International Journal of Educational Research 98 (2019) 146–158

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Improving student outcomes through professional development: Protocol for a cluster randomised controlled trial of quality teaching rounds

Andrew Miller*, Jennifer Gore, Claire Wallington, Jess Harris, Elena Prieto-Rodriguez, Max Smith

Teachers and Teaching Research Centre, School of Education, University of Newcastle, Australia

ARTICLE INFO

Keywords:
Teacher development
Intervention
Professional development
Teaching rounds
Quality teaching

ABSTRACT

Translation of teacher professional development into improved student outcomes is of global interest, with experimental methods required to demonstrate potential professional development intervention effects. This protocol for a four-arm cluster randomised controlled trial is designed to test the efficacy of a structured collaborative approach to professional development called Quality Teaching Rounds (QTR). Linear mixed models will be used to compare a QTR group to two time-equivalent intervention groups, and a usual-practice control group. The primary outcomes are at the student level (reading, mathematics and science). Secondary outcomes are assessed at student level (quality of school life and aspirations), and teacher level (teaching quality, engagement, morale, teaching efficacy and collective efficacy). Qualitative methods are used to compliment quantitative analysis.

1. Introduction

Given its profound impact on individuals, families, communities, industries, and nations, schooling is appropriately subject to intense government and public scrutiny globally (Carter, 2015; Louden, 2008; Roberts-Hull, Jensen, & Cooper, 2015; Teacher Education Ministerial Advisory Group [TEMAG], 2014). Arguably, the quality of school education makes the most critical contribution to a nation's well-being and prosperity. Therefore, developing the knowledge and skills of the teaching workforce is foundational to achieving good outcomes.

The most critical in-school factor influencing student outcomes is the quality of teaching (Hattie, 2008; OECD, 2005; Rockoff, 2004; Rowe, 2003). Every year, millions of dollars are invested in teacher professional development (PD), and PD policies and practical approaches abound in every education system. However, few studies have shown rigorous evidence of their impact on the performance of teachers, let alone students (Council for the Accreditation of Educator Preparation (CAEP), 2015; Cordingley, Bell, Evans, & Firth, 2005; Darling-Hammond, 2013; Guskey & Yoon, 2009; Kennedy, 2016).

Given the size of the teaching workforce globally, an approach to teacher development that is effective in enhancing the quality of teaching—for both current and beginning teachers—has enormous potential for improving student outcomes. Quality Teaching Rounds (QTR) is a short-term PD intervention with demonstrated significant effects on teaching outcomes (d = 0.4), teacher morale

* Corresponding author at: School of Education, University of Newcastle, HA97, University Drive Callaghan NSW 2308 Australia.
E-mail addresses: andrew.miller@newcastle.edu.au (A. Miller), jenny.gore@newcastle.edu.au (J. Gore), claire.wallington@newcastle.edu.au (C. Wallington), jess.harris@newcastle.edu.au (J. Harris), elena.prieto@newcastle.edu.au (E. Prieto-Rodriguez), maxwell.smith@newcastle.edu.au (M. Smith).

https://doi.org/10.1016/j.ijer.2019.09.002
Received 15 May 2019; Received in revised form 30 August 2019; Accepted 3 September 2019

(d = 0.4) and teacher perceptions of appraisal and recognition (d = 0.4), sustained six months post-intervention (Bowe & Gore, 2017). Furthermore, prior research documented gains in confidence and skill among both early career teachers (Gore, Holmes, Smith, Southgate, & Albright, 2015) and experienced teachers who overwhelmingly report being re-energised, refreshed, and re-engaged through their participation (Bowe & Gore, 2017; Gore & Rickards., under review). Building the capacity of teachers to deliver high quality teaching using the QTR approach should translate into significant return on investment in terms of enhanced student outcomes (academic outcomes, aspirations, wellbeing). We will test this premise under experimental conditions, as per the protocol outlined in this paper.

1.1. Aims and hypothesis

The primary research question of the project is:

1 To what extent does teacher participation in QTR improve student academic achievement in numeracy, literacy and science relative to other forms of professional development?

Secondary research questions are:

2 To what extent does teacher participation in QTR improve student perceptions of their own academic ability, their relationships with teachers and the relevance of schooling relative to other forms of professional development?
3 To what extent does participation in QTR change teaching practice?
4 To what extent does participation in QTR affect teacher morale, engagement, individual efficacy, and collective efficacy?
5 Does QTR professional development delivered to teachers by trainers have the same impact on student and teacher outcomes as QTR delivered by the research team?

We hypothesise that students taught by teachers who have undertaken QTR will display greater positive achievement growth in mathematics, reading and science and hold more positive perceptions of their own academic performance, their relationships with teachers, and the relevance of their schooling, compared with students whose teachers undertake alternative approaches to PD. We also hypothesise that teachers undertaking QTR will display gains in teaching quality, and improvements in morale, engagement, individual efficacy, and collective efficacy above those of teachers undertaking the alternative forms of PD.

2. Methodology

2.1. Intervention

Quality Teaching Rounds is a PD process in which a form of “rounds” (Elmore, 2007; Goodwin, Del Prete, Reagan, & Roegman, 2015) is undertaken by a group of teachers in a professional learning community (PLC) (Lave & Wenger, 1991), with analysis and discussion guided by a pedagogical model, the Quality Teaching model (NSW Department of Education and Training (NSWDET), 2003). Groups of four (or more) teachers, working in a PLC, undertake a set of rounds together over a period of 3–6 months. A “Round” is comprised of sequential sessions that occur on a single day, involving:

1 Reading discussion: Designed to support the group in developing a shared theoretical basis for professional conversations and build a sense of professional community (typically one hour)
2 Observation: One PLC member teaches a lesson that is observed by all other members of the PLC (a full lesson length, typically 30–80 minutes); and
3 Coding and discussion: Individual coding of the observed lesson, including coding by the observed teacher, using the Quality Teaching model (NSW Department of Education & Training (NSWDET), 2003), is followed by discussion whereby all PLC members contribute (typically one to two hours).

In QTR, at least one lesson is observed for every member of the PLC. This is considered a “full” set of rounds. PLC members stay together for an entire set of Rounds. The intent of QTR is to focus on the relationship between classroom practice and student learning and to show respect for the teacher and the teaching-learning process by watching a whole lesson each time (Bowe & Gore, 2017). The use of a framework for observation and discussion provides teachers with a common language and set of conceptual standards with which to engage in rigorous diagnostic professional conversations with colleagues, whilst the protocols of turn taking during discussion seek to flatten power structures within the group and give all participants a chance to express their professional understandings of the concepts addressed in the model (Bowe & Gore, 2017; Gore et al., 2015).

