Research paper

Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds

Jennifer Gore [a, *, 1], Adam Lloyd [a, 1], Maxwell Smith [a, 1], Julie Bowe [a, 1], Hywel Ellis [a, 1], David Lubans [a, b, 2]

[a] Teachers and Teaching Research Centre, School of Education, University of Newcastle, Australia
[b] Priority Research Centre for Physical Activity and Nutrition, University of Newcastle, Australia

HIGHLIGHTS

- This study tested a pedagogy-based PD approach for impact on quality of teaching.
- The cluster randomised controlled trial involved 192 teachers in 24 schools.
- Significant positive effects ($d = 0.4$) were found for teaching quality.
- Positive impacts on teacher morale and sense of recognition were also observed.
- Effects were sustained 6 months post-intervention.

ARTICLE INFO

ABSTRACT

Robust evidence of the effectiveness of professional development for teachers is limited. This study tested a pedagogy-based, collaborative PD approach for impact on the quality of teaching. A cluster randomised controlled trial involving eight teachers at each of 24 schools found significant positive effects on teaching quality ($d = 0.4$), independent of school type (primary/secondary), school location (urban/rural), and years of teaching experience. These effects were sustained six months later. Qualitative data are used to illustrate mechanisms underpinning the success of the intervention. This study illuminates how to support teacher learning for measurable positive impacts on teaching quality and teacher morale.

1. Introduction

School systems throughout the world acknowledge that the quality of teaching is the most critical in-school factor impacting on student outcomes (Darling-Hammond, 2000; Hattie, 2008; Organisation for Economic Co-operation and Development [OECD], 2005; Rockoff, 2004; Rowe, 2003; Timperley & Alton-Lee, 2008).

Every year, in the name of improving teaching quality, millions of dollars are invested in teacher professional development (PD) and elaborate regulatory systems have been designed to ensure that teachers engage in ongoing professional learning activities. Yet few studies show rigorous evidence of the impact of such activities (Avalos, 2011; Cordingley, Bell, Evans, & Firth, 2005; Guskey & Yoon, 2009; Vescio, Ross, & Adams, 2008). Moreover, PD programs often lack clear and direct links with classroom practice. The result is what some have referred to as an "evidence void" when it comes to documenting the sustained effects of PD on either teaching practice or student outcomes (Council for the Accreditation of Educator Preparation [CAEP], 2015; Cuban, 2013; Darling-Hammond, 2013; Hill, Beisiegel, & Jacob, 2013; Kennedy, 2016).

In the absence of strong evidence for PD, concerns about the quality of teaching in schools are increasingly being addressed in two alternative ways, each of which has gained significant political

* Corresponding author.
   E-mail addresses: jenny.gore@newcastle.edu.au (J. Gore), adam.lloyd@newcastle.edu.au (A. Lloyd), maxwell.smith@newcastle.edu.au (M. Smith), julie.bowe@newcastle.edu.au (J. Bowe), hywel.ellis@newcastle.edu.au (H. Ellis), david.lubans@newcastle.edu.au (D. Lubans).
[1] University of Newcastle, Hunter Building HA80, University Drive, Callaghan NSW 2308 Australia.
[2] University of Newcastle, Advanced Technology Centre ATC301, University Drive, Callaghan NSW 2308 Australia.

traction. One approach seeks to improve quality by restricting entry into teaching to only the "best and the brightest" (Teacher Education Ministerial Advisory Group [TEMAG], 2014; Weldon, McKenzie, Kleinhenz, & Reid, 2013; Wilson, Dalton, & Baumann, 2015). The logic here is that the quality of teaching is a function of the quality of the teacher—typically understood as academic credentials, some-times with dispositions and personality traits also considered (Bowles, Hattie, Dinham, Scull, & Clinton, 2014). As an approach to "fixing" the quality of teaching however, the proposed tightening of selection criteria fails to provide a short-term solution and ignores the impacts of preservice teacher development. It also lacks feasi-bility if dramatic economic changes are not simultaneously imple-mented (Goldhaber, 2015; see also Gore, Barron, Holmes, & Smith, 2016). A second approach centres on finding sophisticated ways to evaluate the quality of teaching (Grissom & Youngs, 2015). This line of work seeks to improve teaching by using robust measures of quality to weed out the worst teachers and learn from the best (Hanushek & Rivkin, 2010; Metzler, 2014). However, even leaving aside complex and unresolved measurement issues, *evaluating* teaching quality will have limited impact on *improving* teaching quality unless linked to an effective approach to PD.

Given the limitations of these approaches, the question of how teaching can be improved remains one of fundamental significance internationally. In Australia alone, the teaching workforce amounts to 250,000 teachers (Weldon, 2015) and there are more than 76,000 prospective teachers enrolled in undergraduate teacher education programs.[3] Finding ways to support and develop teach-ers (i.e., inservice) and student teachers (i.e., preservice) thus re-mains a strategy worth pursuing with some urgency—both for moral reasons, in support of these teachers and their many stu-dents, and for pragmatic reasons, in terms of the exorbitant re-sources required to re-build a teaching workforce from the point of recruitment. This paper reports on one attempt to impact on the development of practising teachers, based on a particular peda-gogical framework, Quality Teaching (QT), and a particular approach to using the framework, Quality Teaching Rounds (QTR).

