



Shifting the focus of research on effective professional development: Insights from a case study of implementation

Sally Patfield¹ · Jennifer Gore¹ · Jess Harris¹

Accepted: 13 October 2021 / Published online: 29 October 2021
© The Author(s) 2021

Abstract

Globally, teacher professional development is heralded as a key mechanism for educational reform. With governments investing heavily in PD programs, the aim of these interventions is not only enhanced teacher knowledge and practice but, ultimately, improved student outcomes. A substantial body of research has attempted to identify characteristics of effective PD, generating a growing list of features that ostensibly ‘work’. As such, program design has become the dominant analytic focus. In this paper, we shift attention squarely to program implementation as necessary in conceptualising and evaluating effective PD. We apply the lens of implementation science to a case study of how one regional secondary school in NSW, Australia, implemented a robust PD program called Quality Teaching Rounds that has strong evidence of effectiveness. Despite the school’s attempts to remain true to the spirit of the PD, a combination of remoteness, lack of casual relief teachers, high teacher turnover, and negative perceptions of peer observation result in a form of QTR that is almost unrecognisable from its intended design. We argue greater attention must be given to understanding and supporting successful implementation within and across diverse school contexts in order to take effective forms of PD to scale.

Keywords Teacher professional development · Implementation · Effective professional development · Implementation science · Quality Teaching Rounds · Educational reform

✉ Sally Patfield
sally.patfield@newcastle.edu.au

¹ Teachers and Teaching Research Centre, School of Education, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

Introduction

Globally, policymakers and governments have set ambitious targets for educational reform. While improvement agendas vary widely across nations and jurisdictions, two key commonalities are embedded within most large-scale reform efforts. First, improving student outcomes is positioned as an important goal, spurred on by both large-scale international assessments that underpin global comparisons of performance and social justice imperatives to alleviate disparities in achievement (Meissel et al., 2016). Second, teachers are unequivocally positioned as fundamental to—even inseparable from—reform. Teachers are crucial ‘enactors’ of educational policy (Ball et al., 2012) and, ultimately, the facilitators of any changes to classroom practice (Borko, 2004; OECD, 2019).

Accordingly, we have seen significant investment in teacher professional development (PD), heralded as a key catalyst for improving student outcomes. PD is now ‘big business’ (Hill, 2009), with governments and educational jurisdictions investing heavily in a host of initiatives and interventions, varying in scope and content (OECD, 2019). While a wide range of learning experiences fall under the umbrella of ‘PD’ (Hill et al., 2013; OECD, 2019; Wei et al., 2009), scholarly attention has increasingly been directed at ‘effective PD’—“structured professional learning that results in changes in teacher practices and improvements in student learning outcomes” (Darling-Hammond et al., 2017, p. v). As such, it is now commonplace for the ‘final test’ of PD to be whether or not an intervention leads to better academic outcomes for students—not just teachers’ knowledge, skills or pedagogy (Darling-Hammond et al., 2017; Desimone, 2011).

In this paper, our focus is ‘effective PD’, as distinct from the broader spectrum of activity conceptualised as ‘teacher PD’ or ‘professional learning’ (PL). In light of the current climate of reform and desire for strong return on investment by governments (Gore et al., in press), research on effective PD has flourished over the past decade, triggering a methodological shift from small-scale studies using teacher self-reports of satisfaction and change, to experimental designs measuring student outcomes (Hill et al., 2013). Despite this agenda, however, effective PD—as measured by student achievement—remains somewhat elusive, with many studies failing to demonstrate positive gains in academic outcomes and/or criticised for lacking scientific rigour (Borko et al., 2010; Yoon et al., 2007).

In order to better understand effective PD, the dominant line of enquiry focuses on program design (Darling-Hammond et al., 2017; Hill et al., 2013). Such studies seek to identify the features of PD initiatives associated with positive gains in teacher knowledge and practice and, most importantly, student outcomes. These features include: a focus on discipline-specific content knowledge and pedagogy; sustained duration; coaching; collaboration; opportunities for feedback and reflection; and active learning (Darling-Hammond et al., 2017; Desimone, 2009, 2011; Garet et al., 2001). This area of research has risen in prominence to such an extent that some scholars refer to an informal consensus on the core characteristics of PD that ‘work’ (Desimone, 2009, 2011), or what others describe as a

‘new orthodoxy’ grounded in the view that for PD to be effective, these specific features must be included (Gore et al., in press).

We contend, however, that this consensus is problematic for a variety of reasons, not least because of the weak evidence—often based more on conjecture than empirical evidence—underpinning its claims (Gore et al., in press; Sims & Fletcher-Wood, 2021). Indeed, even when studies have used rigorous randomised controlled trial (RCT) designs, PD encompassing many of these features has rarely shown success in improving student outcomes (Gore et al., 2021; Hill et al., 2013; Yoon et al., 2007). Effective PD also often works as a ‘package,’ such that it is difficult to isolate which specific design features are important when an intervention is successful, or how particular features work together to engender positive outcomes (Hill et al., 2013; Opfer & Pedder, 2011).

Furthermore, the so-called consensus has not attended carefully to context, instead bringing together characteristics of interventions that have worked ‘somewhere’ for ‘someone’ (Bryk, 2015). This kind of generality does little to illuminate how effective forms of PD will translate into outcomes across diverse school communities and student populations, and contexts with different political, social, cultural and material elements (Ball et al., 2012). The importance of context is already well-established in many comparable fields of research, including policy enactment (Ball et al., 2012), school reform (Datnow et al., 2002) and the use of professional learning communities for improvement (Wenger, 1998). It is surprising, then, that research on effective PD has largely ignored the context of implementation, giving much greater credence to program design. As a result, little analytic attention has been paid to examining how effective PD can be implemented across diverse settings (Borko, 2004; Borko et al., 2010) and how, therefore, implementation might be conceptualised and evaluated across sites.