Derived from work on Authentic Pedagogy (Newmann, Marks, & Gamoran, 1996) and Productive Pedagogy (Gore, Griffiths, & Ladwig, 2004; Gore, 2007; Lingard et al., 2001), the Quality Teaching framework incorporates three dimensions of pedagogy: Intellectual Quality, Quality Learning Environment, and Significance. Each dimension comprises six elements, as shown in Table 1. This pedagogical framework has been widely implemented within Australia in recent years, and has also been utilised by other researchers, who have reported the transformative capacity of the framework in terms of teachers’ practice (e.g., Aubusson, Steele, Dinham, & Brady, 2007; Ewing et al., 2010; Hammond, 2008; Penney, Brooker, Hay, & Gillespie, 2009; Plummer, Nyholm, Quince, &

Table 1
Quality Teaching model (3 dimensions and 18 elements).

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

Dione, 2010; Rushton, 2008; Treble, 2009; Whalan, 2012).

2.2. Study design

There has been a recent increase in the number of studies seeking to provide substantive quantitative data on the effects of teaching practice on student outcomes through the use of randomised controlled trials in school settings (Connolly, Keenan, & Urbanska, 2018). While this methodology is increasingly accepted as providing a new level of rigour in educational research, concerns have been raised about the low effect of many interventions and, in turn, null result of many evaluations (Lortie-Forgues & Inglis, 2019). Specific concerns raised by examinations of uninformative randomised controlled trials, include the adequacy of the evidence base for the interventions, the translation of that evidence into effective interventions, and the design of evaluations for testing interventions (Lortie-Forgues & Inglis, 2019).

QTR is founded on a strong evidence base from prior research on effective pedagogy (Ladwig, 2007; Newmann et al., 1996), translated into an intervention based on principles of effective professional development (Bowe & Gore, 2017). It is informed by a 'comprehensive' (Indig, Lee, Grunseit, Milat, & Bauman, 2018) program of prior studies, with systematic attention to processes of development, proof of concept, efficacy testing, real-world trials, and dissemination (Gore, 2018). The design of the proposed evaluation adheres to the Consolidated Standards of Reporting Trials (CONSORT) guidelines for group trials (Moher et al., 2010) to promote optimal independence of judgements, minimise contamination within the sample, gather reliable data, and increase the potential for drawing causal inferences. Table 2 outlines the timeline of the trial in relation to CONSORT guidelines, including separation of research tasks between those who are blinded to group allocation and those who have access to group allocation details. Sampling and sample size decisions were made to maximise the likelihood of obtaining an informative result (Lortie-Forgues & Inglis, 2019).

A four-arm cluster randomised controlled trial (RCT) will be undertaken to evaluate the efficacy of QTR for the improvement of student outcomes in comparison to three alternative conditions (See Fig. 1 and Alternative study conditions, below). The four arms of the study are:

1 Researcher-led QTR PD
2 Trainer-led QTR PD
3 Peer observation time-equivalent PD active wait-list control
4 PD-as-usual (PDAU) wait-list control

Inclusion of the Trainer-led QTR arm is designed to examine the potential for implementation of this PD at scale, by comparing outcomes for the Researcher-led and Trainer-led delivery of the two-day QTR workshops that precede the school-based implementation of QTR. The active control group is included to increase the capacity for causal inference by matching the time and funding levels provided for the intervention groups. We address ethical concerns by ensuring that control group participants have wait-list access to the QTR intervention within the duration of the project. This project has been approved by the University of Newcastle's Human Research Ethics Committee (Approval No. H-2018-0340) and the NSW Department of Education (NSWDOE) State Education Research Applications Process (SERAP) (Approval No. 2018458), and is funded by the Paul Ramsay Foundation and the Australian Research Council (DP180100285).

2.3. Study conditions – Four arm trial

Schools participating in the study will be randomly allocated to one of the following four experimental conditions.

2.3.1. Researcher-led QTR

In this Researcher-led QTR condition, four teachers from a school will form a PLC to participate in QTR in the 2019 school year. Prior to commencing Rounds, two teachers from the PLC will participate in a two-day QTR workshop to prepare them for conducting QTR within their school. The training workshops will provide background information on the Quality Teaching model and Quality Teaching Rounds, highlighting the intention and importance of each component of the approach (i.e., PLCs, readings, observation, individual coding, group discussion). Teachers will be given opportunities to practise the QT coding process and participate in simulated Rounds using sample video-recorded lessons. An overview of the research design will also be provided.

Table 2
Timeline for the Quality Teaching Rounds intervention.

Phase and Task	Assess Team A	Assess Team B	3rd party	Target Audience		Group Allocation			
				Teacher	Student	QTR	Trainer QTR	Active control	Control
Pre-intervention									
3 months pre-intervention									
Call for EOI	x	x							
Check eligibility	x	x							
Obtain consent	x	x		x	x*				
Assessment training	x								
1-2 months pre-intervention [a]									
Progressive Achievement Tests	x				x	x	x	x	x
Student survey	x				x	x	x	x	x
Observations	x			x		x	x	x	x
Teacher survey	x			x		x	x	x	x
Interviews	x			x		x	x	x	x
2 weeks pre-intervention									
Randomisation			x						
1 week pre-intervention									
QTR workshop		x				x	x		
Peer-observation workshop		x						x	
Intervention period (7-months)									
Intervention fidelity checks		x							
2-month teacher survey	x			x		x	x	x	x
5-month teacher survey	x			x		x	x	x	x
Post-intervention									
Immediately post-intervention[b]									
Progressive Achievement Tests	x				x	x	x	x	x
Student survey	x				x	x	x	x	x
Observations	x			x		x	x	x	x
Teacher survey	x			x		x	x	x	x
Interviews	x			x		x	x	x	x
5-months post-intervention [c]									
Teacher survey	x			x		x	x	x	x
Data analysis	x			x	x	x	x	x	x
Post 12-month follow-up									
QTR workshop		x		x				x	x

Note: Assess Team A blinded to group allocation; Assess Team B knows group allocation. QTR = Quality Teaching Rounds.
* Opt-out consent.
[a] Baseline.
[b] 7-month follow-up.
[c] 12-month follow-up.

Workshops for the schools in the intervention group are provided at no charge, with each school funded for the release of the two teachers attending training for the two-days (4 funded release days). In addition, these schools are funded for the four participating teachers to be released from class to carry out a full set of rounds across the intervention period (4 teachers x 1 day per round x 4 rounds = 16 funded release days).

2.3.2. Trainer-led QTR

Participants in the Trainer-led condition will also receive the QTR intervention, however the two-day QTR workshop will be delivered by a trainer, who has been trained by the members of the research team. QTR Trainers were recruited from a pool of exemplary teachers with experience in QTR, who responded to an employment advertisement from the NSWDOE and the University of Newcastle (UON). These trainers have undertaken 25 days of training in the theoretical foundations and practical delivery of QTR in order to prepare them to lead the two-day training workshops.

Two teachers from each school assigned to the Trainer-led QTR condition will be funded for the equivalent amount of teacher release as the Researcher-led condition and will undertake QTR in 2019. That is, other than the delivery of the QTR PD workshop by trainers, the two QTR conditions are equivalent.

