schön, 1974, 1996; Robinson, 2011). If middle leaders are to successfully assist their teachers to raise achievement and reduce disparity, more deliberative problem-solving interventions are needed to enable leaders to avoid the task relationship dilemma by building trust with teachers while systematically investigating the causes of, and solutions to, teaching and learning problems (Argyris & Schön, 1974).

For problem-solving interventions to be effective, they need to simultaneously address leaders’ interpersonal effectiveness and procedural knowledge, be research informed, engage and build on existing practice, bridge the world of theory and practice to ensure learned skills can be applied to the work, and be supported by multiple opportunities to learn with ongoing intervention and feedback (Ackerman et al., 2011; Bransford et al., 2000; Cavanagh et al., 2014; Day & Dragoni, 2015; Mourshed, Chijioke, & Barber 2010; Spillane et al., 2002; Timperley, Wilson, Barrar, & Fung, 2007).

Our model of CPS

Our model of CPS (see Fig. 1) integrates a theory of interpersonal effectiveness (portrayed in the central box) with a theory of deliberative problem solving (portrayed in the outer circle) to avoid the task/relationship dilemma. It is collaborative because it requires deep engagement with, rather than bypassing, others’ interests (Robinson, 2018).

Our theory of interpersonal effectiveness

Our theory of interpersonal effectiveness (a revision and elaboration of Argyris and Schön’s Model 2 theory) was used to evaluate the quality of the leaders’

Fig. 1 Our model of collaborative problem-solving

CPS conversations during *Leading by Learning* (LbL). LbL is a theory of ethical instructional leadership that enables leaders to build trust while making significant progress on problems encountered in their work. LbL develops learning-focused (as opposed to control-focused) thinking and action through the development of leaders who are open-minded and thus motivated to seek valid information, be respectful, and create internal commitment (Argyris & Schön, 1974; Robinson & Le Fevre, 2011; Schwarz, 2017; Timperley, 2015).

LbL motivations

When motivated to seek valid information, leaders advocate their views about the causes of, and possible solutions to a problem in ways that make their thinking public and testable. Leaders speak in concrete, rather than abstract language, give reasons for their views, and provide frequent examples and illustrations to make their meaning clear (Argyris & Schön, 1974; Robinson, 2011; Robinson, Sinnema & Le Fevre, 2014; Schwarz, 2017). Their open-minded desire to test their views requires acceptance that their ideas are fallible and that they need to be critically receptive to the alternative ideas of others. Open-mindedness requires the acknowledgement and suspension of prejudgement and personal biases (Hare, 2003; Schwarz, 2017). Leaders who seek valid information are also motivated to genuinely inquire into their teachers' beliefs, particularly if their views differ. Validity increases when differences between leaders' and teachers' theories are treated as opportunities to learn, rather than opportunities to persuade the other of the superiority of one's views (Argyris & Schön, 1974; Le Fevre, Robinson & Sinnema 2014; Robinson et al., 2014; Sinnema, Meyer, Le Fevre, Chalmers & Robinson, (2021)). The ability to be open-minded encourages and supports open-communication, the free flow of communication necessary for informed decision making, and a psychologically safe environment in which to commit to the hard work of improvement (De Nobile, 2010).

To be respectful requires leaders to genuinely engage with others by giving equal consideration to both their own and others' contributions. Leaders recognise their views may be fallible and are committed to being respectfully truthful in sharing them, while at the same time conveying they are open to others' differing views. They listen deeply to their teachers' beliefs, and demonstrate this by providing accurate summaries of what they have heard (Argyris & Schön, 1974; Robinson 2011).

To create internal commitment, leaders build mutual accountability for the choices made, rather than merely seeking compliance. Leaders probe the reasons for doubts and disagreements so they can be understood and resolved, rather than brushed aside. When leaders and teachers become internally committed to a course of action, they follow through and monitor its implementation and outcomes (Argyris & Schön, 1974; Robinson, 2011).

LbL behaviours

Leaders' motivations are revealed by the statements they make (advocacy), the questions they ask (inquiry), and their ability to demonstrate their accurate understanding of others' views (listening). Learning-focused leaders respectfully advocate their own views, invite others to do similarly, and make it safe for doubts and disagreements to be raised. Learning-focused leaders genuinely inquire to learn about others' views and seek reaction to their own ideas; actively listen to deeply understand all perspectives prior to rushing to solve a problem; listen to identify untested assumptions, create pauses to allow thinking, promote critical evaluation of the quality of

thinking, and provide accurate summaries of important points; and are committed to high-quality solutions that integrate the needs of all involved (Argyris & Schön, 1974; Robinson, 2011).

Our theory of deliberative problem solving

Our theory of problem solving is deliberative, because the inequitable reading achievement problems experienced by learners are longstanding and previous attempts to solve them have repeatedly not yielded desired results (Mintrop & Zumpe, 2019; Robinson et al., 2020). Our theory comprises four dimensions, three of which have been revised and adapted from Leithwood and Stager's (1989) multi-component problem-solving model. Dimension 1 (problem identification [PI]) encompasses the interpretation component from Leithwood and Stager's (1989) model; dimension 2 (problem cause [PC]), encompasses the constraint component; dimension 3 (problem solutions [PS]) encompasses the goals and solution strategies components, while dimension 4 (problem outcomes [PO]) has no equivalent in the Leithwood and Stager model.

Dimension 1: Problem identification

The complexity and importance of PI should not be underestimated as weak problem definition impacts all subsequent stages. For example, in many schools, problems are commonly expressed as underachievement in a curriculum area (e.g., low reading achievement). This broad and general description complicates thinking about causes and solutions, because low reading achievement could suggest a myriad of possible problems including, but not restricted to, poor comprehension, weak decoding, or poor inferencing. When identifying and defining problems, leaders need to precisely describe their perception of the problem, including their perception of its seriousness, to seek others' responses and ideas, and check the demand to solve it. For example, reframing the broad definition of 'low reading achievement' to a more precise 'inability to simultaneously decode and comprehend text' enables greater precision in the next stages of the problem-solving process (Mintrop & Zumpe, 2019; Robinson, 2018).

Dimension 2: Problem cause

Most achievement problems leaders grapple with are complex and have obscure causes (Mintrop & Zumpe, 2019; Mumford et al., 2000; Robinson et al., 2020; Weber & Khademian, 2008). The causes of students' difficulties to simultaneously decode and comprehend text, for example, could be due to under or over-use of phonics, insufficient focus on building vocabulary, or problems with the text they are predominantly exposed to. Without carefully diagnosing and testing likely causes, finding a well-matched solution is a 'best guess' (Leithwood & Steinbach, 1991). Leaders' engagement in this dimension requires them to explicitly identify and test possible causes, including their hunches about how the relevant students are taught,

before reaching for solutions. This requires leaders to be respectfully honest about perceived problematic teaching practice and to access their teachers' reactions and alternative theories (Robinson, 2018; Robinson et al., 2020).

Dimension 3: Problem solutions

A good solution is strongly aligned to the cause of the problem and to the evidence about the likely effectiveness of any chosen strategy (Leithwood & Steinbach, 1991; Mintrop & Zumpe, 2019; Robinson et al., 2020; Timperley et al., 2007). For example, if the cause of the same decoding/comprehending text problem is identified as insufficient explicit teaching of vocabulary, the solution will require more front-loading and deliberate teaching of vocabulary. That solution alone, however, will not address the multiple causes of the problem. Successful solutions are those that match proposed strategies with all causes, rather than prioritising one (Leithwood & Steinbach, 1992; Nickles, 1981; Robinson et al., 2020). In this case, while front-loading vocabulary will assist, to be effective teachers need to pay attention to other strategies (e.g., phonics, choice of text, or deliberate acts of teaching comprehension) to effectively improve reading achievement (Nickles, 1981; Robinson, 2011).

Dimension 4: Problem outcomes

Problem outcomes explain the impact of the leaders' work with their teachers in the first three dimensions. Of importance to this study is to monitor task outcomes (reading achievement) and the leaders' relationships with their teacher, because we tested whether leaders can be assisted to avoid the task relationship dilemma and simultaneously give priority to improving the task, while maintaining and strengthening the leader/teacher relationship. Our theory of CPS suggests that the combination of *LbL* and deliberative problem solving will achieve this.

Methodology

A concurrent mixed methods design was used to quantify and explain shifts in the effectiveness of leaders' behaviour in CPS conversations subsequent to the intervention (Burke-Johnson & Onwuegbuzie, 2004; Leech & Onwuegbuzie, 2009). Next, we describe how schools, middle leaders, and teachers were recruited and how data was collected and analysed.

Sampling

We invited schools to participate whose senior leaders had attended an *LbL* workshop, because we anticipated that middle leaders would need the support of their senior leaders to learn and implement the new CPS approach. Schools were only included in the sample if two or more middle leaders responsible for Year 3–8 teachers gave consent to participate. These year levels were targeted because a multi-year

pattern of student underachievement in reading could be established, and the measures required to track reading progress were available. To mitigate the risk of rival explanations for any observed/reported change in our data, we required schools that had not planned professional learning in reading.