This paper offers precisely this kind of analysis through a case study of the implementation of one form of rigorously-tested, effective PD called Quality Teaching Rounds (QTR). Under RCT conditions, QTR has already produced significant positive effects for both teachers and students, including notable increases in teacher morale, teaching quality, and student academic achievement (Gore et al., 2017; Gore et al., 2021). However, less is known about how to support high-quality implementation in diverse contexts outside of research settings or how to support different kinds of school communities to successfully implement QTR.

With a focus on depth and particularity rather than breadth, we adopt a case study approach to examine the implementation of QTR in one school community—Olsen Valley High School (pseudonym)—located in the state of New South Wales, Australia. The analysis is anchored in ‘implementation science’, an approach that is embedded in clinical, health and community-based research (Moir, 2018) but a relatively recent phenomenon in education (Centre for Evidence & Implementation, 2017). Specifically, we draw on Proctor et al.’s (2011) heuristic of eight implementation outcomes—acceptability, adoption, appropriateness, feasibility, fidelity, cost, penetration, and sustainability—to conceptualise how QTR was implemented at Olsen Valley High and consider the merits of drawing on implementation science to evaluate the implementation of effective PD more broadly. We begin with a brief discussion of implementation science, its value to the study of effective PD and the

specifics of Proctor’s heuristic. Next, we present an overview of QTR, the intervention that forms the basis of the paper.

Implementation science and effective PD

Implementation science has been defined as the “scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice” (Eccles & Mittman, 2006, para 2). It draws attention to the temporal gap between research and practice, with its focus on how evidence-based interventions are adopted and sustained in real-world contexts (Bauer et al., 2015; Thomas et al., 2017). Applied to the field of effective PD, we can ask how *are* evidence-based PD interventions implemented within schools and systems and, thus, how *can* they be implemented to increase their effectiveness and maximise outcomes (Kelly, 2012)?

In contrast to research that synthesises and generalises core program features of effective PD, implementation science is fundamentally concerned with the specificity of context. That is, the implementation of an intervention is seen to be entwined with the unique set of circumstances associated with where—and even when—it takes place (Damschroder et al., 2009; Kelly, 2012). In this way, context is positioned as active in shaping both program implementation *and* program outcomes; it is much more than just a passive backdrop for an intervention (Datnow et al., 2002). Although context is defined in varying ways within implementation science theories and frameworks, the ideas of ‘outer’ and ‘inner’ setting (Damschroder et al., 2009) are particularly applicable to school contexts. The outer setting encompasses the economic, political, social, and cultural climate in which a school is situated while the inner setting focuses attention on the characteristics of the school itself (Damschroder et al., 2009). The distinction between these two layers is somewhat arbitrary, however, as their interaction is often dynamic, permeable, and reciprocal.

In this light, the implementation of any intervention can be thought of as the “product of the context in which it is implemented” (The Design-Based Research Collective, 2003, p. 5). Consideration of context is not new in education (see, for example, Ball et al., 2012; Datnow et al., 2002). However, the systematic application of principles of implementation science to the investigation of implementation quality is in its infancy. The potential value of applying such a lens in studies of effective PD lies in the notion that positive gains in student outcomes are related to both the quality of an intervention *and* the quality of its implementation (Centre for Evidence & Implementation, 2017). Increasing pressure on schools to evaluate PD drawing on sophisticated standards of evidence (Desimone, 2011) makes such analysis timely.

Proctor et al.’s (2011) heuristic of eight implementation outcomes, derived from a major synthesis of implementation literature, is useful in conceptualising and evaluating implementation efforts. Table 1 sets out these outcomes as an overarching framework and set of concepts to guide implementation science research. While the heuristic is derived primarily from health and behavioural sciences research, we find it to be a valuable starting point for studying the implementation of effective PD, especially in the absence of a robust field of

Table 1 Implementation outcomes (Proctor et al., 2011)

Implementation outcome	Definition
Acceptability	The degree to which stakeholders view an intervention as acceptable or satisfactory
Adoption	The initial decision to employ the intervention
Appropriateness	The perceived fit and relevance of the intervention to the given setting
Feasibility	The extent to which the intervention can be successfully executed
Fidelity	The degree to which the implemented intervention has adhered to program protocols
Cost	The costs associated with the implementation effort
Penetration	The integration of the intervention in the given setting
Sustainability	The extent to which the intervention is maintained within ongoing operations

implementation science research in education more generally (Centre for Evidence & Implementation, 2017) and the lack of attention given to implementation within effective PD specifically (Hill et al., 2013).

By and large, Proctor et al. (2011) argue that these implementation outcomes represent the effects of deliberate and purposive actions to implement an intervention. In this way, they can be thought of as ‘preconditions’ for attaining the desired changes brought about by a specific intervention, such as positive gains in student achievement in the case of effective PD. As such, implementation outcomes can be seen as impacting other kinds of outcomes, although they are not interchangeable.

While Proctor et al. (2011) suggest that the eight implementation outcomes are discrete conceptual categories, we note a degree of overlap and correspondence between many of them. For example, the perceived *acceptability* of an intervention among stakeholders is closely related to its perceived *appropriateness* for the setting. Likewise, *penetration* and *sustainability* both relate to the integration of an intervention in a particular setting, although sustainability is generally observed further along—or even after—the implementation process.

Implementation must therefore be viewed as a process (Proctor et al., 2011), one that involves a sequence of interrelated activities over time. Indeed, in practice, each outcome is interconnected in complex and dynamic ways, such that one aspect can influence most, if not all, of the other outcomes. In our paper, we use Proctor et al.’s (2011) heuristic to examine the different phases of QTR implementation, from the initial decision to implement QTR, through to ongoing implementation efforts within the school and attempts to incorporate QTR as part of normal routine.