2.3.3. Peer observation time-equivalent PD active wait-list control

Teachers in this condition will participate in an alternative form of PD called Peer Observation. This method of PD is endorsed by the Australian Institute for Teaching and School Leadership (AITSL), the regulatory body responsible for development of the professional standards for Australian teachers. As with the two QTR intervention groups, two of the four teachers in a PLC will be funded to attend a two-day workshop to prepare them to undertake Peer Observation in their schools, according to a specific protocol. Four participating teachers from each school in this condition will be funded for the equivalent number of release days as teachers in the intervention condition for 2019.That is, apart from a different protocol, the Peer Observation condition is equivalent to the two QTR

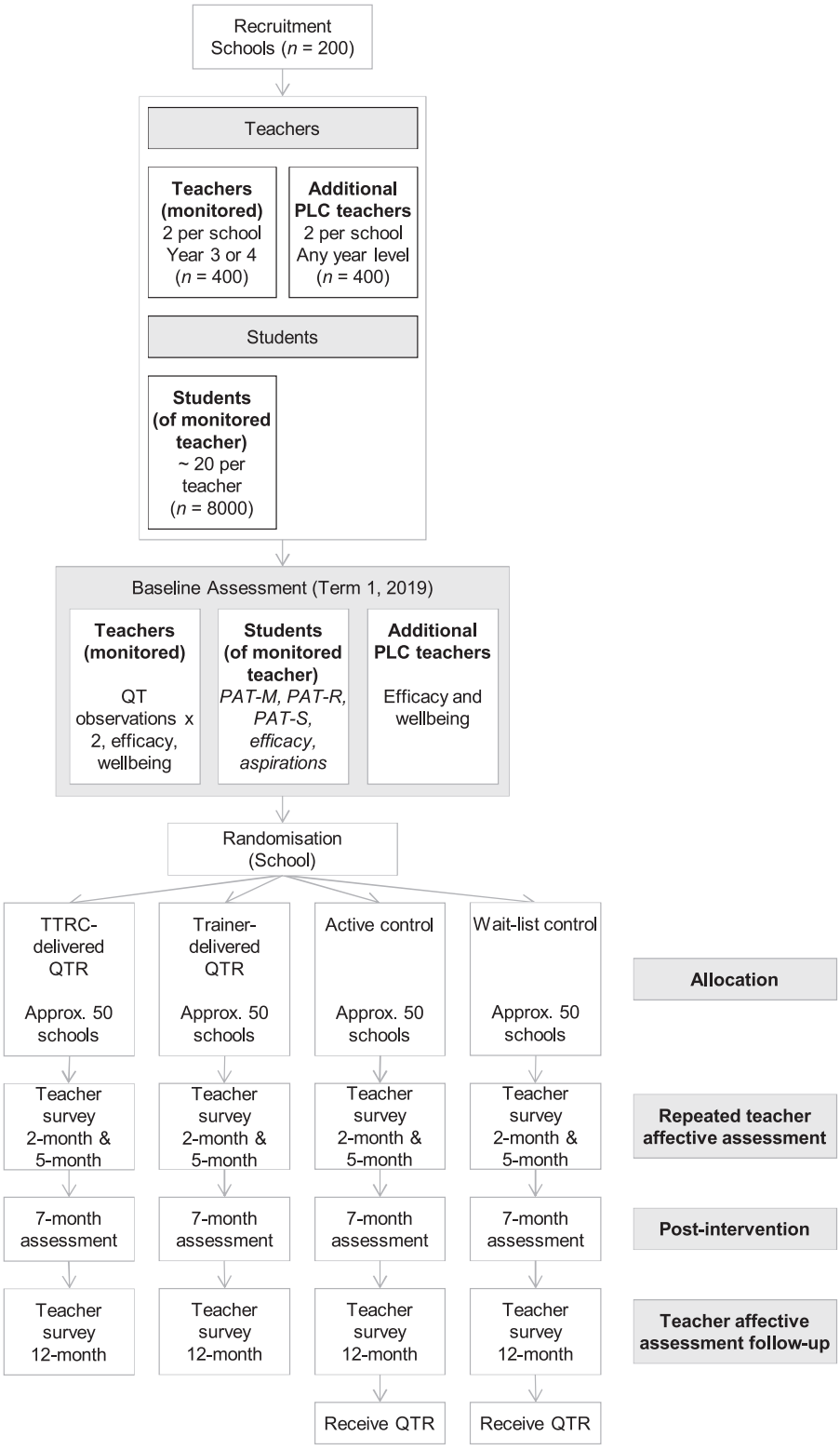

Fig. 1. Participant flowchart.

intervention conditions. In addition, this group will receive QTR PD and funding for QTR as specified for the intervention group in the school year following the intervention period (2020).

2.3.4. Professional-development-as-usual wait-list control

Schools in this condition will continue their PD as usual, with no additional funding or release time provided for PD as part of this research in 2019. This group will receive PD and funding for QTR as specified for the intervention groups in the school year following the intervention period (2020).

2.4. Sample

The study aims to involve 200 NSW primary schools. From these schools, 800 teachers (four from each school) and approximately 8000 of their students will form the sample.

In this study, the effects of QTR on student outcomes will be sought from a sample of Stage 2 students (school Years 3–4, age 8–10 years) from NSW primary schools, whose teachers are involved in the QTR intervention. Four teachers will be recruited per school to form a PLC to participate in the randomly-allocated condition. Student sampling is calculated on school level cluster sizing of approximately 40 students per cluster, drawn from two intact class groups (teacher and students). The combination of sampling strategy and intervention requirements will result in two Stage 2 teachers (monitored teachers) and their students and two teachers from any stage ('PLC' teachers) being sought from each school. An example of the PLC participant structure at a school is provided below:

- Teacher 1 (Stage 2) – Monitored teacher (Students invited)
- Teacher 2 (Stage 2) – Monitored teacher (Students invited)
- Teacher 3 (Stage 3) – PLC teacher
- Teacher 4 (Stage 1) – PLC teacher

2.4.1. Sample size calculations

Power calculations were conducted to determine the sample size required to detect changes in the primary outcomes in students (i.e. Progressive Achievement Test scores). As the goal of this study is to detect an effect for those receiving the Researcher-led QTR intervention (Group 1) compared with those in the PD-as-usual wait-list control (Group 4), calculations are based on comparison of these groups, with sampling extended to the additional arms of the study (Trainer-led QTR and Peer Observation wait-list control). In order to detect a practically significant effect (d = 0.20) of approximately 3 months additional growth (Higgins et al., 2018) approximately 788 students are required to detect an effect at 80% power with alpha 0.05. To adjust for the hierarchical structure of the data, the following correction factor is applied [1 + (m – 1) x ICC] (Donner & Klar, 2000), where m = students per class and ICC = the intra-class correlation coefficient. Assumptions are based on the hierarchical levels of students within classes, and classes within schools (two classes recruited per school, with approx. 20 students per class). A conservative ICC of 0.38 is assumed based on the combined partitioning of variance at the between-school and between-class levels (Lamb & Fullarton, 2001), resulting in a correction factor of 8.22. The resulting student sample is 6477 students at 161 schools for a minimum detectable effect size of d = 0.20. This sample is extended across the four arms of the study for a total of 12,954 students from 644 teachers' classes at 322 schools.