Three schools meeting all criteria were invited and agreed to participate. School A was a large urban primary school catering for 500 students from a mix of high and low socioeconomic communities, 15% of whom performed below age-related reading standards. School B was a mid-sized urban primary school catering for 350 students from a lower socioeconomic community, 45% of whom performed below age-related reading standards. School C was a mid-sized semi-urban middle school catering for 400 students from a mix of high and low socioeconomic communities, 40% of whom performed below age-related reading standards.

Four middle leaders from both Schools A and C and two from School B who met four selection criteria were invited to participate. Firstly, they could not have previously attended an *LbL* workshop. This was important because to confidently attribute any shifts in leaders' behaviour to the interventions, the learning had to be new (Shadish, Cook, & Campbell 2002). Secondly, they could pair with a teacher in their team who was willing to engage in the study. Thirdly, to mitigate attrition, we required they anticipated remaining in the school for the duration of the study. Fourthly, they accepted being randomly assigned to one of two interventions.

Two of the 10 leaders withdrew 5 months into the research due to employment changes, leaving a final sample of eight leaders. The eight leaders had been in their roles for an average of 4 years and their teaching experience ranged from 7 to 19 years.

The middle leaders (referred to as team leaders or senior teachers) were paid extra in recognition of their responsibilities, but only School B leaders were released from their classroom teaching (20% release). Each leader was responsible for up to four classroom teachers. The middle leaders' role descriptions had 10 key responsibilities in common; four relevant to this study: communicating student progress to the senior leadership team; leading their team to achieve the annual improvement focus; analysing data to set goals to improve student achievement; and overseeing curriculum planning, delivery, and reporting.

The two criteria used to identify teachers were a willingness to work with their middle leader to improve reading, and their intention to remain in the school for the study period. Twenty-eight teachers agreed to participate. We identified our final sample of 10 by inviting those with the greatest number of students reading below age-related standards (excluding students who were cognitively delayed, second language speakers, or receiving additional support). Pseudonyms have been used to protect participant confidentiality.

The design of the interventions

To resolve the task/relationship dilemma experienced by leaders, the design of our interventions and associated tools intentionally focused on simultaneously improving leaders' interpersonal effectiveness and building their knowledge and application

of a deliberative problem-solving process (Ackerman et al., 2011; Argyris & Schön, 1974; Day & Dragoni, 2015). We did not include building domain-specific knowledge and skill in reading assessment and pedagogy, because we assumed that given the New Zealand Government's intensive focus on literacy instruction for a sustained period leading up to the intervention, and the commitment of leaders and teachers to literacy outcomes, such knowledge and skill would be sufficient to improve the reading of the target students if applied in a conscientious manner (Education Review Office, 2015; Ministry of Education, 2014; Pont, Figueroa, Zapata & Fraccola 2013; Timperley & Parr, 2009). We were also aware that four of the eight middle leaders (one in each of Schools A and B and two in School C), held or had previously held responsibility for literacy leadership.

In accordance with best practice in professional learning for leaders, our design incorporated three important elements. Firstly, we taught leaders how to access and analyse their existing practice. One criticism of interventions is they are commonly disconnected from participants' current knowledge, experiences, and theories-of-improvement (Cavanagh et al., 2014; Timperley et al., 2007). Helping leaders to deeply understand their current practice, the rationale for it and its consequences, highlighted what they needed to learn, unlearn, and relearn to be more effective. Secondly, to strengthen the immediate applicability of the learning, 50% of the teaching focused on leaders' real and current problems (Bransford et al., 2000; Day & Dragoni, 2015; Leithwood et al., 2011; Timperley et al., 2007). Assisting a leader or teacher to solve their real day-to-day problems builds efficacy and motivation and reduces feelings of inadequacy (Twyford et al., 2017). Thirdly, leaders were provided with multiple learning/practice opportunities supported by specialist coaching and targeted feedback to deeply embed change, experience success, and be assisted in making sense of any implementation issues (Ackerman et al., 2011; Bransford et al., 2000; Huber, 2011; Timperley et al., 2007; Twyford et al., 2017).

Two interventions were designed to improve the quality and outcomes of the leaders' CPS conversations. By designing two interventions we could compare the impact of both a more and less intensive focus on the interpersonal components of our CPS model. Our assumption was that both interventions would lead to improvement, but those in intervention 2 would improve more. Both interventions had the same theory of problem solving taught in the same way.

For intervention 1 participants, *LbL* (interpersonal theory) was implicitly taught through the language used in the questionnaire that integrated our theories of problem solving and interpersonal effectiveness. For example, the questionnaire item '***I respectfully*** disclosed my concerns about the problem and ***checked that the teacher understood them***' (rather than 'I disclosed my concerns'), implicitly teaches the *LbL* motivations of valid information and respect.

For intervention 2 participants, *LbL* was explicitly taught. Intervention 2 leaders had the same access to the questionnaire tool, but were also taught to diagnose and understand their theory-in-use for CPS conversations. Theories of action describe links between behaviour, (actions/inactions), beliefs, and motivations that explain the behaviour, and the intended and unintended consequences of the behaviour (Argyris & Schön, 1974; Hannah, Sinnema, & Robinson 2019, 2021; Robinson, 2014, 2018; Sinnema, Hannah, Finnerty & Daly,

Table 1 Day 1 activities

What the interventionist did	What intervention 1 and 2 leaders did
Explained the research	First self-rating of the problem-solving questionnaire (naïve rating)
Taught the standard for the first three dimensions of our theory of problem solving (PI/PC/PS) using a video exemplar	Watched the video and used a template to analyse how the leader systematically gathered information about the first three dimensions
Processed the video activity with participants using the questionnaire to reinforce and illustrate the requisite behaviour in each dimension	
Showed the video for a second time	Reviewed and refined their analysis Second self-rating of the problem-solving questionnaire (informed rating)
Discussed dimension 4–problem outcomes	
Taught procedure for the completion of the weekly log	
Explained the use and role of an independent assessor to collect the target students' reading achievement	

2021). The causes of many complex problems lie in leaders'/teachers' theories of action, and consequently, improvement efforts are compromised if there is little or no engagement with them (Argyris & Schön, 1974; Cavanagh et al., 2014; Mintrop & Zumpe, 2019; Robinson, 2018; Robinson et al., 2020). For example, a teacher may believe that a reason for a student's reading difficulty is their lack of enjoyment in reading that contributes to their low vocabulary and low reading mileage. This belief may lead the teacher to only provide the student with texts that are directly connected to their interests (actions). The consequence of the teacher's beliefs and actions are that the student's reading mileage improves, but their vocabulary remains limited to the narrow range found in the books.

The interventions, requiring two days of delivery, were led by the third author with the first author observing and taking field notes. Leaders were randomly assigned to the interventions.

Day 1 activities

On Day 1, leaders from both groups attended for 5 h and were explicitly taught our theory of problem solving. Day 1 activities are explained in Table 1.

Additionally, we provided 3-monthly check-ins (coinciding with the interviews) to respond to questions/concerns raised by leaders and teachers and provide feedback to the leaders on their CPS practice.

Day 2 activities

Day 2 of the intervention (held separately for each group) involved intervention 1 participants in an additional 3.5 h of learning and intervention 2 participants in an additional 5 h of learning (resulting in a 90-min increased intervention dosage). The activities for each group are explained in Table 2.

Data sources

Data were collected through individual interviews with all leaders and teachers, questionnaires completed separately by leaders and teachers, audio recordings of leaders' conversations with teachers, and leaders' weekly logs. Two types of student achievement data were collected: running records and the results of a standardised test of reading achievement.

Interviews

Eighty interviews were held in total—five for each leader and teacher. Initial interviews occurred at the end of the school year for Schools A and B and the beginning of the following year for School C. A critical incident approach focused on the students' reading achievement was used (Flanagan, 1954). In the first interview, leaders were asked to visualise and recall the names of students (in their team but not their class) who were persistently underachieving in reading. Teachers were asked to do similarly for their own students. Questions focused on the interviewee's beliefs about the nature, seriousness, and cause of the reading problem; what guided their decisions and processes to solve it; and the effectiveness of their efforts. Leaders were asked to explain the decisions they made about their approach to these conversations. Post-intervention interviews followed the same process, occurring at 3-monthly intervals. These interviews focused on the progress of the target group (including what may have caused/inhibited the change for students), patterns emerging from the leader's log (e.g., infrequent conversations, rationale for choice of strategies), and perceived intervention impact on the leader's CPS conversation behaviour.