When an intervention is unsuccessful in practice, policy actors often move to the next idea, the next ‘fad’ or the next reform initiative that can be transplanted from ‘somewhere else’ (Bryk, 2015; Datnow et al., 2002). Instead, we take the view that developing a more robust understanding of effective PD necessitates a dual focus on both program features *and* program implementation.

The intervention: Quality Teaching Rounds

Quality Teaching Rounds (QTR) is a rigorously researched approach to PD that has been widely used in New South Wales, Australia, and increasingly adopted in other state educational jurisdictions. At its core, QTR is underpinned by four interrelated components. First, it is collaborative, with teachers working in professional learning communities (PLCs) to observe, analyse, and discuss one another's practice. Teachers in a PLC can come from any Year level, teaching specialisation, or career stage (Gore & Rickards, 2020a; Gore & Rosser, 2020b). Second, it is an approach to teaching rounds (City et al., 2009; Elmore, 2007)—similar to the idea of medical rounds—which supports teachers to discuss and develop a shared understanding of 'good teaching'. The goal is instructional improvement guided by teachers, rather than an external facilitator. Third, it uses a pedagogical framework, the Quality Teaching (QT) Model, which scaffolds the Rounds process. The QT Model provides a comprehensive set of concepts and associated language for deep, professional conversations about teaching practice (Bowe & Gore, 2017; Gore et al., 2017; Gore et al., 2021). And fourth, it is underpinned by a set of protocols that have been designed to address power relations among teachers (Gore et al., 2021), encouraging full participation, turn-taking and confidentiality by members of the PLC.

Logistically, QTR usually consists of four 'Rounds', with each Round taking place over a single day. Each Round begins with discussion of a chosen professional reading. The aim of this initial session is to support teachers to engage in professional conversation and build a sense of community within their PLC. Next, a full lesson is taught by one member of the PLC and observed by all others. Each member is required to be fully present throughout the Rounds process, and have a lesson observed on a rotational basis over the course of the Rounds. After each observation, lesson coding occurs. Each teacher in the PLC (including the 'host' teacher) codes the lesson individually using the QT Model, which consists of three overarching dimensions and 18 elements (see Table 2). To conclude the day, the PLC members discuss the observed lesson, and pedagogy more broadly, drawing on the language, concepts, and structure of the QT Model. The purpose of this final session is to support meaningful analysis of practice, which is achieved through a process whereby teachers discuss their codes and associated evidence, and try to come to agreement as a group about the appropriate code for each element of the QT Model.

Table 2 Dimensions and elements of the Quality teaching Model

Intellectual quality	Quality learning environment	Significance
Deep knowledge	Explicit quality criteria	Background knowledge
Deep understanding	Engagement	Cultural knowledge
Problematic knowledge	High expectations	Knowledge integration
Higher-order thinking	Social support	Inclusivity
Metalanguage	Students' self-regulation	Connectedness
Substantive communication	Student direction	Narrative

These codes remain confidential within the PLC and are less important than the rich conversation generated about practice.

Research design

Before commencing QTR in their schools, at least two teachers from each school attend a two-day workshop designed to support implementation. In 2019, 687 teachers who participated in a QTR workshop between July 2014 and May 2018 were sent an email by the project team which included a link to a short, online questionnaire, administered via SurveyMonkey. The questionnaire included a series of questions designed to ascertain if, and how, QTR was implemented in their schools after attending the workshop.

Overall, 177 survey responses were received from teachers at 81 schools. From this pool of responses, schools were categorised into three implementation categories: ‘QTR embedded’, where QTR was embedded throughout school processes after attendance at a workshop; ‘QTR introduced’, where QTR was implemented for some or many staff at a school, but was not yet embedded in school planning and processes; and ‘QTR discontinued’, where QTR had been implemented but discontinued at a school. Principals from each school were subsequently sent an email inviting them to participate in a study examining the implementation and sustainability of QTR, with the aim of recruiting schools from each of these three implementation categories. Six schools were recruited (two from each category), however, the two schools where QTR had been discontinued withdrew after providing organisational consent. While there are lessons to be learned from termination or suspension of an intervention and we remain interested in studying such schools as part of our broader research agenda, for this analysis, our primary interest was schools that adopted QTR.

Data collection and analysis were informed by case study methodology, aiming for richness and depth rather than breadth (Yin, 2013). Given the nature of the research, the unit of analysis was the school, with data collected from multiple participants at each site for the purposes of triangulation (Yin, 2013). All teachers within a school were invited to participate, with written consent provided by participants at the time of interview. Two researchers visited each school in late 2019 and conducted interviews with the principal and a sub-sample of volunteer teachers available on the scheduled interview dates. Interviews were semi-structured and focused on: experiences of QTR; enablers and barriers to implementation; adaptations (if any); overall impressions; and perceived impact. Interviews were audio-recorded and lasted approximately 60 minutes. Schools and participants were allocated pseudonyms to protect anonymity. Transcripts were coded using the NVivo 12 software program, drawing on a two-step case-oriented approach to analysis (Yin, 2013): (1) open coding, where a line-by-line reading of each transcript was undertaken to define and develop categories or ‘nodes’; and (2) abstraction and interpretation, where nodes were grouped, and subsequently reduced, at higher levels of meaning.

This paper focuses on one school community only, Olsen Valley High School, where interviews were conducted with the Principal and eight teachers from a wide array of subject specialisations. Olsen Valley High represents an example of ‘QTR embedded’, highlighting a degree of sustainability. We selected it for this analysis, however, because it represents an ‘extreme case’ (Jahnukainen, 2009) as a school that has substantially modified aspects of QTR against recommended implementation, providing a powerful opportunity to consider the implementation of effective PD more broadly. Our case study explores potential benefits of using implementation science for evaluating and subsequently enhancing implementation, taking into account the nuances and complexities of context.