1.1. Background to the study

Despite the substantial corpus of research on PD, few studies directly link specific teacher development activities to changes in teaching practice and/or improved student outcomes (Desimone, 2009; Guskey & Yoon, 2009; Guskey, 2002; Kennedy, 2014, 2016). Where studies have found positive effects, the PD has typically been limited to a small part of teaching practice (Hill et al., 2013), a small group of teachers (Korthagen, 2016), or single subject area (Penuel, Fishman, Cheng, & Sabelli, 2011). The reported effects have been on outcomes such as teachers' satisfaction (Ullah & Jundran, 2014), attitude change and commitment to innovation (Desimone, 2009), and self-efficacy (Tzivinikou, 2015), or effects on student achievement gains, often in mathematics or science (Blank & de las Alas, 2009). Nonetheless, some scholars speak of an emerging "consensus" that effective approaches to PD involve teachers as both learners and teachers (Darling-Hammond & McLaughlin, 1995), are needs-supportive (Aelterman et al., 2013), take place within the school day (Garet, Porter, Desimone, Birman, & Yoon, 2001), are integrated into practice (Armour & Yelling, 2007), cohere with school and system policies (Desimone, 2009; Ingvarson, Meiers, & Beavis, 2005), and promote transformative practice, rather than accountability (Kennedy, 2005). Most

attempts to implement PD that meets these criteria have been expensive and yielded weak return on investment (Harris & Sass, 2011; Hill et al., 2013). Leading researchers conclude that in order to deliver the highest quality PD, investment needs to be limited to fewer teachers or additional resources must be found (Garet et al., 2001).

The approach reported in this paper not only builds on these so-called "principles of effective PD," but can be applied across all subject areas and all levels of schooling (K—12) at a relatively low cost. It aims to support teachers in improving their practice while also developing their efficacy, well-being, and professional engagement. This is in contrast to approaches that subject teachers to greater levels of accountability, evaluation, and performance review. Simultaneously, and ambitiously, our approach aims to provide evidence of a kind that is persuasive to governments and education systems that want to be sure their investments have pay-off, especially in terms of student outcomes.

The approach, Quality Teaching Rounds, is designed to bring together the benefits of teachers working in professional learning communities (PLCs) (Lave & Wenger, 1991) and engaging in a form of instructional "rounds" (Elmore, 2007), with the Quality Teaching pedagogical framework (NSW Department of Education and Training [NSW DET], 2003). By adding a pedagogical framework to collaborative PD activities, Quality Teaching Rounds provides teachers with a common language and set of conceptual standards with which to engage in rigorous diagnostic professional conver-sations with colleagues (Bowe & Gore, 2017; Gore & Bowe, 2015). The framework and the approach to using the framework are grounded in an analysis of prior research (Ladwig & King, 2003; Newmann, 1996) and more than a decade of our own research into how QT functions in supporting teachers and improving teaching practice (Gore, 2014).

Quality Teaching Rounds, the PD approach, involves four (or more) teachers working in a PLC. A "Round" is comprised of three sequential sessions that occur on a single day:

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 1 h)
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 min); and
3 *Coding and discussion*: Individual coding of the observed lesson, including coding by the observed teacher, is followed by dis-cussion whereby all PLC members contribute (typically one to 2 h). Coding and discussion are centred on constructs of the Quality Teaching framework (described below)

In Quality Teaching Rounds, at least one lesson is observed for every member of the PLC, and PLC members stay together for an entire set of Rounds. The intent of Quality Teaching Rounds 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).

Fundamental to Quality Teaching Rounds is the structuring of observations and post-lesson discussions through the research-based constructs of the Quality Teaching framework (NSW DET, 2003). This pedagogical framework has been widely used during the past decade in Australia (particularly in New South Wales and the Australian Capital Territory). Derived from work on Authentic Pedagogy (Newmann, 1991; Newmann, Marks, & Gamoran, 1996) and an extensive synthesis of research on aspects of pedagogical practice that make a difference for student outcomes (Ladwig & King, 2003), the QT framework focuses teachers' attention on three dimensions of pedagogy: (i) Intellectual Quality, (ii) Quality

---

[3] From data that is publically available via uCube (Australian Government Department of Education and Training, 2016) under the Creative Commons Attri-bution 3.0 Australia Licence.

Table 1
Dimensions and elements of the quality teaching framework.

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

Learning Environment, and (iii) Significance. These dimensions are comprised of six elements each (see Table 1), with each element framed as an inquiry question to focus teachers' observations and discussions rather than as a rigid directive about what every lesson or every classroom must look like. The concepts, associated language, and fine-grained indicators of quality (using a 1—5 scale) help teachers describe and analyse classroom-based evidence, thus forming the framework for discussion within the Quality Teaching Rounds process.

While there is growing advocacy for pedagogical frameworks to guide the improvement of teaching, the QT framework differs in several respects from other widely used frameworks like CLASS (Pianta, LaParo, & Hamre, 2008) and Framework for Teaching (Danielson, 2007). First, the QT framework offers a comprehensive account of teaching, addressing matters of curriculum, student engagement, and social justice, as well as pedagogical practice (Gore, 2007). In this way, it avoids reducing the complex, multi-dimensional enterprise of teaching (Jackson, 1968) to a set of teaching skills or practices. On the contrary, the QT framework is more about a conception of "the practice of teaching" (Lampert, 2010, p. 29) (akin to the practice of law or practice of medicine) than it is a specific set of skills to be mastered. Arguably, such an approach has potentially profound consequences for teachers' intellectual and attitudinal learning as well as their behavioural learning (Evans, 2014), the latter which tends to be the focus when frameworks are designed and/or used primarily for evaluative purposes. In operationalising multiple theoretical perspectives, the QT framework provides structure and direction for the many small decisions that teachers make daily. Importantly, with its three key principles of Intellectual Quality, Quality Learning Environment, and Significance, the QT framework avoids the proliferation of discrete practices in a way that becomes unwieldy (like the 1001 teaching activities identified in a 1929 study by Charters and Waples, as cited in Zeichner, 2012).