2.4.2. Eligibility criteria

Schools with any number of Stage 2 classes will be eligible to participate in this research project, providing there are at least 15 Stage 2 students in the school and four teachers who are willing to participate. Teachers and students will be sourced from schools where the principal has agreed to participate in the research. Teachers who are available for the full period of the study, have full-time access to a class of students, and have not undertaken QTR previously are eligible to participate. Of the eligible teachers, two of the four must be teaching Stage 2 students. Students of consenting Stage 2 teachers will be eligible to participate.

Teachers who have previously participated in QTR will be excluded from the study. Students for whom the school receives DOE funding for significant identified learning difficulties will be excluded from analysis (note. These students are not excluded from involvement in class related activities).

As the study is designed around PLCs of four undertaking QTR, there may be some cases in which small schools are disadvantaged. In the case of small schools, if two to four schools are willing to combine to form a 'small school network' PLC of four teachers, these schools will be accepted into the study. In the case of small school networks, at least one Stage 2 "monitored" teacher is required within a constructed PLC. The exclusion criteria still apply in small school networks.

2.4.3. Recruitment

In collaboration with the NSWDOE, all public primary schools in NSW will be invited to participate in the study through a formal expression of interest process. Schools expressing interest will receive an information statement (i.e., letter to the Principal detailing the study) and consent form. A video providing an overview of the study design will summarise the complex design, (http://bit.ly/2VQzyO8) with a period of communication made available for principals, nominated school contacts, and teachers who want additional information.

Participating schools will be compensated financially to release teachers for the time spent receiving training and participating in QTR or Peer Observation. These funds will be available to schools in the QTR, QTR Trainer-led, and Peer Observation groups in 2019

after the completion of baseline data collection, and to the wait-list control and Peer Observation groups so they can participate in QTR in 2020. Schools will receive funding commensurate with the costs of substituting the participating teachers ($500 per day, per teacher), and so will compensate the time commitments of participating teachers but is not a financial inducement to participate.

2.4.3.1. Teachers. Once principals provide institutional consent for their school to participate, they will be asked to select four teachers from among volunteers, in-line with the eligibility criteria, who will be invited to participate in the study. Emphasis will be placed on communicating the voluntary nature of participation for teachers, to ensure similar levels of motivation among participants across the study (Kennedy, 2016). Selected teachers will receive a consent form and participant information statement detailing the study design, what will be required of them, benefits, data collection, and confidentiality. Teachers will also be given access to the study design video and will be able to contact the research team for further information prior to providing consent. Consenting teachers form the sample of teachers to participate in the study.

2.4.3.2. Students. After the principal and teachers have provided consent, parents of students taught by the two 'monitored' (Stage 2) teachers within each school will be informed about the research via a Parent Information Statement. This document will be sent home with the regular school correspondence to parents and will outline the study design, including details of what will be required of students, benefits, data collection/protection, and confidentiality. The Parent Information Statement will also contain a link to a student version of the study design video (http://bit.ly/2Lr1h4i) where the requirements of the study will be explained to children. Parents will also be provided a link to the full study design video, and can contact the research team for information prior to making a decision regarding consent. Students will be shown the student video by their teacher to ensure their informed participation.

Student participation is 'opt-out' in nature. Data will be collected for all students to ensure no student feels excluded, however students who have opted out will be tagged as "OUT" in all data collection systems, and their data will be removed prior to any analysis. To opt out, parents must return a student withdrawal form to the school office. Even where parents have not opted out at the time of data collection, they will be able to withdraw their child from the study at any time.

2.4.4. Cohort management

Due to the large number of schools required, and the specific inclusion criteria for this study, eligible schools will be accepted into the study upon provision of consent (rather than selecting a stratified group of schools from all those recruited). Recruitment will be closed when a sample of 200 schools is obtained. In the event of under-recruitment in the initial year of the study (2019), an additional cohort will be recruited in the subsequent year to complete the required sample (2020). In the case of an additional cohort, all study protocols outlined in this description will be adhered to for the additional cohort, one year after the initial cohort.

2.4.5. Randomisation procedures

The school is the unit of randomisation in this study. Consenting schools will be stratified based on school location (i.e., urban or rural), and socio-economic status using the continuous variable Index of Community Socio-Educational Advantage (ICSEA). These two stratification variables were selected due to their use in classifying schools nationally and availability on the Australian Government My School website (https://www.myschool.edu.au/). For school location, schools classified as Remote, Very Remote, Outer Regional and Inner Regional will be grouped into the 'rural' strata, and schools classified as Major cities grouped into the 'urban' strata. For socio-economic status, schools within each school location strata are ranked by their ICSEA and grouped into blocks of four (e.g., ICSEA 1, 2, 3, 4 = block 1, ICSEA 5, 6, 7, 8 = block 2) (Table 3).

Following stratification and blocking, schools within blocks will be allocated to one of four conditions by an external researcher using a computerised random number generator. Schools within the allocation blocks will be ranked ascendingly by their random number, with allocation of groups by rank (1 = QTR; 2 = QTR-Trainer; 3 = Peer Observation; and 4 = Control). Any blocks with less than four schools will include dummy schools to complete the block in order to maintain the same probability of the schools in these blocks being allocated to one of the four conditions. Condition allocation will be provided to each school as closely as possible following the completion of baseline data collection, and only after baseline data has been collected at an individual school. This is the preferred method of randomisation in cluster randomised trials (Murray, 1998).

Table 3
Randomisation schedule with example values.

School ID	Location strata	Ranked ICSEA	ICSEA block	Random number	Group allocation
1	Urban	1180	1		
12	Urban	1167	1		
32	Urban	1152	1		
6	Urban	1109	1		
115	Rural	980	2		
92	Rural	922	2		
3	Rural	915	2		
65	Rural	908	2		

2.4.6. Independent oversight

To ensure that evaluation processes align with CONSORT guidelines, a senior analyst from the RAND Corporation will be employed to undertake independent oversight. This role will include the following tasks:

1 Monitor randomisation processes, including:
   - Monitor all correspondence between CIs and statistician (external researcher) during the randomisation process;
   - Check that alignment of allocation is maintained between allocation and any follow-up datasets.
2 Provide ongoing advice to the project team to ensure data collection is compliant with CONSORT guidelines, including:
   - Blinding;
   - Allocation concealment (i.e., allocation reveal post baseline data collection).
3 Provide ongoing advice to the project team to ensure minimisation of bias throughout the study, including:
   - Potential split cohort design;
   - Decision making from any interim analysis.
4 Provide independent review of statistical procedures in line with statistical protocols.