Adopting QTR: Perceptions of acceptability and appropriateness

Olsen Valley High School is a comprehensive, co-educational secondary school situated in the inland community of Olsen Valley, a regional township in NSW. Located a vast distance from major metropolitan centres, the community is geographically isolated and predominantly surrounded by desert. According to data from the Australian Bureau of Statistics, the median weekly family income is far less than the state’s average, similarly reflected in the relative socio-educational advantage of the school which is below the national mean. These characteristics are echoed in the responses of the teachers we interviewed, who describe the community as ‘remote’ and depict the student population as primarily from lower socio-economic backgrounds.

Aligning with Departmental priorities, Olsen Valley High School currently has three strategic directions centred around quality teaching, learning, and distributed instructional leadership. These goals are clearly stated in the recent school plan, which explicitly references QTR as a core whole-of-school mechanism to enhance both the quality of teaching and student outcomes. Although QTR is a relatively new practice in the school, peer observation has been a part of the school’s culture for many years, initially driven by an aim to deprivatise classroom practice:

Our current Deputy Principal, he led a team that we called the ‘lesson observation team’ and the aim of that team was to try and open the doors of classrooms because it kind of felt that teachers pretty much kept to themselves in their classrooms. And teaching being such a complex practice that nobody really went and watched anyone else teach, it was just you were teaching or you were madly preparing your stuff—you know—flat out, there was no time to go and watch someone else or take anything else in. (Rick)

Although geographically isolated, Rick emphasises the school’s attempt to overcome the professional isolation that can characterise any teaching context. The deliberate aim has been to interrupt teaching as a ‘private act’ (Cochran-Smith, 2015)—*teachers pretty much kept to themselves in their classrooms*—by making teaching more public and open. This constitutes a shift in both culture and practice—the literal *opening of*

classroom doors—by creating time and space for teachers to participate in peer observation, primarily driven at this point in time by the lesson observation team.

This effort began at Olsen Valley High with proformas to guide observation. After attending a QTR workshop, however, the leadership team became convinced that QTR would take them to a new level:

We started the lesson observation team purely with the intention to get teachers comfortable with being observed. So we had a few tools that we used that were more tick-box proformas, that they could say ‘yes that’s happening in the classroom’. It was all around rules and routines, praise and consequence and things like that. Then myself and another staff member went and did the training for QTR and came back to the school and sort of said ‘this is where we need to go. This is such a good model, we can really dive into this.’ (Jerry)

Jerry’s description of the original *tick-box* and *yes/no* observation tools used at Olsen Valley suggests a process underpinned by appraisal and judgement. Indeed, the original observations were about *rules, routines, praise, and consequence*, signalling how easily the deprivatisation of teaching can become a means of surveillance and accountability when executed without a broad understanding of the culture and ecology of a school (Charteris & Smardon, 2018; Cochran-Smith, 2015). By contrast, Jerry exalts QTR as offering greater depth—we *can really dive into this*—indicating a level of acceptability and appropriateness needed for this new approach to be adopted.

More specifically, QTR was perceived as offering the school community an explicit focus on teaching and learning. In discussing the impetus for initially adopting QTR at Olsen Valley High, Rick identifies two interrelated characteristics of QTR which he believes make it a powerful form of professional learning:

Well, it focused on teaching [and] it was a model that everybody could use that focused on improving teaching. So regardless of what level of experience... like I see the value in it and I’ve been teaching for 30 years. (Rick)

What stands out in Rick’s account is a strong belief that QTR focuses on the core business of schools and, therefore, is for *everybody*, thus adding to its acceptability and appropriateness. Observational frameworks are often subject-specific, as in mathematics, English/language arts, or science (Gore & Rosser, 2020b; Kane et al., 2013), thus narrowing the pool of teachers able to work with, and learn from, their colleagues. However, QTR’s focus on pedagogy makes it appropriate for whole-of-school implementation. Furthermore, as Rick notes, it is relevant for both beginning and experienced teachers (Gore & Bowe, 2015; Gore & Rickards, 2020a), providing an important foundation for garnering teacher buy-in to this form of PD.

Struggles with feasibility and fidelity

While QTR was perceived to be the right fit for Olsen Valley High, the school community immediately faced structural constraints that affected its feasibility and the degree to which it could be implemented with fidelity. Interestingly,

several teachers used the phrase *Rolls Royce* to signal the logistical impossibilities of implementation created by the school's context, particularly in terms of geographic isolation and the subsequent lack of casual relief teachers (CRTs) in the area:

The nature of being out here, with casual cover being non-existent, is that it's very difficult to get the scale of what we wanted with that sort of 'Rolls Royce' model. So we were lucky enough that we already had scheduled, within our teaching load, one 'professional learning' period a cycle. So we were able to use that as a sort of trade-off with QTR, in that one of those periods was designated for you to go and observe a teacher and another one of those periods was designated for you to code that lesson and then on a Tuesday afternoon staff meeting was when we would come together to do that group coding. (Jerry)

The use of the term *Rolls Royce* positions QTR as a luxury; one that is elusive, unattainable, even an *impossibility*. By contrast, Jerry's description of the Olsen Valley community emphasises a poverty of resources due to the tyranny of distance—*being out here*—especially the lack of CRTs (*non-existent*) which profoundly impacts implementation. Thus a compromise—*trade-off*—is made to balance the requirements of the PD with contextual limitations. Each Round is now conducted over a number of days (instead of during a single day), separating out the observation, individual coding, and group coding/discussion components of QTR, and removing the initial reading discussion altogether.