Questionnaires

Two parallel 31-item questionnaires (see "[Appendix](#)") were developed for leaders and teachers to independently rate the effectiveness of the leader's CPS behaviour. The parallel items in the leader and teacher versions of the questionnaire were worded differently to suit the two types of respondents. For example, leaders rated *'I genuinely inquired into what the teacher believed was contributing to the problem'* and teachers rated *'My leader genuinely inquired into what I believed was contributing to the problem'*. We used skip logic in the PI dimension to allow for the possibility the conversation would be initiated by the

Table 2 Day 2 activities

What the interventionist did	What intervention 1 leaders did	What intervention 2 leaders did
Checked and reinforced understanding of the dimensions of problem solving	Reflected on their understanding and sought clarification and/or further teaching where necessary	Reflected on their understanding and sought clarification and/or further teaching where necessary
Taught the theory of LbL	N/A	Engaged in discussion and supported reflection about own interpersonal effectiveness
Taught leaders to describe a theory-in-use using video exemplars	N/A	Viewed video exemplars and described (with support and coaching from the interventionist) two theories-in-use for a leader—one illustrating control-focused motivations and the other more learning-focused motivations
Provided critique of each leader's completed theory-in-use	N/A	Described their theory-in-use for the baseline conversation recorded prior to the workshop
Modelled and coached the behaviour for: Beginning a conversation (disclosing own view of a problem)	Practised (multiple times) with a peer (coached to replicate as close as possible their likely paired teacher's response), an imminent and important CPS conversation with a focus on:	Practised multiple times with a peer (coached to replicate as close as possible their likely paired teacher's response) an imminent and important CPS conversation with a focus on: Beginning a conversation, Suggesting alternative teaching approaches with reasons and whole of conversations
Suggesting alternative teaching approaches with reasons	Beginning a conversation	
Whole of conversations	Suggesting alternative teaching approaches with reasons	
	Whole of conversations	
Provided specialist coaching/feedback to leaders to improve their CPS conversations	Responded to coaching/feedback to reframe and improve their learning-focused thinking and action	Responded to coaching/feedback to reframe and improve their learning-focused thinking and action

leader (set A) or the teacher (set B). For example, '*I respectfully disclosed my concerns about the problem and checked the teacher understood them*' (set A) was reworded as '*I respectfully listened and summarised the concerns about the problem raised by the teacher*' (set B). There was no set B parallel for one set A item: (*I asked the teacher for frank feedback about whether or not they agreed there was a problem*) because as the teacher raised the issue this was redundant.

Leaders and teachers recalled and used the questionnaire to independently rate their conversations about the target students' reading progress over the previous 3 months. Respondents rated the leaders' effectiveness on a 6-point positively-packed agreement scale using four positive and two negative response options (see "[Appendix](#)"). We used a positively-packed scale because pilot data showed that participants responded to a balanced-scale with a tendency for positively skewed responses. In such cases, positively-packed scales can encourage greater attention to the degree of positivity and therefore increase validity of ratings (Brown, 2004; Lam & Klockars, 1982; Meric & Wagner, 2006).

The ratings were repeated six times (two pre- and four post-intervention) resulting in 96 ratings. Ratings were naïve at time point 1 and informed using a video exemplar of our standard for a CPS conversation at time points 2–6. We used an exemplar to demonstrate the standard because the validity of self-report ratings of performance is affected by a participant's understanding of the standard used to make their judgements (Crockett et al., 1987; Heidemeier & Moser, 2009; Mabe & West, 1982). To control for response bias, the first author rated 35 conversation transcripts (time points 2–6) on PI, PC, and PS (PO was not rated because the transcript did not give a basis for rating the quality of the professional relationships). The four post-intervention ratings of the questionnaire occurred in April, June, September, and November.

Conversation recordings

Leaders and teachers recorded one conversation towards the end of each 3-month period. Prior to the intervention, the conversation was about a student discussed during the interview (baseline measure). The four post-intervention conversations were about the progress of the target students. Forty conversations (five for each leader) were recorded in total.

Logs

Post-intervention, leaders logged details about their reading-related CPS conversations with their teacher on a template that included date, duration, conversation focus, and perceived progress with solving the problem (using a 6-point progress scale). The first author studied each leader's log in advance of the interview and formulated questions to probe their CPS approach.

Student achievement data

Data were collected on reading achievement of the 33 target students attached to the eight teachers and leaders. All 33 students had attended school for a minimum of 2 years and had not yet reached and maintained age-related standards.¹ Pre-intervention, we collected running record data, and post-intervention, running record data and results from the Supplementary Test of Achievement in Reading (STAR).

Running record

A running record is a reading assessment tool commonly used in New Zealand schools and in many other jurisdictions, for example, Australia, England and the United States (D'Agostino et al., 2021; Fawson et al., 2006), to monitor a student's reading progress and identify instructional strategies (Clay, 2000). Pre-intervention running record assessment data was collected at the end of the school year. Post-intervention, a trained and independent assessor collected and analysed four additional 3-monthly running record assessments using an unseen text. We used a student's instructional reading age (90–95% accuracy with at least 80% comprehension) as our benchmark. Assessments from the Price Milburn Benchmark Kit (Smith et al., 2008) and PROBE (Parkin & Parkin, 2020) were used because they provided standardised questions and marking/scoring protocols.

STAR

The STAR test provides a standardised measure and has high reliability, validity, and is designed for repeated measurement (Elley, 2001; Elley et al., 2014; Lai et al., 2009). The STAR test assesses how students are: (1) learning the code of written language; (2) making meaning of texts; and (3) thinking critically about texts. Each year level has three tests available—Form A, B, and C, with each test progressively more difficult. The STAR test was administered and scored by the same independent assessor in March, August, and November.

Data analysis

Quantitative data analyses

This section outlines the quantitative analyses undertaken to test the validity of our problem-solving questionnaire, the reliability of the independent raters, the shift in the effectiveness of leaders' problem-solving behaviour, and the outcomes of their CPS conversations.

¹ On a leader's request, we included three students classified as 'at' due to a concern about the validity of the end-of-year data for those students.

Confirmatory factor analyses

Two confirmatory factor analyses were undertaken and confirmed that our theoretical understanding of the four dimensions of problem solving matched our measures of problem-solving conversation effectiveness and problem-solving conversation outcomes (Patuawa et al., 2021). Scale reliabilities for PI, PC, PS, and PO were calculated using Cronbach's alpha and all four scales were above the 0.9 threshold used (PI and PC = 0.93; PS = 0.91; PO = 0.92).

Inter-rater reliability

Considering the reliability of leaders' and teachers' ratings is important from a leadership evaluation point of view (Sinnema, Robinson, et al., 2015) and also from a methodological point of view. We checked inter-rater reliability between leaders and teachers by testing for the significance of differences in their independent questionnaire ratings using paired *t*-tests. The paired *t*-tests were conducted on the variables used for our subsequent analyses: the four dimensions of problem identification (PI), problem cause (PC), problem solutions (PS) and problem-solving outcomes (PO), and the single problem-solving construct. There were no significant differences; all *p*-values were < 0.5.

Analysis of effectiveness ratings

To determine the effectiveness of the leaders' CPS conversations, an average rating for PI, PC, and PS was calculated as rated by leaders, teachers, and the first author. Added to this was the average ratings of PO as rated by leaders and teachers only. The averages of the scores in the four dimensions were then used to calculate an overall problem-solving effectiveness score for each leader. Averages of those scores were compared at each time point pre- and post-intervention using Welch's *t*-test of 5% significance.

Analysis of logs

Logs were submitted at 3-monthly intervals. Data relating to the frequency and duration of a leader's conversations were entered into an Excel spreadsheet and matched with the record of the reading achievement of each of the leader's target students.

Analysis of student reading data

Running record data was used to measure the shift in reading achievement. We calculated (in months) the difference between each student's reading age at baseline and after 12 months of instruction during which time the leader had been

working with their teacher. We averaged this across all 33 students to determine overall rate of progress.

We applied the Wilcoxon signed-rank test to determine the significance of any differences in the target student's reading age based on the analysis of running records between the two time points. The Wilcoxon signed-rank test was chosen because the difference in the data between the two time points was non-normally distributed making it the most relevant test (Rey & Neuhäuser, 2011).

STAR data was used to calculate the difference in each student's STAR reading scale score at the end of the first term and the end of the final term of the school year. We averaged this across 32 students to calculate the overall rate of progress.²

We conducted pairwise *t*-tests to assess whether the mean reading achievement (as indicated in STAR) at the two time points was statistically different at 5% significance level. We set the null hypothesis as the expected gain in STAR reading by year level instead of the usual zero, as we expected, for example, a Year 3 student to improve by 27.6 STAR reading points over the academic year (Zhu, 2020).

Qualitative data analyses

This section outlines the qualitative data analyses undertaken to describe shifts in leaders' CPS behaviour and the outcomes of their conversations.