2.5. Measures

2.5.1. Primary outcome: student achievement

Student achievement, the primary outcome, will be measured using the most recent editions of the Australian Council for Educational Research's (ACER, 2011) Progressive Achievement Tests (PATs) for Mathematics (PAT-M version 4), Reading (PAT-R version 5) and Science (PAT Science). These measures have been rigorously tested and are widely used in Australian schools and internationally to assess and monitor students' skills, understandings and growth over time. Item reliability produced using Rasch modelling is reported at 0.87 – 0.91 (Lindsey, Stephanou, Urbach, & Sadler, 2010), with this measure interpreted in the same manner as Cronbach's alpha within classical test theory (Bond & Fox, 2015; pp-70). Further, Fogarty (2007) reported that these tests had good predictive validity when administered to 805 Australian secondary students. Each of these tests has three developmentally appropriate levels for each school year that will be tested (in this case, Year 3 and Year 4). Scaled scores will allow for comparison of students across test levels and testing years.

In order to mitigate as far as possible against ceiling and floor effects with testing, we will be using the second of the three tests that are identified as being at developmentally appropriate levels for each year level for baseline data collection (Table 4), and the next progression will be used during follow-up. Tests will be administered electronically with students given 40 min for each of the tests, overseen by members of the research team, blinded to group allocation. Aside from support with technological issues, no assistance will be given to the students during the tests. Students who ask for assistance will be advised to skip the question and come back to it if they have time. To guard against contamination – to hold as much as possible constant – testing will be conducted according to the following strict timetable whereby all maths and reading tests are completed in the morning and all science tests and questionnaires (see below) will be completed in the afternoon. Testing will be undertaken over a two-day period, with mathematics and reading tests undertaken on separate days. The analysis will utilise the test scaled score (interval scale from 0 to 100).

2.5.2. Secondary outcomes: students

A range of secondary outcomes will also be tested to gauge whether improvements in the quality of teaching as a result of participation in QTR not only affect academic outcomes but a broader range of important social and emotional outcomes of schooling (Ladwig, 2010). Students will complete a questionnaire to assess any changes in how they perceive their school life and educational and occupational aspirations.

Quality of school life (QSL) will be assessed using ACER's Quality of School Life scales modified by Ainley and Bourke (1992), based on the original scales developed by Williams and Batten (1981). The QSL scale has indicated that it has a stable latent structure, and the subscales have reasonable reliabilities (Mok & Mcdonald, 1994). Scale means are calculated for analysis, with scales analysed individually. These scales use a four-point measure (Agree, Mostly agree, Mostly disagree, Disagree) to gauge student perceptions of:

- General satisfaction (6 items): enjoyment in the school environment ($\alpha = 0.83$).

Table 4
Student testing schedule.

Measure	Timing	Year level	Assessment time point	
			Baseline	Follow-up
Mathematics	Morning	Year 3	Level 2	Level 3
		Year 4	Level 3	Level 4
Reading	Morning	Year 3	Level 3	Level 4
		Year 4	Level 4	Level 5
Science	Afternoon	Year 3	Level 1	Level 2
		Year 4	Level 2	Level 3
Questionnaire	Afternoon	All years	N/A	N/A

- Achievement (5 items): success as a student and ability to undertake class work (α = 0.83).
- Teachers (5 items): teacher fairness and willingness to help (α = 0.86).
- Relevance (6 items): schooling as useful for future endeavours (α = 0.85).
- Adventure (5 items): excitement and interest in class work (α = 0.84).

Student aspirations will be assessed using items from the Aspirations Longitudinal Study Student Survey (Gore et al., 2015) and linked with the family background variables of parental occupation and level of education. These items are designed to record:

- Educational aspirations: highest level of education a student plans to complete.
- Occupational aspirations: in response to the question "what kind of work would you like to do when you grow up?"
- Importance of school work (5 items): level of support for and perception of importance of study. The internal consistency of this scale is unknown, and will be established using Cronbach's alpha prior to analysis.

2.5.2.1. Additional student data. Co-variates to be considered in the statistical models are: (1) demographic data obtained from the NSWDOE: age, year level, gender, language background, and Aboriginal and/or Torres Strait Islander status; and, (2) student self-reports on if their teacher provides homework, an estimate of how many books they read during the year, and if they received tutoring in mathematics or reading across the year (and the volume of instruction).

To check the validity of any effects of QTR on the primary outcome, national standardised testing data will also be obtained via the NSWDOE in the form of de-identified National Assessment Program for Literacy and Numeracy (NAPLAN) scores and achievement bands. These data will be the most recent available for each participating student. These data are not used as the primary outcome for students as its two-year collection cycle (Years 3, 5, 7 and 9) falls outside the time students are exposed to the instruction of the "monitored" teacher in this study.

2.5.3. Secondary outcomes: teachers

2.5.3.1. Quality of teaching. Quality of teaching will be based on observation and scoring of monitored teachers' lessons using the Quality Teaching Classroom Practice Guide (NSW Department of Education & Training (NSWDET), 2003) by members of the research team, blinded to allocation. This measure has demonstrated high validity in previous studies (Bowe & Gore, 2017; Ladwig, Smith, Gore, Amosa, & Griffiths, 2007). Two lessons will be observed at each assessment point (baseline and immediate post-test). The coding process will produce a score for each element. The mean score for the 18 elements will be used as the overall score of teaching quality. Using data from a previous study (Bowe & Gore, 2017), the estimate of internal consistency (Cronbach's alpha) for this measure was α = 0.82.

Members of the data collection team will participate in training in the weeks prior to initial assessments, preparing them to make reliable assessments of teaching quality. Training will involve explanation of the QT model elements, the process of coding lessons, and opportunities to practice coding of video recorded lessons. Joint observation coding will be conducted for 20% of observations to assess inter-rater reliability. Joint observations will require a coding discussion between observers to reach an agreed code, serving as ongoing development for assessors. Assessors new to classroom observation assessment will undertake a training period of three joint observations prior to solo assessment of classroom practice.

Table 5
Teacher questionnaire measures and timing of administration.

Construct	Measure	Items	Time				
	(Sub-scale; α)		Base	2-mth	5-mth	7-mth	12-mth
Teacher engagement (Klassen, Yerdelen, & Durksen (2013))	Engaged teachers scale (Cognitive engage; 0.84) (Emotional engage; 0.87) (Social engage - colleague; 0.79) (Social engage - student; 0.83)	16	x			X	x
Teaching efficacy (Tschannen-Moran & Woolfolk Hoy (2001))	Teachers' sense of efficacy scale (Instruction; 0.91) (Management; 0.90) (Engagement; 0.87)	12	x	x	x	x	x
Collective morale (Hart, Wearing, Conn, Carter, & Dingle (2000))	Teacher collective morale (Morale; 0.86) (Appraisal and recognition; 0.90)	11	x	x	x	x	x
Collective efficacy (Goddard, Hoy, & Hoy (2004))	Collective efficacy scale (Collective efficacy; 0.92) (Powerlessness; 0.83) (Trust in teachers; 0.95) (Teacher efficacy; 0.87)	21	x			x	x
Background	Years of experience (teaching and current school), the year they currently teach, qualifications, QTR experience	4	x				
Professional development	Professional development activities for the previous year: type, time, impact and support (OECD, 2014)	6				x	

2.5.3.2. Teacher questionnaire. Teachers will complete a questionnaire to assess any changes in engagement, teaching efficacy, collective morale, and collective efficacy. Mean scale scores will be used for all analysis. Table 5 details the measures used and their timing across the study. Teacher efficacy and morale will be measured at additional time-points across the study as these measures are potentially susceptible to seasonal influence. To obtain a measure of additional PD undertaken and /or clear picture of PDAU, the follow up questionnaire will include validated measures of PD undertaken during the year of the study (OECD, 2014).