This substantial modification to QTR generated further implications for fidelity. To keep costs down and manage logistics, PLCs are formed based on practicality and convenience, with one teacher in each PLC being the 'host' (the observed teacher) of a given Round and the other teachers designated observers due to their professional development period (or 'free period') being timetabled at the same time. Unfortunately, this means that only one Round occurs per term and that a PLC only functions for this single Round:

[After completing one Round] our groups changed. And I didn't realise the groups would change. Thinking the Executive went first—"Oh, that's really good". And then [thinking that] one of us will be next... But after [the Round] we were talking and they're like, "Oh, you won't get to see us, because we'll be in a different group"... Personally, I like staying [in the PLC]. I just think it would be nice to see that Head Teacher that we watched. The purpose [is meant to be that] they went first. But then they didn't see us in return....like you watch a lesson, and you come in, and you do your coding, and you do [the] group code, but then that's it. (Holly)

Having participated in QTR at her previous school, Holly was very surprised to find PLCs at Olsen Valley High would not be sustained or reciprocal (*they didn't see us in return*). Thus, the very basis of QTR—teaching rounds within PLCs—is interrupted (Gore et al., 2017); observation is not mutual, there is no time to develop a group identity (...*then that's it*) and commitment to the PLC is

limited (*because we're in a different group*). These core elements of QTR, typically involving PLC members engaging in mutual observation and ongoing collaboration, were designed to flatten power hierarchies in observation and build a sense of community (Bowe & Gore, 2017). However, at Olsen Valley, Holly describes a QTR experience that is reduced to a simple, single lesson observation—just with a different conceptual lens from the proformas used previously at the school.

Another major adaptation has been the separation of the lesson observation from the individual coding; two components of QTR that are usually undertaken on the same day. Most teachers at Olsen Valley High have to return to their own classrooms straight after conducting an observation, creating a substantial time lag in the process:

It's recommended obviously to do it [the coding] straight away so it's fresh in the mind. At the same time we can't control everyone's free periods and give them two periods off or something to do it... [But] you don't want to leave it too long. And I've found that personally I have done that before and either forgotten about doing it or had a lesson straight after and then didn't have a free period until the next day or something like that. Then I found that quite difficult trying to remember what the lesson was about and code it properly. (Arnold)

Unlike more superficial forms of observation, QTR requires teachers to assign codes to elements of an observed lesson and note associated evidence as a means to collaboratively analyse and discuss practice. As such, being given the time to individually code a lesson is particularly important to facilitating QTR discussions. Importantly, the coding process is conceptualised as a means to an end; a scaffold to generate analytical dialogue (Bowe & Gore, 2017; Bowe, 2016) rather than a quantitative measure of teaching performance (Kane et al., 2013). The time gap between the observation and coding at Olsen Valley, however, is an imposed structural constraint (*we can't control...*) which leads to teachers like Arnold *forgetting* to do the coding or finding it *difficult to remember what the lesson was about*, despite taking notes during the observation.

Similarly, the value of the group coding and discussion—the final component of QTR—also appears to be diminished. This process has been adapted and compressed into a regularly scheduled staff meeting at the end of each term:

They want it to be the hour length, but they were finding that you could never do all dimensions in that amount of time, or the 50-minutes length sorry, because that's the period, for 52 minutes. So yeah, they reduce the amount [of elements] that you do. They went through the school—"What's the most important ones? Well, the top row [see Table 1] is the most important for the school." Then it's, "What do you want to benefit the most from?" And you're meant to pick the ones that you go, "Okay, this one I want to up the most." So, you'd never pick the one that you're going to score a 'one' that wasn't a part of the lesson. So, you're meant to pick the ones that you want to improve the

most and the observers are then picking the ones that they think are interesting. (Pat)

The Quality Teaching Model that underpins QTR represents a holistic and comprehensive framework that is designed to honour the complexity of teaching (Gore, 2021; Gore et al., 2017). Yet Pat's description of the modifications adopted at Olsen Valley High turns the elements comprising the Model into a set of choices, as emphasised by Pat's repeated reference to *picking the ones* that will be discussed. Here the complex practice of teaching is *reduced* to a number of elements, rather than the sum of its parts. Although selection is based on teachers' perceptions of *importance, benefit, and needed improvement*, they are unlikely to explore the multi-dimensional nature of teaching to the extent they do when all 18 elements of the Quality Teaching Model are addressed, again raising issues of implementation fidelity.

Moving forward: Penetration and sustainability

Unsurprisingly, it has been difficult for Olsen Valley High to build momentum integrating QTR into the school's PD program. The school experiences very high levels of teacher turnover, and is seen by some long-term staff as a momentary stopover in a career:

The majority of our teachers are early career teachers in the first or second year of teaching... Because of the nature of the incentive transfer system people will come to Olsen Valley to get a permanent job—they will do their three years, and then they'll transfer back to family on the coast... Unfortunately we kind of see ourselves as a bit of a factory for teachers in that we put lots into them, we produce teachers that go out to other areas of the State and they're just really hitting their straps by the time they leave us. So yeah, but if that's what we're doing, that's what we're doing I suppose and then the next batch comes in. (Rick)

Rick's *factory* metaphor highlights the extent of teacher attrition at Olsen Valley High: teachers usually arrive early in their careers, are shaped and fashioned through PD and other opportunities at the school, and then leave. The cycle continues as *the next batch comes in*. In NSW, graduates often wait years for a permanent teaching appointment, especially in coastal regions. Because of the government's *incentive transfer system* that allots higher 'transfer point ratings' to harder-to-staff schools—usually in rural and remote areas, like Olsen Valley—these schools can represent an attractive stepping stone in securing a permanent job.