Theory-in-use

We used conversation and interview transcript data to reveal the theories-in-use (TiU) for the leaders' conversations with their teachers about their target student's reading achievement. We developed nodes in NVivo for the four components of a TiU: (1) beliefs, (2) motivations, (3) actions/inactions, and (4) consequences. Following our model of CPS, three sub-nodes were developed to individually code beliefs and motivations, actions/inactions, and consequences about PI, PC, and PS. Inactions were inferred from the conversation and interview transcripts and validated with each leader (e.g., in a conversation transcript we noted on three occasions a teacher's request for advice about how to support a student, with no response from the leader). When the leader viewed the transcript and was asked what caused their non-response, the leader replied, "The teacher is more knowledgeable about literacy, and I feel I have nothing to offer". Two nodes were developed for PO, one that related to reading achievement and one related to outcomes for the leaders' professional relationships with their teacher. In addition to these TiU nodes, additional nodes were used to code conversation frequency, duration, and intervention impact. On completion of the TiU for all eight leaders, the drafts were shown to each leader, feedback sought about their accuracy, and revised as negotiated.

² One student was missing a scale score at time point 5 and was consequently excluded from the analysis.

Findings

In this section we use quantitative data derived from leaders and teachers (time points 2–6) to report the shift in the effectiveness of middle leaders' CPS conversation behaviour, and the outcomes of their conversations for the leaders' relationship with their teacher and the task (student achievement in reading). We use qualitative data to illustrate and explain the shifts made. Our quantitative analysis showed no significant differences in the shifts between the two intervention groups, and consequently we report the two groups as one (Kim et al., 2020).

Research question 1: The reported impact of the intervention on leaders' behaviour in problem-solving conversations

Our comparison of leaders' pre-intervention (time point 2) and post-intervention ratings showed that leaders reported a shift from ineffective to more than moderately effective. The teachers' ratings of their leaders' behaviour effectiveness in the same conversations shifted from slightly effective to more than moderately effective. Both shifts were statistically significant (see Table 3). We used the informed baseline rating to investigate shifts because the difference between the naïve and informed ratings was significant (reducing from an average of 4.57–2.81 for leaders and 4.12–2.37 for teachers), highlighting the importance of exposure to a standard (Heidemeier & Moser, 2009; Patuawa et al., 2021).

Problem identification

Leaders' ratings of their approach to describing and seeking agreement about the target students' reading achievement problem, its seriousness, and the demand for it be solved shifted from slightly effective to nearly mainly effective and the teachers' ratings of their leader's behaviour shifted similarly (see Table 4).

Prior to the intervention, leaders typically believed teachers would seek support if required, and consequently did not often have planned conversations about the target students' reading progress. When conversations happened, leaders were typically vague in raising their concerns about student achievement and focused their meetings on student behaviour, because their senior leaders' meeting agendas prioritised behaviour over achievement. Additionally, leaders suggested they grappled with how to talk about achievement issues because they were afraid to be critical of colleagues, fearing defensive responses and upsetting their hard-working teachers (Bassett & Shaw, 2018; Cardno et al., 2018; Harris & Jones, 2017; Robinson et al., 2020). Leaders also saw such discussions of achievement as complex. For example, Fred (leader) said:

I think we are more comfortable to talk about behaviour because it's easier. To get to the nitty gritty of the learning of target group students, I think is more

Table 3 Effectiveness of leaders' pre- and post-intervention behaviour in collaborative problem-solving

Dimension	Intervention	Leader			Teacher		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
						Limit	Limit
						Upper	Lower
Problem solving	Pre	8	2.81	0.32	8	3.01	0.96
	Post	32	4.57	0.71	32	4.38	0.86
	Welch's <i>t</i> -test		<i>t</i>	10.38		<i>t</i>	3.66
			<i>df</i>	26.17		<i>df</i>	9.97
						<i>p</i> -value	<i>p</i> -value

[illegible][illegible]

Table 5 Illustrative quote indicating a shift in leaders' behaviour in problem identification

Problem identification related talk	
Pre-intervention	Post-intervention
Bob (Leader): "We're supposed to have a conversation today about [the student] and her reading. So, what level did she achieve?"	Bob (Leader): "When I reviewed your end of term reading data, I thought some of your kids have made a substantial amount of progress, and yet as I see it, some of your target children have made very little progress, for example [the student A, B and C]. Do you know what I mean?" Jan (Teacher): "Yes I agree that they have made very little progress." Bob: "Can you tell me why you think have they made such little progress?"

threatening. To be honest, I'm more confident talking about behaviour than all the different learning needs because there are so many, and I don't think we have a handle on what to do with them all, so we avoid talking about them.

Post-intervention, leaders' conversations focused more explicitly on target students' progress. They applied their learning from the intervention about the importance of precision and checking assumptions by directly disclosing their perception of the problem, and more explicitly, inquiring into the teacher's perception. Leaders' open and respectfully honest disclosure of their thinking, together with their commitment to learning from the teacher, increased the validity of the information about the reading problem and built trust. For example, the opening of the pre-intervention CPS conversation in Table 5 is very general and implies a requirement to discuss [the student]. The opening of the post-intervention conversation illustrates a shift to a more precise definition of the problem with evidence provided to support the leader's concerns.

Problem cause

At pre-intervention, leaders and teachers rated their CPS conversations as weaker in causal inquiry compared to any other dimensions of our CPS model, and it remained so after the intervention. There was, however, a significant shift in leaders' rating of their behaviour from ineffective to more than slightly effective. While teachers' ratings of leaders' attention to problem cause increased from pre- to post-intervention, the shift was not significant (see Table 6).

During the intervention, as leaders used the CPS tool to reflect on their practice, they became acutely aware that they had leapt to quick fixes and had bypassed causal inquiry. At pre-intervention, only one of the eight leaders made explicit reference to cause. Instead, leaders typically reacted intuitively and focused their conversations on what they would do to improve the target student's reading. This is illustrated in the following excerpt:

Table 6 Effectiveness of leaders' pre- and post-intervention behaviour in problem causes

Dimension	Intervention	Leader				Teacher			
		<i>n</i>	<i>M</i>	<i>SD</i>	Limit	<i>N</i>	<i>M</i>	<i>SD</i>	Limit
					Upper				Upper
					Lower				Lower
Problem cause	Pre	8	2.16	0.78	2.68	8	2.63	1.16	3.41
	Post	32	3.48	1.32	3.87	32	3.54	1.27	3.92
	Welch's <i>t</i> -test		<i>t</i>	3.66	<i>p</i> -value		<i>t</i>	1.95	<i>p</i> -value
			<i>df</i>	18.50			<i>df</i>	11.51	

Interviewer: So, you quickly agreed the student's underachievement was serious. Did you discuss what might be contributing to his lack of progress?

Jenny: Well no, because it was so early in the year. And to us, it's a really important time to get to know the child.

Interviewer: And have you discussed it since?

Jenny: No

Interviewer: Can I ask why?

Jenny: I don't tend to think like that. I am more focused on what we can do to help them.

Although not explicitly discussed, there was often an implied cause in the suggested improvement strategies. For example, when Jenny was asked what she was doing for [the student], she explained she chose to use texts linked to [the student's] interest to get to know and engage them (implicit cause of the reading difficulty—lack of engagement). These implicit causes were commonly assumed to be valid rather than checked.

In post-intervention conversations, the leaders applied their new learning about the importance of causal inquiry and were much more systematic and disciplined about discussing cause with their teachers. They did not however, always give the desired attention to testing the validity of the causal assumptions made. For example, one middle school teacher midway through the research had an unchallenged and untested causal theory about teachers' relationships with students being the most important variable for improving achievement. While quality of relationships is undoubtedly important, effective relationships on their own are insufficient to improve reading achievement (Education Review Office, 2015). With additional feedback and coaching, the leader later challenged this belief and assisted the teacher to integrate positive relationships with high expectations for learning.

Amongst the lowest scoring items in the PC dimension were items relating to leaders respectfully sharing their view with reasons about the cause. A primary school middle leader (Grant), suggested the reason for this was a fear of upsetting the teacher:

Interviewer: I want to probe your understanding of the busyness that disrupts consistent teaching and whether or not you can talk to [the teacher] about your concerns?

Grant: I just feel when I go into her classroom it's a bit loose. I think that's why she hasn't found time to sit with her target learners. To be honest, I'm finding that quite challenging to raise with her.

Interviewer: Why is it challenging?

Grant: I don't want her to believe I think she's a bad teacher.

Interviewer: What makes you think she's going to think that?

Grant: Because she's a lovely person and very sensitive. I suggested she try allocated seating and she took that personally.