Second, the QT framework provides a level of specificity that offers a clear goal toward which teachers can strive. Teachers engage with the definition for each element and then use the focus question and coding scale, as depicted in Fig. 1, to determine the most appropriate code for the lesson observed. After making independent judgments, the PLC members discuss the evidence they each gleaned from the lesson and their rationales for the codes given in the process of attempting to reach agreement about the best code for each element. This coding and discussion process creates a means by which the PLC members engage in fine-grained conversations not only about the observed lesson but about their practice more broadly. Simultaneously, the framework enables acknowledgment that the particular teaching context will shape the suitability or relevance of certain goals. For example, "Student direction"—giving students control over some aspects of their learning—will be less relevant in lessons where risk is high and safety is a concern (such as learning to throw a javelin or working with heat sources in science), or when the teacher wants to relay key messages to all students (such as explaining a key concept or conveying a set of instructions). This combination of specificity and contextualisation, we suggest, means that the QT framework is

"comprehensive enough to encompass teachers' concerns, manageable enough to give focus to their thinking and practice, and open enough to enable their critical engagement" (Bowe & Gore, 2017, p. 358). It seeks to resonate with the realities of the diverse contexts in which teachers work while being practical and achievable.

Moreover, as Sartain, Stoelinga, and Brown (2011) note:

An observation tool cannot promote instructional improvement in isolation. A rigorous instructional rubric plays a critical role in defining effective instruction and creating a shared language for teachers and principals to talk about instruction, but it is the conversations themselves that act as the true lever for instructional improvement and teacher development. (p. 41)

Quality Teaching Rounds is an approach to teacher professional development designed to enable such conversations. It attends to both the substance of conversations about teaching practice (through the QT framework) and the processes by which productive conversations might occur (through ensuring a safe space for their critical analysis), in order to inform ongoing refinement of their practice.

2. Methodology

Building on a series of preliminary studies (Gore, 2014), the study was a cluster randomised controlled trial (RCT) conducted in government schools in New South Wales (NSW), Australia. The methodology has been described in detail elsewhere (Gore et al., 2015). In brief, the project involved a rigorously designed RCT adhering to the Consolidated Standards of Reporting Trials (CONSORT) guidelines for group trials (Moher et al., 2010). The three-arm RCT, conducted from mid-2014 to mid-2015, involved two intervention groups (a "Set" intervention and a "Choice" intervention) and a waitlist control group. Two full lessons per teacher for 192 teachers in 24 schools (8 in each school) were observed by the research team before Quality Teaching Rounds commenced (baseline), six months later when the two intervention groups had finished, and again six months after that in order to begin to consider the sustainability of any effects.

2.1. Participants

The study involved 24 schools,[4] selected from among the 243

---

[4] The sample size required to detect changes in the primary outcome (i.e., Quality Teaching score) was based on 80% power with alpha levels set at $p < 0.05$. Based on our previous studies, we anticipated an effect size of $d = 0.5$ (adjusted mean difference of 0.25) for the Quality Teaching Rounds interventions (assuming a mean QT score of 2.6 and standard deviation of 0.5). Our power calculation was adjusted for clustering using a correction factor of $[1+ (m-1) \times ICC]$, where m = teachers per school and ICC = the intra-class correlation coefficient. Assuming that eight teachers were to be recruited from each of the 24 schools and an ICC of 0.07, the correction factor is 1.5 [i.e., $1 + (8-1) \times 0.07$]. Therefore, the required sample size was determined to be 192 teachers from 24 schools.

To what extent do students exercise some direction over the selection of activities related to their learning and the means and manner by which these activities will be done?

Student direction

① No evidence of student direction. All aspects of the lesson are explicitly designated by the teacher for students.

② Low student direction. Although students exercise some control over some aspect of the lesson (choice, time, pace, assessment), their control is minimal or trivial.

③ Some student direction. Students exercise some control in relation to some significant aspects of the lesson.

④ Substantial student direction. Some deliberation or negotiation occurs between teacher and students over at least some significant aspects of the lesson.

⑤ High student direction. Students determine many significant aspects of the lesson either independent of, or dependent on, teacher approval.

Fig. 1. Descriptors and coding scale for *Student direction*.
Reprinted from Quality Teaching in NSW Public Schools: A Classroom Practice Guide (p. 36), by NSW Department of Education and Training, 2003, Sydney, Australia: Author. Copyright © 2003 by the State of NSW, Department of Education and Training, Professional Support and Curriculum Directorate. Reprinted with permission.

schools that provided expressions of interest in response to an invitation that was sent to all 2200 government schools in the state of New South Wales, Australia. To be eligible for this study, schools were required to have at least eight teachers who were willing to participate, be prepared to accept any group allocation (including a wait-list control group that would wait 12 months to commence the PD activity), and be able to commit to the 18-month study period.[5] Eligible schools were stratified based on school type (i.e., primary or secondary), location (i.e., urban or rural), and socio-economic status[6] (SES) in order to ensure a diverse and representative sample of schools for testing the efficacy of Quality Teaching Rounds. Following stratification, 24 schools were then randomly selected for the trial, 12 each from primary and secondary schools and, within each sector, three from each of the SES categories (low, mid, high) for the urban schools and three from a fourth category (all SES) for the rural schools. Following baseline assessments of the quality of teaching, the 24 schools were matched based on SES, geographical location, and school type, and then randomly assigned by a researcher independent of the project to one of three conditions (see Fig. 2), using a computerised random number generator. The diversity of the selected schools is apparent in the following characteristics: proportion of students with language backgrounds other than English (range from 2% to 92%), proportion of Indigenous students (range from 0% to 62%), and school index of relative socio-educational advantage (range from 766 to 1209, where the average for all schools in Australia is 1000).

From the selected schools, 192 teachers participated in the study. Principals identified eight teachers to be invited, on a voluntary basis, to participate in the study, ensuring a balance of experience, grade level, and subject area, and potential benefits for individual teachers. Recruiting volunteers ensured matched learning motivation profiles for the two groups of participants, a common oversight in comparative PD studies (Kennedy, 2016). Teachers could withdraw from the study at any time. Characteristics of the teachers who participated in each arm of the study are summarised in Table 2. Note that all teachers in Australian schools are certified to teach with a minimum of four years of university study.