In some years, almost half of the school's teaching staff have left at the end of the year to work elsewhere. This has meant an unrelenting process of inducting new staff, including almost having to start from scratch every year with QTR:

You can't create a small community because a small community is constantly changing. It might last for a year but every year there might be 17

teachers leaving. So, you can't get the strong group to stick together... Then it's just the whole re-educating people all the time on what you need to do because you've constantly got large numbers of new teachers all the time. So, every single year, it's like, "Okay, we're going to have to do training to refresh, but also we've got all these new teachers we've got to induct." (Pat)

In speaking about the staff culture at Olsen Valley High, most teachers raised the challenges posed by high attrition. Pat's comments illustrate just how taxing it can be for a relatively small teaching community to lose staff—it is not only *constantly changing* but *large numbers* of teachers are leaving *all the time*—thus inhibiting the development of an ongoing community of practice around the experience of QTR. These circumstances signal an underlying difficulty in sustaining QTR in a school where it is hard to build a community of teachers who are invested, support each other and *stick together*.

While the school has mechanisms in place to induct new staff into QTR, the high turnover also brings the related challenge of overcoming pre-existing negative attitudes about lesson observation:

Mainly with new staff there is still that stigma about 'is this performance-based?' And that's probably the first thing that we do every year is to try and break down those stigmas that "hey, this is not a performance-based thing. There's no one judging you as a teacher" because everyone has their own opinions about what a quality teacher is. But this gives us a really good framework to look at what quality teaching is. So there's so much more to being a teacher than just being in front of a class and what's going on in the classroom. So by no means are we saying "this is you as a teacher". It's all about the *teaching*. (Jerry)

Lesson observation has increasingly been embraced in Australian classrooms as a means to improve practice. Globally, however, observation is used for both 'low-stakes' purposes—such as self-reflection and formative feedback—and high-stakes purposes—namely, decisions about remuneration, tenure, and dismissal (Cohen & Goldhaber, 2016). In an era of accountability, where teachers are often positioned as 'performance workers' (Ball, 2003), it is not surprising that some are wary of observation. However, as Jerry eloquently explains, QTR is the antithesis of this view—it focuses on *teaching*, not *teachers*—which is why it was initially viewed as having a high level of appropriateness for the school. Indeed, in this light, Jerry makes an important distinction between teaching as a practice (*it's all about the teaching*) and the traits of individual teachers (*by no means we are saying... this is you as a teacher*). But with such high turnover rates, there is an added layer of effort to constantly have to remove the stigma attached to lesson observation more broadly.

More recently, the sustainability of QTR has also been hindered by the current political environment in the local community, when the union intervened at another local school where teachers made complaints about observation more generally. Ultimately, this set of circumstances has triggered further adaptations to how QTR is implemented at Olsen Valley:

This year it's been really different because we've had the same teachers observed twice. Because there hasn't been as many people put their hand up. I'm not sure if that's representative of—we've got a lot of new teachers who are sort of like, "we don't want to be observed yet, we don't really understand the process". Particularly that's what happened in our KLA [Key Learning Area]. I sort of said to our new staff member, "do you want to get observed?" and he's like "not yet because I don't understand what this is". (Racquel)

Unfortunately, that's where I'm not happy with it because, at the moment, we have the same sort of ten to fifteen staff members volunteering each term. And in my point of view, that's not how it should be run. Everyone should be having a turn at hosting... To me that's not the ideology of the process. You should be hosting a Round if you're going and watching other people. You should be comfortable enough to have them come and watch you as well. (Jerry)

With the same group of teachers now repeatedly observed by their colleagues, there is an imbalance in the way QTR operates at Olsen Valley, affecting both feasibility and fidelity. This disparity revolves around a new opt-in process, where teachers must now come forward and *put their hand up* to be observed, rather than everyone taking a turn within a four-person PLC. Jerry understands that this approach conflicts with a key premise of QTR, which is the need to build reciprocity and trust (Bowe & Gore, 2017; Gore et al., 2017); *you should be hosting a Round if you're going and watching other people*. However, when coupled with high teacher turnover, and hence fewer teachers to pass on their positive experience, new teachers are understandably hesitant to volunteer to host a Round. They are unfamiliar with QTR (*we don't really understand the process*) and cautious about being observed by strangers soon after starting at the school. Although the school leadership team is unhappy with this imposed adaptation, they are doing what they can to keep QTR running, given their perception of its value as a form of professional development.

Discussion

This paper sought to go beyond broad generalisations about effective PD by shifting the focus to program implementation. In the current climate of educational reform, the key measure of an intervention's success is increasingly the oft-elusive goal of academic achievement; that is, positively influencing student outcomes (Darling-Hammond et al., 2017; Gore et al., in press; Hill et al., 2013). To date, however, the long list of program features advocated as central planks to 'best practice' (Darling-Hammond et al., 2017; Desimone, 2011) has largely overshadowed an analytic focus on what happens at the point of PD implementation. As the case study presented in this paper illustrates, even effective and robust forms of PD will not necessarily translate into effective implementation.

Applying the lens of implementation science—specifically Proctor et al.’s (2011) heuristic of implementation outcomes—highlights both the possibilities and constraints of translating effective PD like QTR across diverse contexts. The initial adoption of QTR at Olsen Valley High was underpinned by the best of intentions—to deprivatise classroom practice using a rigorous observational framework. Indeed, QTR was perceived to be both acceptable and appropriate among staff and leaders at the school, particularly when compared to the observational tools previously used. However, the combination of remoteness, lack of CRTs in the area, high teacher turnover, and negative perceptions of lesson observation (both locally and more broadly within the teaching profession) had major repercussions for all other implementation outcomes (Proctor et al., 2011)—feasibility, fidelity, cost, penetration, and sustainability—ultimately resulting in a form of QTR that is almost unrecognisable from its intended design.