2.5.3.3. Teacher interviews. Researcher-conducted interviews will take place with one monitored teacher in 10% of schools, randomly selected from all four arms. These semi-structured, phone interviews will typically last for around 45 min and will take place at two data collection points (pre-intervention and post-intervention). Pre-intervention interviews will focus on issues of teaching experience, career plans and aspirations, school culture, professional development experiences and any prior experiences with the Quality Teaching model and QTR. Post-intervention interviews with the same sample of teachers will focus in more detail on their experiences of QTR or peer observation and any other professional development over the course of the year. All interview participants will be asked to share their experiences of participating in this research and invited to contribute views that fall outside of the predetermined list of topics covered by the interview questions.

2.5.4. Intervention fidelity checks

All PLCs involved in intervention and peer observation arms during the trial will be asked to provide regular data on implementation fidelity to evaluate if the interventions are being undertaken as designed, and to evaluate any adaptation within and across settings. Fidelity processes are:

1 Intervention protocols will be outlined during workshops for (at least) two teachers from each school who will facilitate the intervention within their schools.
2 An online checklist will be completed by a member of each PLC on each of the four days during which the in-school PD is undertaken (QTR or peer observation) addressing the degree to which they have complied with the main components of the intervention. This online survey includes key items relating to the conduct of a PLC event (e.g., QTR: Did PLC members individually code all QT elements prior to the lesson discussion? How long was the post-lesson discussion?) A fidelity score will be calculated for each PLC (i.e., level of compliance).
3 Fidelity checks (one per school) will be conducted by a member of the research team (with the researcher staying for the whole event on each occasion) to monitor how interventions are being conducted against the relevant protocols, and fidelity checklist for the intervention being undertaken.

2.6. Data analysis

The study has been designed to provide both quantitative and qualitative data. Statistical analyses of the primary and secondary outcomes will be conducted with linear mixed models using IBM PASW Statistics 25 (SPSS Inc. Chicago, IL) software. Impacts are estimated using an intention-to-treat (ITT) approach, with alpha levels set at $p < 0.05$. Linear mixed models will be fitted to compare continuous outcomes for each of the intervention groups (Researcher-led QTR, Trainer-led QTR and Peer observation) against those of the PDAU control condition. Group, time, and group-by-time interaction will be assessed as fixed effects within the model, with covariates of gender and year level (Year 3 or Year 4) also included as fixed effects. The class a student belongs to will be included as a random intercept within the model to account for clustering of students within classes, and subject (student ID) will be included as a random intercept to model repeated measures at the individual level. Differences of means and 95% confidence intervals (CIs) will be determined using the linear mixed models. Moderators of intervention effects will also be explored using linear mixed models with interaction terms for the following: (a) SES (based on school ICSEA values), (b) geographic location of school (urban versus rural), (c) years of teaching experience, and (d) sex of teacher.

2.6.1. Effect size calculation

Hedges' g will be used to determine effect sizes of the change in mean score for each group relative to the baseline value (effect of intervention on the mean change score), where $\bar{x}_1$ is the conditional estimate of the control group, and $\bar{x}_2$ is the conditional estimate of the intervention group being compared (QTR or peer observation).

$$g = \frac{\bar{x}_2 - \bar{x}_1}{s^*}$$

The conditional estimate of $\bar{x}_2 - \bar{x}_1$ is recovered from β of the Group x Time interaction in the primary ITT analysis model; s^* is estimated from the analysis sample as follows:

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

wheren $_1$ is the sample size in the control group, n_2 is the sample size in the treatment group, s_1 is the standard deviation of the control group, and s_2 is the standard deviation of the treatment group (all estimates of standard deviation used are unconditional).

Ninety-five per cent confidence intervals (95% CIs) of the effect size will be computed using the compute.es function (AC Del Re, 2013) in R version 3.4.4 (R Core Team, 2019). This function computes the confidence intervals using the variance in g derived by the

Hedges & Olkin (Hedges & Olkin, 1985, p. 86) formula:

$$\text{var}(g) = \frac{n_1 + n_2}{n_1 n_2} + \frac{g^2}{2(n_1 + n_2)}$$

2.6.2. Non-compliance and protocol adherence analysis

In a randomised controlled trial the intervention effect in an intention-to-treat (ITT) analysis may be biased when there is 'contamination' by control group participants who received the intervention protocol and/ or 'differential compliance' when participants in the intervention condition receive less than their prescribed allocation (e.g., for reasons such as non-compliance, partial compliance, or issues with adherence to protocol). An Instrumental Variables (IV) approach (Angrist & Imbens, 1995) may be used to estimate the unbiased treatment effect by using the received treatment as an instrument for group allocation.

A Two Stage Least Square (2SLS) approach will be undertaken using the SYSLIN procedure in SAS version 9.4 (SAS Institute, Cary NC). The first stage of the 2SLS involves regression of group allocation (intervention = 1; control = 0) on the compliance instrument (teacher achieved a full set of intervention events = 1; teacher did not achieve a full set of events = 0), with covariates baseline PAT scaled score, gender and year level. The second stage involves regression of the dependent variable (Follow-up PAT scaled score) on the predicted values obtained from the first stage, with covariates baseline PAT scaled score, gender and year level.

The correlation between group allocation and compliance variables will be reported alongside the model F test from the first stage. The suggested rule of thumb for checking the strength of the instrument is that the F-statistic from the first stage regression should be greater than 10, or the t value of the instrument above approximately 3 (Angrist, 2006). Otherwise, the instrument is considered to be weak, the consequence of which is that the sampling distribution of the 2SLS estimator might not be approximately normal even in large samples and is very likely to be biased.

The parameter estimates and corresponding effect size from the second stage will be reported. It is also important to note that the standard errors produced in the second stage regression are not correct; the standard error for the 2SLS estimates must take into account the additional uncertainty due to performing two stages of regression. The SYSLIN procedure in SAS contains an algorithm to adjust the standard errors in the second stage.

In addition to an IV approach, intervention participants may choose to modify the intervention protocols, producing lower levels of fidelity to the intended protocols. Per-protocol analyses will be performed for those PLCs that met at least 75% of the pre-specified fidelity standards (based on fidelity reporting). The linear models used in the intention-to-treat analysis will be fitted during the per-protocol procedure.

2.6.3. Qualitative analysis

Qualitative data from interviews with teachers will be transcribed. Recurrent themes within and beyond the a priori areas of focus from the interviews will be categorised and coded using NVivo (QSR International, 2014). Specific attention will be paid to such issues as: (1) aspects of the intervention that were most valued by the participants; (2) what, how, and why the intervention made an impact on which teachers and students; (3) how differences between intervention groups interacted with teaching culture, teacher identity, and teachers' career plans and aspirations; and (4) how differences between intervention groups relate to issues of implementation and scalability. For participants in the PDAU control group, questions are modified to relate to professional development that they have undertaken during the intervention period.