2.2. The intervention

Eight teachers per school participated in the Quality Teaching Rounds intervention, forming PLCs of between three and eight teachers. Prior to commencing Quality Teaching Rounds, at least four of the eight teachers from each school in the intervention groups participated in a two-day workshop which provided background information on the Quality Teaching framework and Quality Teaching Rounds, and highlighted the intention and importance of each component of the approach (i.e., PLCs, readings, observation, individual coding, group discussion). Teachers were given opportunities to practise the QT coding process and to participate in simulated Rounds using sample video-recorded lessons. They then led the Rounds back in their schools, with no further input from the research team, except for informal conversations that occurred during data collection and implementation fidelity checks.

During the time between baseline and 6-month follow-up data collection, teachers participating in the intervention carried out either the Set or Choice form of Rounds in accordance with the

---

[5] Intervention groups completed participation in the study after 12 months while the waitlist control group was involved for an additional 6 months during which time they implemented Quality Teaching Rounds.

[6] The measure used for school SES was based on the school Index of Community Socio-Educational Advantage (ICSEA) values (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2013).
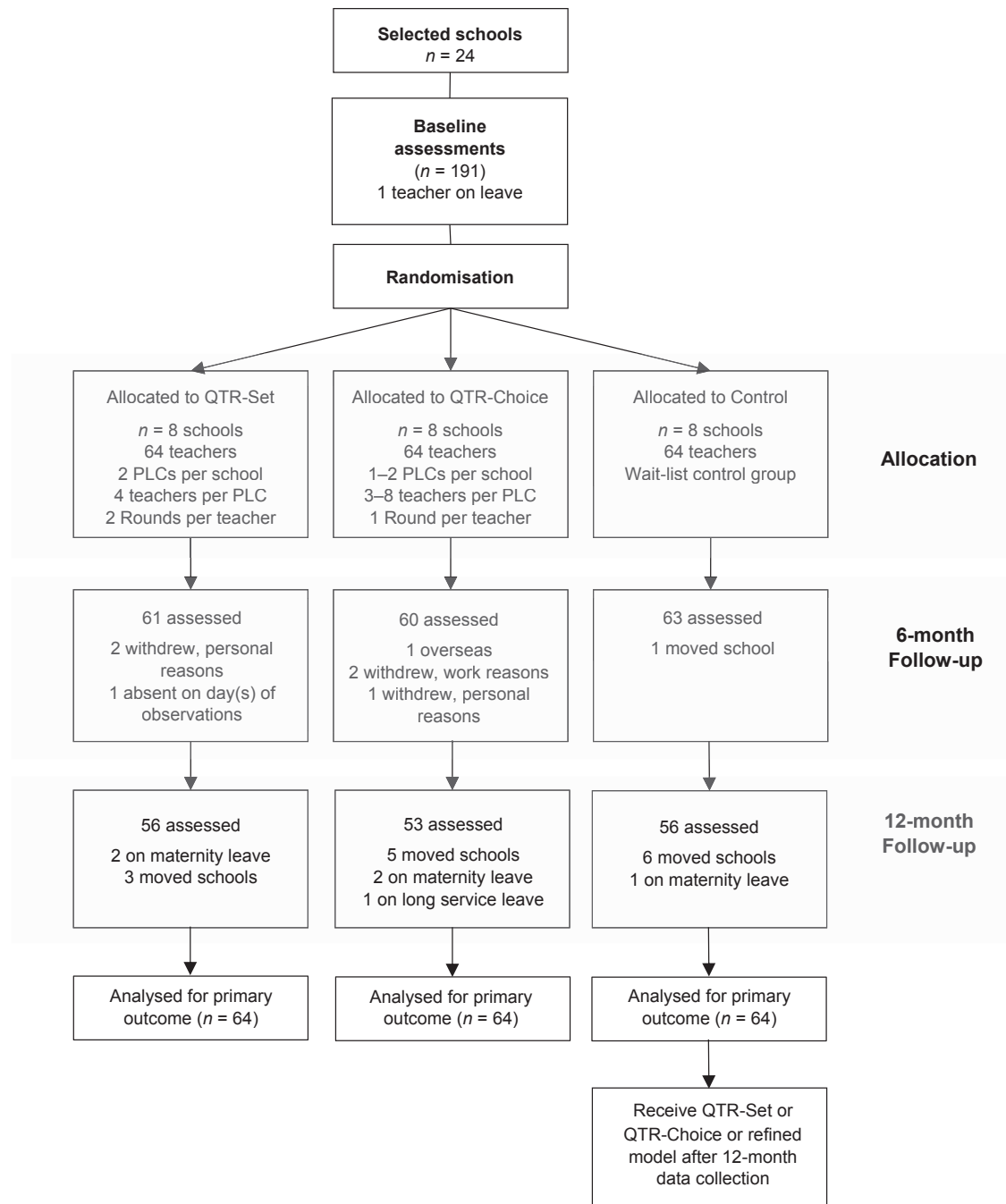
Fig. 2. Participant flow through the trial and analysed for the primary outcome (quality teaching).

allocation of their school. In the QTR-Set group, teachers formed two PLCs of four teachers each and conducted two full sets of Rounds, such that each teacher "hosted" two lesson observations. In the QTR-Choice group, teachers had more flexibility to adapt Quality Teaching Rounds to suit their school situation, with PLCs comprising between three and eight teachers and each PLC choosing whether to conduct one or more full sets of Rounds.

Intervention quality was monitored through fidelity check observations conducted by members of the research team on two separate occasions in each of the 16 intervention schools, with the researcher staying for the whole round on each occasion. The fidelity observations aimed to check compliance with the following eight essential features of QTR:

1 Was a reading discussion conducted?
2 Were PLC members present throughout the lesson?.
3 Did PLC members individually code all QT elements prior to the lesson discussion?
4 Was the post-lesson discussion at least 60 min in duration?
5 Was the host teacher included in the discussion?
6 Were PLC members (including the host teacher) present throughout the discussion?
7 Did PLC members (including the host teacher) provide their codes and justification using evidence from the lesson for each QT element?
8 Was the Quality Teaching Classroom Practice Guide a consistent point of reference throughout the discussion?