While leaders' behaviour in causal inquiry remained weaker than in the other dimensions, our qualitative analyses indicated some notable shifts. Table 7 contrasts how Lucy (a middle school team leader) mentions cause only fleetingly in her

Table 7 Illustrative quote indicating a shift in leaders' behaviour in problem cause

Problem causes related talk	
Pre-intervention	Post-intervention
Lucy (Leader): Thinking about literacy, because we haven't got that up and going yet so we can do anything with our target kids, what do you think (pause)?"	Lucy (Leader): Why do you think inferencing is contributing to her underachievement in reading?
Graham (Teacher): "What do you think?", do you mean, "What shall we do?"	Graham (Teacher): In this STAR assessment, [the student] is missing some important information for inferencing. For example, "Belinda waited with impatience at the end of the platform", and the question is, "What do you think Belinda said when she called out to Nick?", and [the student] said, "Be careful". The answer should be something like "Come on, hurry up", because she's impatient. So, she's missing those cues. Another one, "Nick sat up, staring into the darkness" and the question is, "Where was Nick when he heard the blast of the whistle?" [The student] said "In the house, having breakfast". Again, missing cues
Lucy: Yeah, what shall we do, or what's causing them to underachieve and then what is it we think might work? Because we've already had these kids for a year	Lucy: I agree and looking at the data, that seems to be a common pattern for several of our target students, so we should work to improve their inferencing
Graham: Yeah. Well, with [the student] her language barrier is probably still a factor and [the students A and B] don't get a lot of support for reading from home and don't necessarily think deeply about what is read	Graham: I agree with the focus on inferencing, but I think I am going to need to also focus on vocabulary because without strong vocab knowledge they cannot make inferences
Lucy: So, what do you want to do...?	Lucy: I thought you had done lots of vocabulary work with these students—is that not the case?
	Graham: Not as much as I should have, I definitely think I need to tie vocabulary and inferencing together

pre-intervention conversation with the expectation in her post-intervention conversation, that Graham (teacher) will use available data to explain with reasons his causal hypotheses and intended teaching focuses (maximise valid information).

Problem solutions

There were significant shifts in leaders' ratings of their effectiveness in designing solutions to improve the target students' reading achievement (see Table 8). Leaders' ratings of their behaviour shifted from less than effective to more than moderately effective and their teachers agreed.

At pre-intervention, leaders commonly moved quickly from a general description of the problem to discussing actions/solutions or addressed the two simultaneously. The common question asked by leaders was, "How do we get the student to improve?" The intervention taught leaders the difference between single and double-loop learning, and leaders quickly identified their default behaviour was the former—that they typically iterated through a series of different actions to try and remediate the achievement problem. Without exploration of the cause, suggested

Dimension	Intervention	Leader			Teacher					
		<i>n</i>	<i>M</i>	<i>SD</i>	Limit	<i>N</i>	<i>M</i>	<i>SD</i>	Limit	
										Upper
Problem solution	Pre	8	2.63	0.71	3.10	2.15	2.88	1.25	3.71	2.04
	Post	32	4.13	1.38	4.54	3.72	4.12	1.28	4.50	3.73
	Welch's <i>t</i> -test		<i>t</i>	4.32	<i>p</i> -value	0.000	<i>t</i>	2.51	<i>p</i> -value	0.031
			<i>df</i>	22.05			<i>df</i>	10.98		

Table 9 Illustrative quote indicating a shift in a leaders' behaviour in discussing problem solutions

Problem solutions related talk	
Pre-intervention	Post-intervention
Lucy (Leader): So what do you want to do in your reading programme?	Lucy (Leader): So, if inference and vocab are to be our foci, and we specifically want them to notice important information, I have some thoughts about what we could do... What I have had them do is to read an article first, and I've got them to write up the whole sentence where there's a word they don't know. We then workshop together, but before we do, we talk about those unknown words and work out their meaning. Then we go to learning about the inference
Graham (Teacher): There are probably some things we could do around building on their decoding and word attack skills and some strategies to get them thinking more deeply about what they read	Graham (Teacher): Definitely clarifying those unknown words is important
	Lucy: We've also been talking about using picture books. It could be with those target students that the picture book is read at the beginning, as a whole class and together we could surface and work out the unfamiliar words. Then we could do an inference activity
	Graham: So, they've already had that vocabulary built for them

solutions were likely to be ineffective in achieving desired results. This was reflected in the frustration of the teacher in the following excerpt:

James: He has made no progress, so, I'm trying to ramp it up. I have considered what other things I can do. I've done blends, I've done sight words and I'm now trying word study, but he is still going backwards. He's very stilted. It's his fluency, it's very shaky and he hasn't got to that point where he's running through. I've even got some easier books and he's doing some buddy reading.

During post-intervention conversations, leaders worked hard to interrupt the single-loop cycles they had recognised during the intervention as their default, and were consequently more likely to discuss whether or not their proposed solutions would match the problem and its causes. For example, continuing the conversation between Lucy and Graham, in the pre-intervention conversation, Graham (teacher), on the request of Lucy (leader), very quickly made suggestions for programme inclusions. In their post-intervention conversation (see Table 9), Lucy provided a summary of the agreed teaching focus (based on their causal inquiry), and then discussed possible strategies to specifically address the identified learning needs of the target students.

The lowest scoring item in this dimension was the need to critically evaluate the quality of the chosen solutions. For example, Lucy and Graham could have discussed why they believed their suggested strategies would improve student's inferential reasoning and asked each other whether there was anything they had missed or that could be more effective.

Research question 2: Reported outcomes of the intervention on problem outcomes

Leaders' ratings of the outcomes of their CPS conversations shifted significantly from less than effective to more than moderately effective and teachers agreed (see Table 10).

Leaders and teachers reported that they persisted more in their efforts to improve reading for the target students as evidenced by the increased frequency of their conversations, the deliberate learning focus they brought to them, and their greater attention to monitoring progress. The increased reported frequency of conversations focused on learning was an intended outcome of the intervention design, which set expectations for such conversations as a requirement for participating in the research. Leaders and teachers suggested this enabled them to learn more about the problem and how to solve it. Carol (middle school teacher) and Sam (leader) reflected together on how they found their joint work:

Carol: I have enjoyed working with you on reading this year. The biggest highlight for me is actually understanding what I'm teaching and the reason why I'm teaching it. I think I've got a more in-depth understanding of where my students sit, what to do to help them and I'll continue tracking progress. I think your openness to my ideas got me on board. Thank you for not coming and saying, "This is what we're doing", without asking for my input.

Sam: I agree...I've looked at your results and compared your mathematics, reading and writing, it's easy to see your children have progressed most in reading.

The highest scoring item in this dimension was 'We established a strong level of trust'. Jan, a teacher who was initially reluctant to participate in the research based on a prior experience that left her lacking in self-esteem, efficacy, and confidence, shared how although she experienced the conversations with her leader as evaluative and challenging, she felt supported to be more effective, rather than being labelled as a poor teacher.

Reading achievement of the target students also improved significantly and is reported next.

Assessed impact of the intervention for the reading achievement of the target students

At pre-intervention, running records showed 30 of the 33 students were reading below their chronological age (the standard) and 16 of them had been below for 2 years or more (see Table 11). At time point 5 (the final post-intervention data collection point), while 18 students remained reading below their chronological age, 29 students had made accelerated progress (18 months or more), 3 students had made expected progress (12 months) and 1 student made less than expected

Table 10 Effectiveness of leaders' pre- and post-intervention behaviour in problem outcomes

Dimension	Intervention	Leader				Teacher					
		<i>n</i>		<i>M</i>	<i>SD</i>	Limit	<i>n</i>		<i>M</i>	<i>SD</i>	Limit
		Upper	Lower	Upper	Lower						
Problem outcome	Pre	8	2.83	0.31	3.03	2.62	8	2.78	0.67	3.22	2.33
	Post	32	4.38	1.29	4.77	4.00	32	4.41	0.95	4.69	4.12
	Welch's <i>t</i> -test		<i>t</i>	6.15	<i>p</i> -value	0.000		<i>t</i>	5.61	<i>p</i> -value	0.000
			<i>df</i>	38.00				<i>df</i>	14.96		

Table 11 Target students' reading achievement pre- and post-intervention

Category	Description	Pre-intervention <i>N</i> =33	Post-intervention (12 months later) <i>N</i> =33
Well above	More than 12 months above chronological age	0	5
Above	Between 6 and 12 months above chronological age	0	6
At	Between 0 and 6 months above chronological age	3	4
Slightly below	Between 0 and 6 months below chronological age	7	8
Below	Between 6 and 12 months below chronological age	7	5
Well below	More than 12 months below chronological age	16	5

progress (0 months) in reading. The average progress across the cohort of 33 students during the 12-month timeframe was 23 months of reading progress.

The relationship between these quite marked shifts in the target students' reading achievement and the interventions is the focus of the third and final phase of the research and will be further discussed in a subsequent article.

The Wilcoxon signed-rank test showed the post-intervention shift in achievement in running records was statistically significant at 5% (p -value < 0.05). The paired t -test showed that post-intervention progress in the results of the STAR test were statistically significant with an average acceleration of 9.4 star reading scale scores ($n=32$, $t=5.37$, p -value < 0.001).

Discussion

This study adds to the limited pool of intervention studies on educational leaders' problem-solving, by reporting the impact of an intervention to assist middle leaders to collaborate with their teachers to solve complex teaching and learning problems preventing their students from reaching age-related reading standards.

The major contribution of this research is that it has shown how a leadership intervention designed to improve middle leaders' problem-solving capability had positive consequences for the leaders' and teachers' practice, professional relationships, and their students' reading outcomes. The interventions improved leaders' capability in talking with teachers in ways that built trust, while simultaneously improving their inquiry into the classroom-based causes of, and solutions to, students' reading problems.