3. Discussion

The project design presented in this paper has a number of strengths. In particular, the intervention under examination, QTR, is founded on a strong evidence base of research into effective pedagogy, which generated the QT model (NSW Department of Education & Training (NSWDET), 2003), and effective PD, which led to the development and early testing of QTR (Bowe & Gore, 2017, 2017). Furthermore, the mixed method design brings a level of granularity to the analysis, addressing a common critique of RCTs in education whilst also offering a level of rigour that addresses a common limitation of PD studies that rely solely on quasi-experimental designs or the self-reports of teachers and students.

A potential limitation of the design is that we are only testing the efficacy of QTR relative to one alternative form of PD. On the other hand, the inclusion of an active control group goes beyond the typical intervention/control group design of most RCTs in education, which are often constrained by the high costs of this form of research. Another potential limitation is that the development and delivery of the active control PD is to be conducted by members of our research team. However, this alternative PD will strictly adhere to the Peer Observation information provided by AITSL and the associated workshop will be delivered by members of the research team not involved in the delivery of the QTR workshops. Finally, (Gore & Bowe, 2015; Gore, Lloyd, Smith, Bowe, & Ellis, 2017; Gore, 2018) central involvement in the development of the QTR intervention and hence lack of independence from the trial may be seen as a limitation. However, we have attempted to mitigate this concern by: having researchers not involved in the development of QT or QTR leading key aspects of the study; establishing a rigorous form of independent oversight; and, strictly following CONSORT protocols.

4. Conclusion

This study extends previous research into QTR, with a protocol rigorously designed to assess the potential of QTR for system-wide impact on student academic achievement. While there exists a large body of evidence on the impact of different forms of PD on teacher change, direct links between teacher PD and improvement of student academic outcomes have proved elusive. This challenging relationship between PD and student outcomes is the primary focus of this study. The design also enables assessment of important secondary outcomes that may contribute to students' educational achievement including students' academic self-perception, perceived relevance of schooling, and relationships with teachers. In addition, the study has been designed to examine the relationship between participation in QTR and changes to teaching practice, teacher morale, engagement and efficacy.

If the study demonstrates a measurable effect of QTR on student outcomes relative to other forms of PD, it would be a major breakthrough for research on teacher professional learning – a field in which effects have too often been asserted rather than demonstrated. Moreover, the comprehensive design of the study should illuminate not only whether the intervention 'works' but also provide important insights into the questions of 'how', 'why', 'for whom', and 'under what conditions'.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgements

This research is funded by the Paul Ramsay Foundation and the Australian Research Council (DP180100285). Each funding body has reviewed the methodology of this trial.

References

AC Del Re (2013). *compute.es: Compute effect sizes. R package version 0.2-2.* http://cran.r-project.org/web/packages/compute.es.

ACER (2011). *Interpreting ACER test results, Melbourne: Australian council for educational research.* Available at:https://www.acer.org/tsa/progressive-achievement-tests-pat.

Ainley, J., & Bourke, S. (1992). Student views of primary schooling. *Research Papers in Education, 7*(2), 107–128. https://doi.org/10.1080/0267152920070202.

Angrist, J. (2006). Instrumental variables methods in experimental criminological research: What, why and how. *Journal of Experimental Criminology, 2*(1), 23–44. https://doi.org/10.1007/s11292-005-5126-x.

Angrist, J., & Imbens, G. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *American Statistical Association, 90*(430), 431–442. https://doi.org/10.1080/01621459.1995.10476535.

Aubusson, P., Steele, F., Dinham, S., & Brady, L. (2007). Action learning in teacher learning community formation: Informative or transformative? *Teacher Development, 11*(2), 133–148. https://doi.org/10.1080/13664530701414746.

Bond, T., & Fox, C. M. (2015). *Applying the rasch model : Fundamental measurement in the human sciences* (third edition). London, UNITED KINGDOM: Taylor & Francis Group.

Bowe, J. M., & Gore, J. M. (2017). Reassembling teacher professional development: The case for quality teaching rounds. *Teachers and Teaching Theory and Practice, 23*(3), 352–366. https://doi.org/10.1080/13540602.2016.1206522.

Carter, A. (2015). *Carter review of initial teacher training (ITT). Report prepared for the UK Secretary of State for Education*Retrieved fromhttps://gov.uk/government/publications/carter-review-of-initial-teacher-training.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research, 60*(3), 276–291. https://doi.org/10.1080/00131881.2018.1493353.

Cordingley, P., Bell, M., Evans, D., & Firth, A. (2005). *The impact of collaborative CPD on classroom teaching and learning. Review: What do teacher impact data tell us about collaborative CPD?* Retrieved fromhttp://eppi.ioe.ac.uk/cms/Default.aspx?tabid=395&language=en-US.

Council for the Accreditation of Educator Preparation (CAEP) (2015). *CAEP evidence guide version 2.0.* Retrieved from http://caepnet.org/~/media/Files/caep/knowledge-center/caep-evidence-guide.pdf?la=en.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York, NY: Teachers College Press.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research.* London: Arnold Publishers.

Elmore, R. F. (2007). Professional networks and school improvement. *School Administrator, 64*(4), 20–25. Retrieved from https://eric.ed.gov/?id=EJ759044.

Ewing, R., Groundwater-Smith, S., Mockler, N., Loughland, T., Simpson, A., Smith, D., & Brooks, D. (2010). *Meta analysis of quality teaching action learning project*Retrieved from. Sydney, Australia: University of Sydney2003–2009. http://sydney.edu.au/education_social_work/professional_learning/resources/papers/AGQTP_reports/2010_MetaAnalysis_Quality_Teaching_Action_Learning_Project.pdf.

Fogarty, G. (2007). *Research on the Progressive Achievement Tests and academic achievement in secondary schools*(Technical report). Retrieved from Melbourne:https://www.acer.org/files/ACERPress_PAT_Supp.pdf.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3–13. https://doi.org/10.3102/0013189x033003003.

Goodwin, A. L., Del Prete, T., Reagan, E. M., & Roegman, R. (2015). A closer look at the practice and impact of "rounds". *International Journal of Educational Research, 73*, 37–43. https://doi.org/10.1016/j.ijer.2015.06.006.

Gore, J. M. (2007). Improving pedagogy: Challenges of moving teachers toward higher levels of quality teaching. In J. Butcher, & J. Asha (Eds.). *Making a difference: Challenges for teachers, teaching, and teacher education, (15–33).*. New York, NY: Sense.

Gore, J. (2018). Making a difference through Quality Teaching Rounds: Evidence from a sustained program of research. *Paper presented at the annual Research Conference of the Australian Council for Educational Research Research Conference*.