Adaptation to create a better fit between an intervention and local conditions is a widely acknowledged need of educational reform at scale (Borko, 2004; Datnow et al., 2002; Quinn & Kim, 2017). However, extreme variation in implementation, such as occurred at Olsen Valley, highlights how program integrity can be lost through modifications to, and removal of, core components. First, the reading discussion was removed, thus limiting the building of community through shared engagement in professional discussion of ideas. Second, each Round took place over a number of days—often weeks—rather than a single day, losing coherence in the process of observing, coding, and discussing a lesson. Third, PLCs were not sustained for a set of Rounds (typically four days spread over a period of weeks), instead operating fleetingly based on teaching schedules and availability, thereby limiting the building of trust among participants and reciprocity that comes from observing and analysing each other’s teaching. Finally, the coding discussion was both compressed in terms of time allocated and reductionist in the selection of elements, losing adherence to the protocol of addressing all dimensions and elements in the Quality Teaching Model. This adaptation limits the deep professional learning that comes from comprehensive analysis and discussion of teaching practice.

While it might therefore be easy to characterise this case study as an ‘implementation failure’ (Thomson, 2014), we argue that the lens of implementation science helps to understand the situation differently. Thomson (2014) argues that ‘failure’ in educational reform is too often attributed to those involved in the implementation—teachers, leaders, schools—or to the context itself, which is blamed for posing too many difficulties. It is certainly true that the structural limitations faced by Olsen Valley have had serious consequences for the uptake and sustainability of QTR. However, we neither see the environment as ‘too difficult’ for implementation nor do we criticise the people involved. Instead, implementation science offers a framework to systematically assess the outcomes of implementation and identify what is needed to enhance the effectiveness of PD.

When implemented with fidelity, QTR has wide-ranging benefits for both teachers and students (Bowe & Gore, 2017; Gore et al., 2017; Gore et al., 2021; Gore & Rickards., 2020a; Gore & Rosser., 2020b). We ask, therefore, how can these benefits be realised in school communities like Olsen Valley? One possibility is ‘QTR Digital’, a modified version of QTR recently trialled specifically for regional and remote

contexts, utilising digital technologies to support implementation. This form of QTR is comprised of the same core features, but with a few critical changes to support uptake and sustainability. In particular, teachers digitally video record a lesson to be observed by the other members of their PLC—rather than the observation occurring face-to-face—supporting implementation in schools that face issues with securing CRTs and/or facilitating teacher release. Teachers can also form PLCs either *within* or *across* schools, which further supports small schools and schools situated in remote areas that struggle to release four teachers for a full set of Rounds.

At a policy level, the recent announcement by the NSW Department of Education of a trial to permanently employ CRTs to ‘cover’ classes in regional and remote areas of the state (NSW Department of Education, 2020) is a critical step to support the implementation of effective PD. Such a strategy provides not only a means for teachers to engage in effective PD, like QTR, but creates an incentive for casual teachers to teach in hard-to-staff areas by appointing them to a permanent position. Given that this trial just commenced at the time of writing, future research into how it might impact the implementation of effective PD in regional and remote areas of Australia is an important area for further investigation.

In sum, many of these implementation challenges are not new in the field of PD or in educational reform more broadly (Ball et al., 2012; Datnow et al., 2002; Hill et al., 2013). However, implementation science offers a new mechanism for conceptualising, evaluating and enhancing implementation through a systematic focus on context. It is insufficient to examine changes in teachers’ knowledge and practice or even student achievement associated with different program features. Instead, for PD to be effective, both the design of a program and the quality of its implementation are critical (Centre for Evidence & Implementation, 2017). When ‘effective PD’ fails to improve student outcomes, it may well be because implementation has deviated from best practice (Hill et al., 2013). Heuristics such as Proctor et al.’s (2011) can be highly beneficial by providing evidence with which to support implementation at scale. We argue that a dual focus on program features and program implementation is critical in the ongoing quest for effective PD.

Conclusion

Empirically and conceptually, this case study has significant implications for the study of effective PD. Rather than continuing to showcase interventions that ‘work’ and pinpoint their key design features, there is clearly an unmet and critical need to understand how programs are—and can be—implemented across diverse school contexts and how, in turn, implementation can be evaluated and enhanced. As Bryk (2015) argues, “the latter is what practitioners typically want to know—what will it take to make it work for *me*, for *my* students, and in *my* circumstances?” (p. 469; emphasis added). The money invested in PD—estimated to be billions of dollars annually (Kraft et al., 2018)—demands that, in going forward, understandings of effective PD must be accompanied by knowledge of effective implementation.

Funding This research was funded by the Paul Ramsay Foundation and the NSW Department of Education. We wish to thank the schools and teachers involved in the research, as well as the project staff involved in data collection. We particularly wish to thank Claire Wallington for her oversight in project management.

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

- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. Taylor & Francis.
- Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*. <https://doi.org/10.1186/s40359-015-0089-9>
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. <https://doi.org/10.3102/0013189x033008003>
- Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (Vol. 7, pp. 548–556). Elsevier.
- Bowe, J. (2016). *Quality Teaching Rounds: Strengthening the knowledge base and collaborative processes for teacher professional development [dissertation]*. The University of Newcastle, Newcastle, Australia. Retrieved from <https://nova.newcastle.edu.au/>
- Bowe, J., & Gore, J. (2017). Reassembling teacher professional development: The case for Quality Teaching Rounds. *Teachers and Teaching*, 23(3), 352–366. <https://doi.org/10.1080/13540602.2016.1206522>
- Bryk, A. S. (2015). 2014 AERA distinguished lecture: Accelerating how we learn to improve. *Educational Researcher*, 44, 467–477. <https://doi.org/10.3102/0013189x15621543>
- Centre for Evidence and Implementation. (2017). *Implementation in education: Findings from a scoping review*. Centre for Evidence and Implementation. Retrieved from <https://evidenceforlearning.org.au/>
- Charteris, J., & Smardon, D. (2018). "Professional learning on steroids": Implications for teacher learning through spatialised practice in new generation learning environments. *Australian Journal of Teacher Education*, 43(12), 12–29. <https://doi.org/10.14221/ajte.2018v43n12.2>
- City, E., Elmore, R., Fiarman, S., & Teitel, L. (2009). *Instructional rounds in education*. Harvard Education Press.
- Cochran-Smith, M. (2015). A tale of two teachers: Learning to teach over time. *Learning Landscapes*, 8(2), 111–134. <https://doi.org/10.36510/learnland.v8i2.699>
- Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378–387. <https://doi.org/10.3102/0013189x16659442>
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*. <https://doi.org/10.1186/1748-5908-4-50>
- Darling-Hammond, L., Hyster, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. Retrieved from <https://learningpolicyinstitute.org/>

- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. RoutledgeFalmer.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189x08331140>
- Desimone, L. M. (2011). A primer on effective professional development. *The Phi Delta Kappan*, 92(6), 68–71. <https://doi.org/10.1177/0031721711109200616>
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science*. <https://doi.org/10.1186/1748-5908-1-1>
- Elmore, R. (2007). Professional networks and school improvement. *The School Administrator*, 64(4), 20–24. Retrieved from <https://www.aasa.org/>
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945. <https://doi.org/10.3102/00028312038004915>
- Gore, J. (2021). The quest for better teaching. *Oxford Review of Education*, 47(1), 45–60. <https://doi.org/10.1080/03054985.2020.1842182>
- Gore, J., & Bowe, J. (2015). Interrupting attrition? Re-shaping the transition from preservice to inservice teaching through Quality Teaching Rounds. *International Journal of Educational Research*, 73, 77–88. <https://doi.org/10.1016/j.ijer.2015.05.006>
- Gore, J., & Rickards, B. (2020a). Rejuvenating experienced teachers through Quality Teaching Rounds professional development. *Journal of Educational Change*, Advance online publication. <https://doi.org/10.1007/s10833-020-09386-z>
- Gore, J., & Rosser, B. (2020b). Beyond content-focused professional development: Powerful professional learning through genuine learning communities across grades and subjects. *Professional Development in Education*, Advance online publication. <https://doi.org/10.1080/19415257.2020.1725904>
- Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education*, 68, 99–113. <https://doi.org/10.1016/j.tate.2017.08.007>
- Gore, J., Miller, A., Fray, L., Harris, J., & Prieto, E. (2021). Improving student achievement through professional development: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education*, 101. <https://doi.org/10.1016/j.tate.2021.103297>
- Gore, J., Patfield, S., & Fray, L. (in press). Questioning the consensus on effective professional development. In R. Tierney, F. Rizvi, G. Smith, & K. Gutierrez (Eds.), *International Encyclopedia of Education* (4th ed.). Oxford, UK: Elsevier.
- Hill, H. C. (2009). Fixing teacher professional development. *Phi Delta Kappan*, 90(7), 470–476. <https://doi.org/10.1177/003172170909000705>
- Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher*, 42(9), 476–487. <https://doi.org/10.3102/0013189x13512674>
- Jahnukainen, M. (2009). Extreme cases. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research (online)*. SAGE Publications.
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Bill & Melinda Gates Foundation. Retrieved from <https://usprogram.gatesfoundation.org/>
- Kelly, B. (2012). Implementation science for psychology in education. In B. Kelly & D. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 3–12). Cambridge University Press.
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Meissel, K., Parr, J. M., & Timperley, H. S. (2016). Can professional development of teachers reduce disparity in student achievement? *Teaching and Teacher Education*, 58, 163–173. <https://doi.org/10.1016/j.tate.2016.05.013>
- Moir, T. (2018). Why Is implementation science important for intervention design and evaluation within educational settings? *Frontiers in Education*, 3, 1–9. <https://doi.org/10.3389/educ.2018.00061>
- NSW Department of Education. (2020). *More casual teachers for regional schools*. NSW Department of Education. Retrieved from <https://www.nsw.gov.au/news/more-casual-teachers-for-regional-schools>
- OECD. (2019). *TALIS 2018 results (volume 1): Teachers and school leaders as lifelong learners*. OECD Publishing. Retrieved from <http://www.oecd.org/education/>

- Opfer, D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376–407. <https://doi.org/10.3102/0034654311413609>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health, 38*(2), 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Quinn, D. M., & Kim, J. S. (2017). Scaffolding fidelity and adaptation in educational program implementation: Experimental evidence from a literacy intervention. *American Educational Research Journal, 54*(6), 1187–1220. <https://doi.org/10.3102/0002831217717692>
- Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: A critical review. *School Effectiveness and School Improvement, 32*(1), 47–63. <https://doi.org/10.1080/09243453.2020.1772841>
- The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8. <https://doi.org/10.3102/0013189x032001005>
- Thomas, D. C., Berry, A., Djuricich, A. M., Kitto, S., Kreutzer, K. O. K., Van Hoof, T. J., Carney, P. A., Kalishman, S., & Davis, D. (2017). What is implementation science and what forces are driving a change in medical education? *American Journal of Medical Quality, 32*(4), 438–444. <https://doi.org/10.1177/1062860616662523>
- Thomson, P. (2014). ‘Scaling up’ educational change: Some musings on misrecognition and doxic challenges. *Critical Studies in Education, 55*(2), 87–103. <https://doi.org/10.1080/17508487.2014.863221>
- Wei, R. C., Andree, A., & Darling-Hammond, L. (2009). How nations invest in teachers: High-achieving nations treat their teachers as professionals. *Educational Leadership, 66*(5), 28–33. Retrieved from <http://www.ascd.org/>
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.
- Yin, R. (2013). *Case study research: Design and methods* (5th ed.). SAGE Publications.
- Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from <https://ies.ed.gov/ncee/edlabs/>

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