A second contribution is our CPS model which builds on previous studies (Graesser et al., 2018; Leithwood & Stager, 1989; Mintrop & Zumpe, 2019; Robinson et al., 2015) and a tool (currently rare) for assessing leaders' CPS capability (the questionnaire). The model and associated tool could be used to build knowledge of effective CPS behaviours and to self-assess and/or seek feedback on CPS capability. Of importance, however, is the need to build shared understanding of the standard implicit in the tool so users know what constitutes effectiveness. We achieved this through explicit teaching of the model and exemplified the standard with a video of

a CPS conversation. This is especially important if the tool is used to measure intervention impact because leaders cannot be expected to objectively rate effectiveness in the absence of an exemplar (Heidemeier & Moser, 2009).

A third contribution is our description of middle leaders' actual relational and problem-solving capabilities as they worked to solve real and pervasive school-based teaching and learning problems. Most existing problem-solving studies use scenarios designed to elicit leaders' reasoning about the problem and how to solve it (Brenninkmeyer & Spillane, 2008; Leithwood & Steinbach, 1992; Marcy & Mumford, 2010). While the scenarios are based on events common to the experience of leaders, they bypass the conditions that leaders work under such as competing role demands, fast-paced work, interpersonal dynamics, and pressure to perform (Graesser et al., 2018; Mintrop & Zumpe, 2019; Robinson et al., 2020). Because there is less at stake in a scenario response, leaders can be more clinical in their approach, but as Mintrop and Zumpe (2019) found, leaders find it difficult to shift their mindsets. Solving the problem of inequitable student reading outcomes is not a solo endeavour and requires a sustained effort by all involved (Robinson, 2011). Leaders must over time manage team dynamics, doubts and disagreement, resistance to change, and be confident enough to directly challenge and improve ineffective teaching practice. Middle leaders report that dealing with difficult staff and talking about problematic teaching practices is one of their biggest challenges (Bassett, 2018; Cardno, Robson, Deo, Bassett & Howse, 2018; Harris & Jones, 2017).

A fourth contribution is our detailed quantitative and qualitative description of leaders' behaviour in four discrete dimensions of problem solving. Studying leaders' real-life approaches to problem solving helps to better understand their thinking, their resultant behaviour, and consequently clarifies the required target of interventions. For example, our results suggest the need for increased emphasis on causal inquiry and critical evaluation as they were the lower rated behaviours. Causal inquiry is highlighted in the problem-solving literature as a skill that is frequently overlooked (Graesser et al., 2018; Leithwood & Stager, 1989; Marcy & Mumford, 2010; Mintrop & Zumpe, 2019; Robinson et al., 2015; Robinson et al., 2020). Certainly, this study provides evidence that prior to intervening, leaders did not explicitly think about or discuss cause, but were instead focused on the actions that could be taken to improve the target students reading. This pattern may be perpetuated by a system-wide default to reach for a 'quick fix' as seen in the introduction of system-wide improvement strategies prior to a careful diagnosis of the problem that needs fixing. For many years in New Zealand, for example, it was assumed that lower achievement results in writing were a consequence of students lacking motivation to write. This causal assumption was not checked, and schools spent much time, resource, and energy investing in resources and programmes to increase the motivation of students, only to find much later, that the lack of motivation was a consequence of a more significant cause—the capability of teachers to effectively teach writing (Timperley & Parr, 2009).

Another reason the leaders avoided causal inquiry was their reluctance to engage in critical evaluation. Critical evaluation is rarely overtly deployed. We say 'rarely overtly', because critical evaluation happens but is mostly locked in the private thinking of the evaluator (Sinnema, et al., 2013; Sinnema et al., 2021). In this study,

two reasons emerged for the leaders' lack of critical evaluation and were both a consequence of control-focused motivations. The first was that leaders were largely self-professed conflict avoiders (protecting self and others). The intervention helped leaders to overcome this fear of conflict by illustrating that in not raising issues directly and respectfully with their teacher, they contributed to the problem by not learning sufficiently about, and providing, the targeted support the teacher required to better assist their students. The second reason for the absence of critique was leaders fearing they did not have the knowledge to meaningfully evaluate or advise the teachers. In this situation, leaders were encouraged to be honest about this and to be a conduit for seeking the help required.

Finally, we discuss the likely reasons that leaders in intervention 2 did not improve more than those in intervention 1 as anticipated. Firstly, there was sufficient behavioural information embedded in the questionnaire items to encourage leaders in intervention 1 to think about and moderate how they raised concerns, engaged in causal inquiry and discussed preferred teaching approaches. Secondly, the explicit teaching of CPS conversations in the deliberative problem-solving process, combined with opportunities to practice conversations with feedback and specialised coaching from the interventionist, focused the leaders on better and more respectful inquiry, advocacy, and listening. This was also reinforced in the check-in sessions coinciding with the interviews. A final possible reason is that the two leaders who withdrew from the research were both randomly assigned to intervention 1, leaving only three leaders (as compared to five) in that sample. This may have reduced the diversity in the leaders' results and thus reduced the difference overall.

Limitations

Our sample size was smaller than desired, necessitated by the complexity of the study and the volume of data collected. There is sufficient evidence in this study to warrant a larger study to test the replicability of our findings. Our intent was initially to understand through highly detailed analyses the effectiveness of leaders' CPS conversations, however, we believe despite the small sample, we can generalise our findings. Existing studies about the problem-solving behaviour of leaders suggests our results are indicative of a typical pattern of CPS behaviour—particularly leaders' reliance on intuitive approaches and minimal consideration given to causal inquiry (Brenninkmeyer & Spillane, 2008; Graesser et al., 2018; Leithwood & Steinbach, 1992; Marcy & Mumford, 2010; Mumford et al., 2000; Robinson et al., 2020).

A second limitation is the lack of a control group. We initially considered an experimental design, but the timeframe and available resources did not allow for a delayed intervention that might have overcome the ethical issue of leaving one group untreated. Instead, we chose to compare two different interventions. While our design of these interventions included controls for the confounding effects of history and maturation, a control group would have enabled us to further control for the results being attributed to better leadership performance resulting simply

from the intensive focus and attention given to how the middle leaders supported their teachers to shift the target students' reading achievement.

The duration of the intervention is another possible limitation. We found leaders needed commitment and support to learn, unlearn, and relearn strongly ingrained patterns of behaviour (e.g., conflict avoidance, reliance on inquiry, rushing for a quick fix). Schwarz (2017) suggests that 98% of people when confronted by threat, default to the control-focused motivations previously explained. If we were to redesign this intervention, we would plan for more regular follow-up activities and include a more extensive mentoring component focused on helping leaders to reframe their theories of action for CPS conversations.

The singular focus on reading and on students who had not reached age-related standards is a further limitation. This raises questions about the replicability for other domains (e.g., writing) and other groups of students (e.g., second language learners). We assume that because this was not a reading intervention per se, but a leadership intervention to support middle leaders to be more effective in their CPS conversations, that the same interventions may similarly impact different domains and different types of target student.

A final limitation is that we did not investigate leaders' domain-specific knowledge; specifically, their curriculum or pedagogical knowledge in reading. Future research could include the identification and remediation of gaps in leaders' and teachers' domain-specific knowledge, because such knowledge is likely to enhance a leader's ability to diagnose and resolve in the relevant domain (Timperley et al., 2007).

Conclusion

The quality of leaders' CPS is important in addressing the inequity issue many jurisdictions grapple with. Every educational leader strives to improve the engagement and achievement of their students, and yet every educational leader, despite their best efforts, frequently encounters situations where this is difficult because the complexity of their work and contexts exceed the resources and capabilities they can deploy (Mintrop & Zumpe, 2019; Patuawa et al, 2021; Robinson et al., 2020). Inequity will persist unless school leaders are more effectively supported to deliberately diagnose and solve school-based problems that contribute to it. This complex work requires attention to improving leaders' interpersonal and problem-solving capabilities, because both are necessary for leaders to engage productively with others in respectful meetings and conversations about issues affecting the quality of teaching and learning. Leaders make daily decisions that impact the opportunities of their students and staff to achieve the goals they aspire to, and it is therefore essential that those decisions are based on high-quality information and not just well-intended guesses. To achieve this, we recommend that those with responsibility for leadership programmes designed to support senior and middle leaders in schools include a professional learning focus on building CPS capability.