Table 2
Characteristics of participating teachers.

Teacher characteristics	Group allocation			Total (n = 191)
	Set (n = 64)	Choice (n = 63)	Control (n = 64)	
Sex, n (%)				
Male	13 (20.3)	19 (30.2)	15 (23.4)	47 (24.6)
Female	50 (78.1)	44 (69.8)	49 (76.6)	143 (74.9)
Average teacher age, years (SD)[a]	36.6 (9.6)	39.2 (8.9)	39.0 (9.7)	38.3 (9.4)
Total years of teaching, n (% of group)[a]				
< 4 years	18 (30.5)	7 (11.7)	11 (17.5)	36 (19.8)
4—15 years	27 (45.8)	38 (63.3)	36 (57.1)	101 (55.5)
16 years or more	14 (23.8)	15 (25.1)	16 (25.4)	45 (24.7)
Years of teaching at current school, n (% of group)[a]				
< 4 years	32 (54.2)	24 (40.0)	29 (46.0)	85 (46.7)
4—15 years	20 (33.9)	32 (53.3)	29 (46.1)	81 (44.4)
16 years or more	7 (11.9)	4 (6.7)	5 (7.9)	16 (8.7)
School type, n (% of group)				
Primary	29 (49.2)	34 (56.7)	35 (55.6)	98 (53.8)
Secondary	30 (51.8)	26 (43.3)	28 (44.4)	84 (46.2)
LBOTE, n (% of group)	8 (13.6)	4 (6.7)	4 (6.3)	16 (8.8)

Note. LBOTE = language background other than English; SD = standard deviation.
[a] Based on valid survey responses: Set group, n = 59; Choice group, n = 60; Control group, n = 63.

Table 3
Overview of the measures used.

Measure	Description
Observation measure —Quality Teaching	The Quality Teaching scoring instrument from the Quality Teaching Classroom Practice Guide (NSW DET, 2003) was used. Scores from 1 to 5 are generated in response to a focus question for each of the 18 elements and specific descriptors for each score. The mean of the scores for QT overall is then calculated. The Cronbach's alpha for the Quality Teaching overall scale in this study was good (18 items; α = 0.82).
Teacher survey	Teachers reported the number of lessons they have observed and had observed during the past year, the time spent during the work day in planning and preparation, and the frequency of collaboration with colleagues to share feedback, suggestions, and strategies. Other survey items form scales that provide estimates of the degree to which teachers: believe there is trust among them (6 items; α = 0.82; Bryk & Schneider, 2003), believe they take collective responsibility for student learning (6 items; α = 0.66; Louis & Marks, 1998; Louis, Marks, & Kruse, 1996), receive appraisal and recognition (6 items; α = 0.86; Hart, Wearing, Conn, Carter, & Dingle, 2000), and have high levels of morale (6 items; α = 0.86; Hart et al., 2000).
	The survey also uses scales to gauge the degree to which teachers believe that their teaching aligns with the Quality Teaching framework, by dimension and overall. These scales have the prefix 'Teaching for' to distinguish them from the observation instrument scales. The overall Teaching for Quality Teaching scale has 18 items and a Cronbach's alpha of 0.87. The subscales are Teaching for Intellectual Quality (6 items; α = 0.79), Quality Learning Environment (6 items; α = 0.67), and Significance (6 items; α = 0.71).
Demographic and background information	Details collected by questionnaire included; sex, age, identification as Aboriginal or Torres Strait Islander, language background, level of education, years of teaching at current school and overall, subject specialisation (if any), and employment status (i.e., full-time, part-time, casual).

In addition, an online fidelity checklist was completed by teachers, documenting how they had conducted Quality Teaching Rounds (one per PLC for each Round completed), in accordance with the essential features.

2.3. Outcome measures

The primary outcome, quality of teaching, was based on two observations of all participating teachers at each time point (i.e., baseline, 6 months, 12 months) through etic-type observations conducted by members of the research team, blinded to participants' group allocation. A total of 1073 whole lessons were coded by the researchers. To determine inter-rater reliability, 122 lessons were simultaneously and independently rated by two or more researchers. Using methods described by Hallgren (2012), the inter-rater reliability score (intraclass correlation) was ICC = 0.76 (95% CI [0.65, 0.83], p < 0.001). A brief description of all outcome measures is provided in Table 3 (for further detail see Gore et al., 2015).

Two teachers per school were selected to take part in semi-structured interviews with researchers during the study; where possible, one with less than three years' teaching experience and the other with more than five years' teaching experience, in order to explore possible effects related to stage of career.

2.4. Data analysis

Statistical analyses of the quantitative data were conducted using linear mixed models[7] with alpha levels set at p < 0.05. The models were used to assess the impact of the two interventions (QTR-Set or QTR-Choice), time (treated as categorical with levels: baseline, 6-months, and 12-months), and the group-by-time interaction, with these three terms forming the base model. The models were specified to adjust for the clustered nature of the data (i.e., teachers located within schools), and the multiple observations conducted at each time point, using random intercepts. Moderators of intervention effects were explored with interaction terms for: (i) type of school (primary versus secondary), (ii) SES (based on school ICSEA values), (iii) geographic location of school (urban versus rural), (iv) years of teaching experience, and (v) sex of teacher. Subgroup analyses were conducted if significant interaction effects (p < 0.1) were identified (Assmann, Pocock, Enos, & Kasten, 2000).

Our protocol (Gore et al., 2015) specified that PLCs needed to

---

[7] Using IBM SPSS Statistics for Windows, Version 22.0 (2010 SPSS Inc., IBM Company Armonk, NY).

meet at least five of the eight pre-specified standards (based on implementation fidelity checks). All schools met the pre-specified level; however, in order to explore effects when stricter standards are applied the per-protocol analysis was conducted with the level set to six of the eight standards.