Gore, J. M., & Bowe, J. M. (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., Lloyd, A., Smith, M., Bowe, J., & Ellis, H. (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. M., Griffiths, T., & Ladwig, J. G. (2004). Towards better teaching: Productive pedagogy as a framework for teacher education. *Teaching and Teacher Education, 20*(4), 375–387. https://doi.org/10.1016/j.tate.2004.02.010.

Gore, J., Holmes, K., Smith, M., Southgate, E., & Albright, J. (2015). Socioeconomic status and the career aspirations of Australian school students: Testing enduring assumptions. *Australian Educational Researcher, 42*(2), 155–177. https://doi.org/10.1007/s13384-015-0172-5.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500. https://doi.org/10.1177/003172170909000709.
Hammond, J. (2008). Intellectual challenge and ESL students: Implications of quality teaching initiatives. *Australian Journal of Language and Literacy, 31*(2), 128–154. Retrieved from https://search.informit.com.au/fullText;dn=180621744400022;res=IELHSS.
Hart, P. M., Wearing, A. J., Conn, M., Carter, N. L., & Dingle, R. K. (2000). Development of the School Organisational Health questionnaire: A measure for assessing teacher morale and school organisational climate. *The British Journal of Educational Psychology, 70*(2), 211–228. https://doi.org/10.1348/000709900158065.
Hattie, J. (2008). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. *Hoboken, NJ: Taylor & Francis.*.
Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.
Higgins, S., Elliot-Major, L., Coleman, R., Katsipataki, M., Henderson, P., Mason, D., & Kay, J. (2018). *Sutton Trust-EEF Teaching and Learning Toolkit & EEF early years toolkit: Technical appendix and process manual (working document v.01)*. Retrieved:London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Toolkit/Toolkit_Manual_2018.pdf.
Indig, D., Lee, K., Grunseit, A., Milat, A., & Bauman, A. (2018). Pathways for scaling up public health interventions. *BMC Public Health, 18*(1), 68. https://doi.org/10.1186/s12889-017-4572-5.
Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980. https://doi.org/10.3102/0034654315626800.
Klassen, R. M., Yerdelen, S., & Durksen, T. L. (2013). Measuring teacher engagement: Development of the engaged teachers scale (ETS). *Frontline Learning Research, 1*(2), 33–52. Retrieved from https://files.eric.ed.gov/fulltext/EJ1090832.pdf.
Ladwig, J. G. (2007). Modelling pedagogy in Australian school reform. *Pedagogies, 2*(2), 57–76. https://doi.org/10.1080/15544800701343919.
Ladwig, J. G. (2010). Beyond academic outcomes. *Review of Research in Education, 34*(1), 113–141. https://doi.org/10.3102/0091732X09353062.
Ladwig, J. G., Smith, M., Gore, J., Amosa, W., & Griffiths, T. (2007). Quality of pedagogy and student achievement: Multi-level replication of authentic pedagogy. *Paper Presented at the Annual Conference of the Australian Association for Research in Education*.
Lamb, S., & Fullarton, S. (2001). *Classroom and school factors affecting mathematics achievement: A comparative study of the US and Australia using TIMSS.* In.http://research.acer.edu.au/timss_monographs/10.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.
Lindsey, J., Stephanou, A., Urbach, D., & Sadler, A. (2010). *Progressive acheivement tests in mathematics* (third edition). Melbourne, Victoria: Australian Council for Educational Research (ACER).
Lingard, B., Ladwig, J., Mills, M., Bahr, M., Chant, D., Warry, M., & Luke, A. (2001). *The queensland school reform longitudinal study.* Brisbane: Education Queensland.
Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher, 48*(3), 158–166. https://doi.org/10.3102/0013189X19832850.
Louden, W. (2008). 101 damnations: The persistence of criticism and the absence of evidence about teacher education in Australia. *Teachers and Teaching Theory and Practice, 14*(4), 357–368. https://doi.org/10.1080/13540600802037777.
Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*. https://doi.org/10.1136/bmj.c869.
Mok, M. M. C., & Mcdonald, R. P. (1994). Quality of school life: A scale to measure student experience or school climate? *Educational and Psychological Measurement, 54*(2), 483–495. https://doi.org/10.1177/0013164494054002023.
Murray, D. M. (1998). *Design and analysis of group-randomized trials.* New York, NY: Oxford University Press.
Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). *Authentic achievement: Restructuring schools for intellectual quality.* San Francisco, CA: Jossey Bass.
NSW Department of Education and Training (NSWDET) (2003). *Quality teaching in NSW public schools: A classroom practice guide.* Sydney, Australia: NSW Department of Education and Training/Professional Support and Curriculum Directorate.
OECD (2005). *Teachers matter: Attracting, developing and retaining effective teachers.* Paris, France: Organisation for Economic Co-operation and Development.
OECD (2014). *Talis 2013 technical report*Retrieved from:. Paris: OECD Publishing. http://oecd.org/education/school/TALIS-technical-report-2013.pdf.
Penney, D., Brooker, R., Hay, P., & Gillespie, L. (2009). Curriculum, pedagogy and assessment: Three message systems of schooling and dimensions of quality physical education. *Sport, Education and Society, 14*(4), 421–442. https://doi.org/10.1080/13573320903217125.
Plummer, F., Nyholm, M., Quince, S., & Dione, M. (2010). Innovative practice in middle years literacy: A New South Wales perspective on professional learning. *Literacy Learning: The Middle Years, 18*(2), 31. Retrieved from https://search.informit.com.au/fullText;dn=131121200357974;res=IELHSS.
QSR International (2014). *NVivo (version 10.0.638.0 SP6). Doncaster.* Australia: QSR International.
R Core Team (2019). *R: A language and environment for statistical computing.* From R Foundation for Statistical Computinghttps://www.R-project.org/.
Roberts-Hull, K., Jensen, B., & Cooper, S. (2015). *A new approach: Teacher education reform.* Melbourne, Australia: Learning First.
Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247–252. https://doi.org/10.1257/0002828041302244.
Rowe, K. (2003). The importance of teacher quality as a key determinant of students' experiences and outcomes of schooling. *Paper Presented at the ACER Research Conference*. Retrieved from https://research.acer.edu.au/research_conference_2003/.
Rushton, K. (2008). Cooperative planning and teaching for ESL students in the mainstream classroom. *TESOL in Context, 18*(1), 21. Retrieved from https://search.informit.com.au/fullText;dn=586801884857523;res=IELHSS.
Teacher Education Ministerial Advisory Group [TEMAG] (2014). *Action now: Classroom ready teachers. Report prepared for the Australian Minister for Education*Retrieved fromhttps://docs.education.gov.au/system/files/doc/other/action_now_classroom_ready_teachers_accessible.pdf.
Treble, V. (2009). 'A love of teaching and learning': A case study of English teaching transformed through quality teaching and professional learning. *Teacher Development, 13*(4), 363–371. https://doi.org/10.1080/13664530903578280.
Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1.
Whalan, F. (2012). *Collective responsibility: Redefining what falls between the cracks for school reform.* Rotterdam, Netherlands: Sense Publishers.
Williams, T., & Batten, M. (1981). *The quality of school life. Hawthorn.* Victoria: Australian Council for Educational Research.