Appendix: Leader's problem-solving questionnaire

	Think about all the students in your paired teacher's class who are reading below their chronological age (the problem). Now think about the conversations you have had with this teacher this term about those students' achievement. During the course of those conversations I believe...	Strongly disagree	Slightly disagree	Slightly agree	Moderately agree	Mainly agree	Strongly agree
Part 1 Establishing agreement to address the problem if raised by you	1. I respectfully disclosed my concerns about the problem and checked that the teacher understood them.						
	2. I was honest about the seriousness of the problem, as I understood it.						
	3. I gave specific reasons and or examples to explain why I thought there was a problem.						
	4. I asked the teacher for frank feedback about whether or not they agreed there was a problem.						
	5. I gained sufficient agreement from the teacher about the seriousness of the problem.						
	6. I established sufficient agreement with the teacher about the importance of solving the problem.						
Part 1B Establishing agreement to address the problem if raised by the teacher	7. I respectfully listened and summarised the concerns about the problem raised with me by the teacher.						
	8. I offered the teacher my view about the seriousness of the problem the teacher had raised.						
	9. I respectfully explained my reasons and or examples to illustrate my view of the seriousness of the problem						
	10. We gained sufficient agreement from the teacher about the seriousness of the problem.						
	11. We established sufficient agreement with the teacher about the importance of solving the problem.						
Part 2 Inquiring into the possible causes of the problem	12. I genuinely inquired into what the teacher believed was contributing to the problem.						
	13. I listened carefully to the teacher's beliefs about what was contributing to the problem.						
	14. I respectfully disclosed my views about how his/her teaching might have contributed to the problem.						
	15. I gave specific reasons/examples to explain why I think his/ her teaching might have contributed to the problem.						
	16. I invited the teacher's honest reaction, including possible disagreement, to my beliefs about how his/her teaching might have contributed to the problem.						
	17. I gained sufficient agreement from the teacher about how to check our beliefs about the impact of teaching on the problem.						
	18. I explicitly requested we consider carefully other possible causes not yet discussed.						
Part 3 Planning and evaluating possible solutions to the problem	19. I respectfully asked the teacher to share his / her thinking about what changes to their teaching might improve the problem.						
	20. I respectfully asked the teacher why he/she thought that his/ her suggested changes would improve the problem.						
	21. I shared my ideas with reasons, about what changes I believe should be made in his/her teaching practice and gave examples of what those changes might look like.						
	22. I genuinely invited the teacher's honest reaction, including possible disagreement, with my ideas.						
	23. I ensured we critically evaluated whether our various suggested changes in teaching practice would address the causes of the problem.						
	24. I checked the teacher was committed to making the changes we had planned to improve the problem.						
Part 4 Reflecting on the outcomes of our problem solving approach	25. We learnt a great deal about how we might solve the problem.						
	26. The teacher made changes to his / her teaching practice.						
	27. The changes the teacher made to his / her teaching practice improved the students' reading achievement.						
	28. We persisted in our efforts to solve the problem.						
	29. We were rigorous in monitoring progress toward solving the problem.						
	30. We established a strong level of trust.						
	31. We felt accountable to each other for solving the problem.						

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

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

References

- Ackerman, R. H., Donaldson, G. A., Mackenzie, S. V., & Marnik, G. F. (2011). Leadership learning that makes a difference in schools: Pushing the frontier at the University of Maine. In T. Townsend & J. MacBeath (Eds.), *International handbook of leadership for learning—part 1* (Vol. 25, pp. 375–396). Springer.
- Argyris, C., & Schön, D. (1974). *Theory in practice: Increasing professional effectiveness*. Jossey-Bass.
- Argyris, C., & Schön, D. (1996). *Organizational learning II: Theory, method and practice*. Addison-Wesley.
- Aygyris, C. (1991). Teaching smart people how to learn. *Harvard Business Review*(May–June), 99–109.
- Bassett, M. (2018). Building the confidence of first-time middle leaders in New Zealand primary schools. *International Journal of Educational Management*, 32(5), 749–760. <https://doi.org/10.1108/IJEM-05-2017-0101>
- Bransford, J., Brown, A., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded). National Academy Press.
- Brenninkmeyer, L. D., & Spillane, J. P. (2008). Problem-solving processes of expert and typical school principals: A quantitative look. *School Leadership & Management*, 28(5), 435–468. <https://doi.org/10.1080/13632430802517209>
- Brown, G. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Report*, 94(3), 1015–1024. <https://doi.org/10.2466/pr0.94.3.1015-1024>
- Burke-Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Cardno, C. (1998). Making a difference by managing dilemmas. *set*, 1, 1–4.
- Cardno, C., Robson, J., Deo, A., Bassett, M., & Howse, J. (2018). Middle-level leaders as direct instructional leaders in New Zealand schools: A study of role expectations and performance confidence. *Journal of Educational Leadership, Policy and Practice*, 33(2), 32–47. <https://doi.org/10.21307/jelpp-2018-011>
- Cardno, C. (2007). Leadership learning—the praxis of dilemma management. *ISEA*, 35(2), 33–50.
- Cavanagh, B., Roberts, A., Nowotarski, J., Garner, J., Kerr, L., Mikaere, M., Waho, T. (2014). *Report of the Professional Learning and Development Group*. Retrieved from <https://www.education.govt.nz/assets/Documents/Ministry/Initiatives/PLDAdvisoryGroupReport.pdf>
- Clay, M. (2000). *Running records for classroom teachers*. Heinemann.
- Crockett, L., Schulenberg, J., & Petersen, A. (1987). Congruence between objective and self report data in a sample of young adolescents. *Journal of Adolescent Research*, 2(4), 383–392. <https://doi.org/10.1177/074355488724006>
- D'Agostino, J. V., Rodgers, E., Winkler, C., Johnson, T., & Berenbon, R. (2021). The generalizability of running record accuracy and self-correction scores. *Reading Psychology*, 42(2), 111–130. <https://doi.org/10.1080/02702711.2021.1880177>
- Day, D. V., & Dragoni, L. (2015). Leadership development: An outcome-oriented review based on time and levels of analyses. *Annual Review of Organizational Psychology and Organizational Behaviour*, 2, 133–156. <https://doi.org/10.1146/annurev-orgpsych-032414-111328>
- De Nobile, J. (2010). Openness as a feature of communication in Catholic primary schools: What are the issues for leadership? *Leading and Managing*, 16(2), 46–61.

- De Nobile, J. (2018). Towards a theoretical model of middle leadership in schools. *School Leadership & Management*, 38(4), 395–416. <https://doi.org/10.1080/13632434.2017.1411902>
- De Nobile, J. (2019). The roles of middle leaders in schools: Developing a conceptual framework for research. *Leading and Managing*, 25(1), 1–14.
- Educational Review Office. (2015). Raising student achievement through targeted actions. Retrieved from <https://www.ero.govt.nz/assets/Uploads/ERO-15298-Raising-Student-Achievement-FULL-v6b-WEB-002.pdf>
- Elley, W. (2001). *STAR Supplementary Test of Achievement in Reading: Years 4–6*. Wellington, New Zealand: New Zealand Council for Educational Research.
- Elley, W., Ferral, H., & Watson, V. (2014). *Star reading test teacher manual* (2nd ed.). NZCER Press.
- Fawson, P. C., Ludlow, B. C., Reutzel, D. R., Sudweeks, R., & Smith, J. A. (2006). Examining the reliability of running records: Attaining generalizable results. *The Journal of Educational Research*, 100(2), 113–126.
- Flanagan, J. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358.
- Graesser, A., Fiore, S., Greiff, S., Andrews-Todd, J., Foltz, P., & Hesse, F. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92. <https://doi.org/10.1177/1529100618808244>
- Hare, W. (2003). The ideal of open-mindedness and its place in education. *Journal of Thought*, 38(2), 3–10
- Hannah, D., Sinnema, C., & Robinson, V. (2019). Theory of action accounts of problem-solving: How a Japanese school communicates student incidents to parents. *Management in Education*, 33(2), 62–69. <https://doi.org/10.1177/0892020618783809>
- Hannah, D., Sinnema, C., & Robinson, V. (2021). Understanding curricula as theories of action. *The Curriculum Journal*. <https://doi.org/10.1002/curj.138>
- Harris, A., & Jones, M. (2017). Middle leaders matter: Reflections, recognition, and renaissance. *School Leadership & Management*, 37(3), 213–216. <https://doi.org/10.1080/13632434.2017.1323398>
- Harris, A., Jones, M., Ismail, N., & Nguyen, D. (2019). Middle leaders and middle leadership in schools: Exploring the knowledge base (2003–2017). *School Leadership and Management*, 39(3–4), 2–23. <https://doi.org/10.1080/13632434.2019.1578738>
- Hattie, J. (2015). *What doesn't work in education: The politics of distraction*. London, UK: Pearson.
- Heidemeier, H., & Moser, K. (2009). Self–other agreement in job performance ratings: A meta-analysis test of a process model. *Journal of Applied Psychology*, 94(2), 353–370. <https://doi.org/10.1037/0021-9010.94.2.353>
- Huber, S. G. (2011). Leadership for learning–earning for leadership: The impact of professional development. In T. Townsend & J. MacBeath (Eds.), *International handbook of leadership for learning–Part 1* (Vol. 25, pp. 635–652). Springer.
- Kim, J. S., Burkhauser, M. A., Mesite, L. M., Asher, C. A., Relyea, J. E., Fitzgerald, J., & Elmore, J. (2020). Improving reading comprehension, science domain knowledge, and reading engagement through a first-grade content literacy intervention. *Journal of Educational Psychology*, 113(1), 3–26. <https://doi.org/10.1037/edu0000465>
- Kin, T. M., & Kareem, O. A. (2019). School leaders' competencies that make a difference in the era of Education 4.0: A conceptual framework. *International Journal of Academic Research in Business and Social Sciences*, 9(5), 214–225. <https://doi.org/10.6007/IJARBS/v9-i4/5836>
- Lai, M. K., McNaughton, S., Amituanai-Toloo, M., Turner, R., & Hsiao, H. (2009). Sustained acceleration of achievement in reading comprehension: The New Zealand experience. *Reading Research Quarterly*, 44(1), 30–56. <https://doi.org/10.1598/RRQ.44.1.2>
- Lam, T., & Klockars, A. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Management*, 19(4), 317–322. <https://doi.org/10.1111/j.1745-3984.1982.tb00137.x>
- Le Fevre, D., Robinson, V., & Sinnema, C. (2014). Genuine inquiry: Widely espoused yet rarely enacted. *Educational Management Administration and Leadership*, 43(6), 1–17. <https://doi.org/10.1177/1741143214543204>
- Leech, N., & Onwuegbuzie, A. (2009). A typology of mixed methods research designs. *Quality and Quantity*, 43(2), 265–275.
- Leithwood, K., Reid, S., Pedwell, L., & Conner, M. (2011). Lessons about improving leadership on a large scale: From Ontario's leadership strategy. In T. Townsend & J. MacBeath (Eds.), *International handbook of leadership for learning–Part 1* (Vol. 25, pp. 337–354). Springer.
- Leithwood, K., & Stager, M. (1989). Expertise in principals' problem solving. *Educational Administration Quarterly*, 25(2), 126–161. <https://doi.org/10.1177/0013161x89025002003>