Cohen's *d* was calculated using the following formula:

3.1. Primary outcome

3.1.1. Impact on quality of teaching

The quality of teaching improved significantly from baseline to post-intervention, with medium effects on overall teaching quality for intervention schools (Cohen's $d = 0.4$). Effects were sustained at

$$d = \frac{[(intervention\ posttest\ mean - intervention\ baseline\ mean) - (control\ posttest\ mean - control\ baseline\ mean)]}{pooled\ standard\ deviation\ of\ change}$$

Qualitative analysis of teacher interviews was conducted, with all interview transcriptions read by at least two members of the research team. Transcripts were coded deductively and inductively to identify themes relating to: the most valued aspects of participating in Quality Teaching Rounds; the impact of participating in Quality Teaching Rounds on teachers and students; and, the impact of participating in Quality Teaching Rounds on teaching culture and teacher identity. Although some qualitative data are used selectively to amplify aspects of our argument, this paper focuses primarily on the quantitative outcomes. Qualitative data relating to teachers' experience of Quality Teaching Rounds have been reported more comprehensively in Gore and Bowe (2015) with further analyses forthcoming.

3. Results and discussion

The flow of participants in the study is presented in Fig. 2. Baseline observations were made for 191 teachers. During the course of the study, six teachers were not available for the six-month follow-up observations and 26 were not available in the following school year for the 12-month observations, yielding 86% retention overall. Attrition from the study was a result of moving schools, going on leave, or withdrawal from the study for reasons unrelated to participation in Quality Teaching Rounds.

the 12-month follow-up assessment (Cohen's $d = 0.2-0.5$). Post hoc analyses revealed that improvements in the QTR-Choice group were not significantly different from those observed in the QTR-Set group. These changes over time are depicted in Fig. 3 and detailed in Table 4. These data indicate the impact of participating in Quality Teaching Rounds with as little as one set of Rounds for participating teachers.

Based on fidelity checks with the level of compliance set to six of eight criteria, three PLCs (two from one school and one from another, all from the QTR-Set group) were excluded. Fig. 4 and Table 5 provide the results for those PLCs that conducted the intervention as per the protocol.

This analysis indicates a stronger effect when schools and teachers carried out the intervention as intended, and cautions against dilution or ad hoc modification of the approach if results are to be replicated.

3.1.2. Moderation effects

Moderators of intervention effects were explored for: type of school, SES, location, teaching experience, and sex of teacher. Moderator analysis indicated time by intervention effects that differed for the subgroups explored for the variables SES ($p < 0.001$) and sex of teacher ($p = 0.034$); therefore subgroup analysis was progressed for each of these variables. A greater effect was observed for female teachers in the intervention group compared to the control group: an increase (adjusted difference)
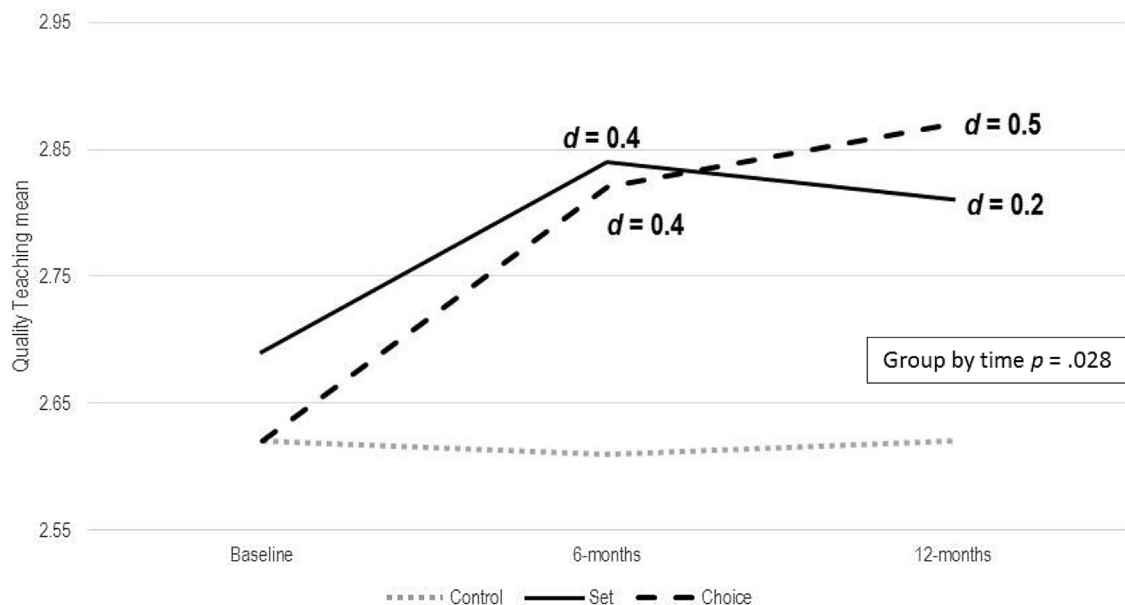

Fig. 3. Teaching quality at baseline, 6 months, and 12 months.

Table 4
Changes in teaching quality (using observation data) for teacher by treatment group from baseline to 6 months and 12 months (whole sample).