- Leithwood, K., & Steinbach, R. (1991). Indicators of transformational leadership in the everyday problem solving of school administrators. *Journal of Personnel Evaluation in Education*, 4, 221–244.
- Leithwood, K., & Steinbach, R. (1992). Improving the problem-solving expertise of school administrators: Theory and practice. *Education and Urban Society*, 24(3), 317–345.
- Mabe, P., & West, S. (1982). Validity of self evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280–296. <https://doi.org/10.1037/0021-9010.67.3.280>
- Marcy, R., & Mumford, M. (2010). Leader cognition: Improving leader performance through causal analysis. *The Leadership Quarterly*, 21(1), 1–19. <https://doi.org/10.1016/j.leaqua.2009.10.001>
- Meric, H., & Wagner, J. (2006). Rating scale format choices for multi-item measures: Does numbering and balanced-ness matter? *B>Quest*. Retrieved from <http://www.westga.edu/~bquest/2006/rating.pdf>
- Ministry of Education. (2014). Aspiration and achievement: Education system: Briefing to incoming minister. Retrieved from <https://assets.education.govt.nz/public/Documents/Ministry/Publications/Briefings-to-Incoming-Ministers/AspirationAndAchievementEducationSystem.pdf>
- Mintrop, R., & Zumpe, E. (2019). Solving real-life problems of practice and education leaders' school improvement mind-set. *American Journal of Education*, 125(3), 295–344. <https://doi.org/10.1086/702733>
- Mourshed, M., Chijioke, C., & Barber, M. (2010). How the world's most improved school systems keep getting better. Retrieved from <https://www.mckinsey.com/industries/public-and-social-sector/our-insights/how-the-worlds-most-improved-school-systems-keep-getting-better>
- Mumford, M. D., Zaccaro, S. J., Harding, F. D., Jacobs, T. O., & Fleishman, E. A. (2000). Leadership skills for a changing world: Solving complex social problems. *Leadership Quarterly*, 11(1), 11–35. [https://doi.org/10.1016/S1048-9843\(99\)00041-7](https://doi.org/10.1016/S1048-9843(99)00041-7)
- Nickles, T. (1981). What is a problem that we might solve it? *Synthese*, 41(1), 85–118.
- Parkin, C., & Parkin, B. (2020). PROBE BLUE reading comprehension assessment manual. Wellington, New Zealand: Triune Initiatives.
- Patuawa, J., Robinson, V., Sinnema, C. & Tzu, T. (2021). Addressing inequity and underachievement: Middle leaders' effectiveness in problem solving. *Leading and Managing*, 27(1), 51–78.
- Pont, B., Figueroa, D., Zapata, J., & Fraccola, F. (2013). Education policy outlook: New Zealand Retrieved from https://www.oecd.org/education/EDUCATION%20POLICY%20OUTLOOK%20NEW%20ZEALAND_EN.pdf
- Rey, D., & Neuhäuser, M. (2011). Wilcoxon-signed-rank test. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1658–1659). https://doi.org/10.1007/978-3-642-04898-2_616
- Robinson, V., Sinnema, C., & Le Fevre, D. (2014). From persuasion to learning. *Leadership and Policy in Schools*, 13(3), 260–296. <https://doi.org/10.1080/15700763.2014.922997>
- Robinson, V., Meyer, F., Le Fevre, D., & Sinnema, C. (2015). *Leaders problem solving capabilities: Exploring the "quick fix" mentality*. Paper presented at the AERA Annual Meeting, Chicago, IL
- Robinson, V., Meyer, F., Le Fevre, D., & Sinnema, C. (2020). The quality of leaders' problem-solving conversations: Truth-seeking or truth-claiming? *Leadership and Policy in Schools*. <https://doi.org/10.1080/15700763.2020.1734627>
- Robinson, V., & Le Fevre, D. (2011). Principals' capability in challenging conversations: The case of parental complaints. *Journal of Educational Administration*, 49(3), 227–255.
- Robinson, V. (2011). *Student-centered leadership*. Jossey-Bass.
- Robinson, V. (2014). Theories of action. In D. C. Phillips (Ed.), *Encyclopedia of educational theory and philosophy* (Vol. 2, pp. 808–810). <https://doi.org/10.4135/9781483346229.n331>
- Robinson, V. (2018). *Reduce change to increase improvement*. Thousand Oaks, CA: Corwin.
- Schleicher, A. (2019). PISA 2018: Insights and interpretations. Retrieved from <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- Schwarz, R. (2017). *The skilled facilitator* (3rd ed.). John Wiley & Sons.
- Shadish, W., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalised causal inference. Belmont, CA: Wadsworth Cengage Learning.
- Sinnema, C. Ludlow, L. H., & Robinson, V. (2016). Educational leadership effectiveness: A Rasch analysis. *Journal of Educational Administration*, 54(3), 305–339. <https://doi.org/10.1108/JEA-12-2014-0140>
- Sinnema, C., Robinson, V. Ludlow, L. H., & Pope, D. (2015). How effective is the principal? Discrepancy between teachers' and principals' perceptions of principal effectiveness. *Educational Assessment, Evaluation and Accountability*, 27(3), 275–301. <https://doi.org/10.1007/s11092-014-9209-y>

- Sinnema, C., Meyer, F., Le Fevre, D., Chalmers, H., & Robinson. (2021). Educational leaders' problem-solving for educational improvement: Belief validity testing in conversations. *Journal of Educational Change*. <https://doi.org/10.1007/s10833-021-09437-z>
- Sinnema, C., Hannah, D., Finnerty, A., & Daly, A. (2021). A theory of action account of an across-school collaboration policy in practice. *Journal of Educational Change*. <https://doi.org/10.1007/s10833-020-09408-w>
- Sinnema, C., Le Fevre, D., & Robinson, V. (2013). When others' performance just isn't good enough: Educational leaders' framing of concerns in private and public. *Leadership and Policy in Schools*, 12(4), 301–336. <https://doi.org/10.1080/15700763.2013.857419>
- Smith, A., Nelley, E., & Croft, D. (2008). *PM benchmark 2 teachers resource book*. Nelson Cengage Learning.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–431.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education.
- Timperley, H. (2015). Professional conversations and improvement-focused feedback: A review of the research literature and the impact on practice and student outcomes. Retrieved from https://www.aitsl.edu.au/docs/default-source/default-document-library/professional-conversations-literature-review-oct-2015.pdf?sfvrsn=fc2ec3c_0
- Timperley, H., & Parr, J. M. (2009). Chain of influence from policy to practice in the New Zealand literacy strategy. *Research Papers in Education*, 24(2), 135–154. <https://doi.org/10.1080/02671520902867077>
- Twyford, K., Le Fevre, D., & Timperley, H. (2017). The influence of risk and uncertainty on teachers' responses to professional learning and development. *Journal of Professional Capital and Community*, 2(2), 86–100.
- Zhu, T. (2020). Building robust statistical models for categorical outcome variables in educational research (Doctoral thesis, University of Auckland). Retrieved from <http://hdl.handle.net/2292/50056>
- Weber, E. P., & Khademian, A. M. (2008). Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. *Public Administration Review*, 68(2), 334–349.

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