Group	Baseline		6-months			12-months			Within group[a] p	Group-by-time p	Post hoc analyses[b]
	Mean	(95% CI)	Mean	(95% CI)	d	Mean	(95% CI)	d			
Control	2.62	(2.45, 2.79)	2.61	(2.44, 2.78)	–	2.62	(2.45, 2.79)	–	0.987	**0.028**	
QTR-Set	2.69	(2.52, 2.86)	2.84	(2.67, 3.01)	0.4	2.81	(2.64, 2.98)	0.2	**0.017**		Set > Control (6M-B), p = **0.046**
QTR-Choice	2.62	(2.46, 2.79)	2.82	(2.65, 2.99)	0.4	2.87	(2.70, 3.04)	0.5	**< 0.001**		Choice > Control (6M-B), p = **0.012**
											Choice > Control (12M-B), p = **0.004**

Note. QTR-Set = Quality Teaching Rounds Set intervention; QTR-Choice = Quality Teaching Rounds Choice intervention; CI = confidence intervals; d = Cohen's d; 6M = 6-months; 12M = 12-months; B = baseline. Significant p-values < 0.05 highlighted in bold.
[a] Within group change from baseline to 12-months.
[b] Post hoc analysis results for adjusted difference between groups over time (non-significant findings not reported). Cohen's d was calculated using the following formula:
$$d = \frac{|(intervention\ posttest\ mean - intervention\ baseline\ mean) - (control\ posttest\ mean - control\ baseline\ mean)|}{pooled\ standard\ deviation\ of\ change}.$$

in QT score of 0.24 (95% CI [0.06, 0.43], p = 0.01) for the QTR-Set group, and 0.34 (95% CI [0.15, 0.53], p = 0.001) for the QTR-Choice group. For male teachers no significant differences between intervention groups and control were observed at 12-month follow-up. This suggests the intervention might have been more effective for female teachers. However, given the limited power with fewer male teachers (see Table 2), the sub-group analysis should be viewed as exploratory and findings interpreted with caution.

Exploration of the effects within SES categories identified significant findings for the high- and mid-SES subgroups. A greater effect was observed for teachers in the high-SES category for both the Set and Choice groups compared to the control group on QT score with an adjusted increase of 0.39 (95% CI [0.10, 0.68], p = 0.009) and 0.37 (95% CI [0.07, 0.66], p = 0.016) respectively. In the mid-SES category the Choice group had an increase (adjusted difference) compared to the control group on QT score of 0.78 (95% CI [0.44, 1.12], p < 0.001). In the low-SES category, the Choice group had a significant within-group effect; however, when compared to the control this was not significant. In the rural (all SES) category there were no significant differences between the intervention groups and control. Although the group-by-time interaction for SES was statistically significant, the sub-group analyses indicate that the differences in intervention effect sizes were not practically

meaningful. Previous research (Gore, Weaver, Lloyd, & Smith, 2016) has indicated the quality of teaching is no better or worse overall in more or less advantaged schools and therefore the small group differences observed in the current study are of little practical or policy consequence.

3.2. Secondary outcomes

From the teacher survey, significant changes were identified for the intervention groups in relation to the two scales of Morale and Appraisal and Recognition; these are detailed below. The scales (and sub-scales) for Teaching for Quality Teaching, as well as Teacher to Teacher Trust and Teacher Responsibility, produced no significant changes. The results for these scales are provided in the appendix. We contend that teachers did not perceive their practices as more aligned with Quality Teaching than did the control group participants in large part because they now had a deeper understanding of the framework and greater capacity to judge the quality of their own practice. The scales for trust and collective responsibility focused on whole school qualities. We contend that there might have been a difference between intervention and control groups had we asked teachers in the intervention groups about these issues within their PLCs only.
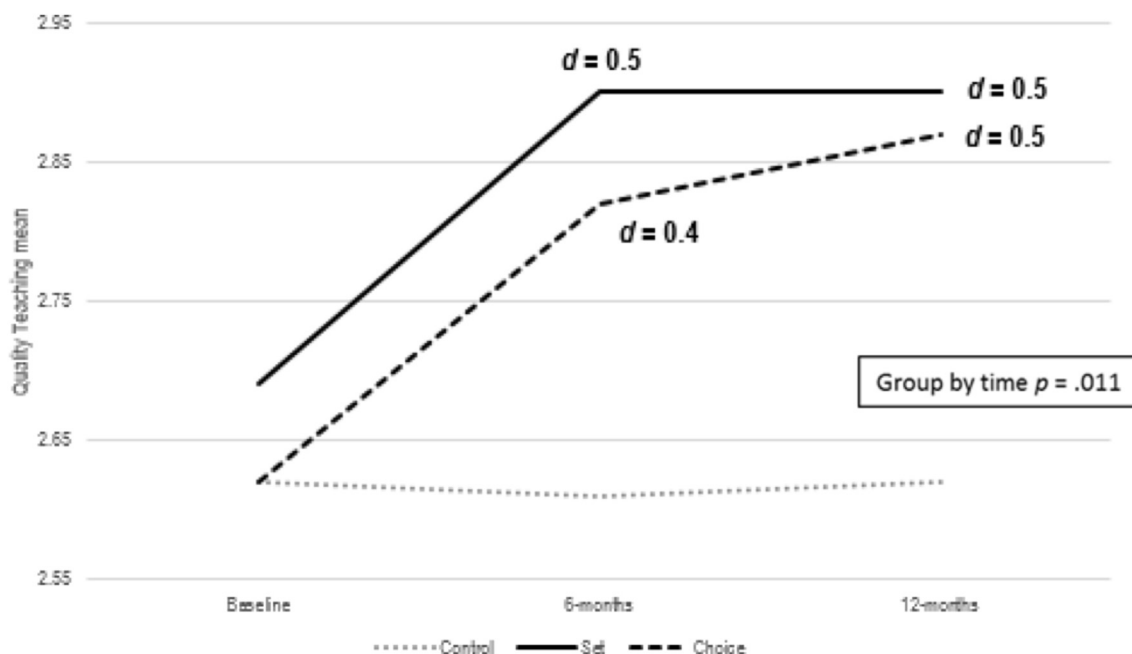

Fig. 4. Per protocol analysis (fidelity check meeting six of eight pre-specified criteria), Teaching quality at baseline, 6 months, and 12 months.

Table 5
Changes in teaching quality (using observation data) for teacher by treatment group from baseline to 6 months and 12 months − per protocol analysis (fidelity check meeting six of eight pre-specified criteria).

Group	Baseline		6-months			12-months			Within group[a] p	Group-by-time p	Post hoc analyses[b]
	Mean	(95% CI)	Mean	(95% CI)	d	Mean	(95% CI)	d			
Control	2.62	(2.44, 2.79)	2.61	(2.43, 2.79)	−	2.62	(2.44, 2.80)	−	0.986	**0.011**	
QTR-Set	2.69	(2.51, 2.86)	2.90	(2.72, 3.09)	0.5	2.90	(2.71, 3.09)	0.5	**< 0.001**		Set > Control (6M-B), $p =$ **0.008**
											Set > Control (12M-B), $p =$ **0.015**
QTR-Choice	2.62	(2.45, 2.80)	2.82	(2.64, 3.00)	0.4	2.87	(2.69, 3.05)	0.5	**< 0.001**		Choice > Control (6M-B), $p =$ **0.011**
											Choice > Control (12M-B), $p =$ **0.004**

Note. QTR-Set = Quality Teaching Rounds set intervention; QTR-Choice = Quality Teaching Rounds choice intervention; CI = confidence intervals; d = Cohen's d; 6M = 6-months; 12M = 12-months; B = baseline. Significant p-values < 0.05 highlighted in bold.
  [a] Within group change from baseline to 12-months.
  [b] Post hoc analysis results for adjusted difference between groups over time (non-significant findings not reported). Cohen's d was calculated using the following formula:
$d = \frac{|(\text{intervention posttest mean} - \text{intervention baseline mean}) - (\text{control posttest mean} - \text{control baseline mean})|}{\text{pooled standard deviation of change}}$.

3.2.1. Staff morale

Participating teachers were asked about staff morale in the surveys at the baseline (mid-2014), 6-month (end 2014), and 12-month (mid-2015) time points. The survey questions related to team spirit, enthusiasm for work, energy, and pride in the school (Hart et al., 2000). The morale scale consisted of five items on a five-point scale from strongly disagree to strongly agree. Fig. 5 below illustrates the mean scores over time for these scales (with a minimum of one, and a maximum of five). The QTR-Control group exhibited a significant decline in morale over the three time points ($p = 0.029$), while significant positive effects on teacher morale were observed for the QTR-Set group at 6-months ($d = 0.4$) and for both QTR-Set and QTR-Choice groups at 12-months ($d = 0.6$ and $d = 0.4$ respectively). It is worth noting that the 6-month data were collected at the end of the school year, when morale among teachers is notoriously low.

3.2.2. Appraisal and recognition

Participating teachers were asked about the level of appraisal and recognition experienced in their schools at the three time points. Questions included items such as receiving feedback on performance, opportunities to discuss performance, recognition of good work, and receiving encouragement (Hart et al., 2000). The scale consisted of five items on a five-point scale from "strongly disagree" to "strongly agree." Fig. 6 illustrates the mean score over time for these scales (with a minimum of one, and a maximum of five).

As with the morale scale, the control group showed a decline over time in appraisal and recognition mean scores (however not significant). In contrast, the QTR-Set and QTR-Choice intervention groups demonstrated significant gains (difference between control and intervention groups), with significant positive effects on sense of appraisal and recognition observed for the QTR-Set group at 6-months ($d = 0.4$) and for both QTR-Set and QTR-Choice groups at 12-months ($d = 0.4$ and $d = 0.5$ respectively).

3.3. Significance of effects

The primary aim of the study was to test the effectiveness of Quality Teaching Rounds as an approach to teacher PD for its capacity to impact on teaching quality. We found that a relatively short-term intervention (i.e., two-day training session for some participants; one or two sets of Rounds) produced, on average, a meaningful effect for teachers in schools randomly assigned to the intervention conditions (Cohen's $d = 0.4$), with the effect sustained six months post-intervention (Cohen's $d = 0.2-0.5$). It is worth noting that we found no difference between the two forms of intervention, indicating that gains in teaching quality resulted from a single set of Quality Teaching Rounds involving teachers in as little as four half days of professional development. Such effects on the quality of teaching have rarely been reported in other studies.

3.3.1. Distinguishing features
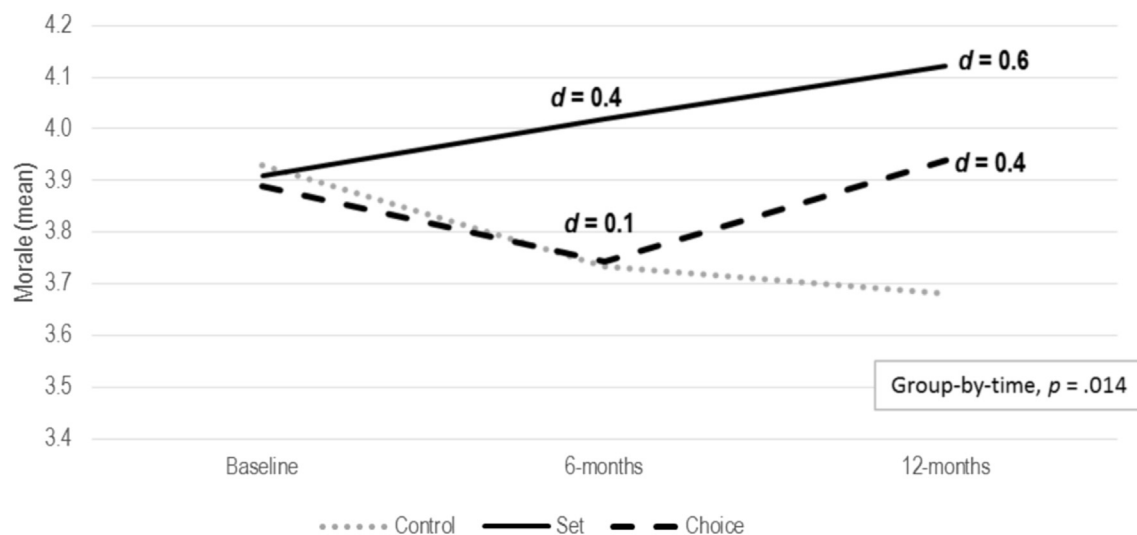
Where specific effects on teaching have been reported, most

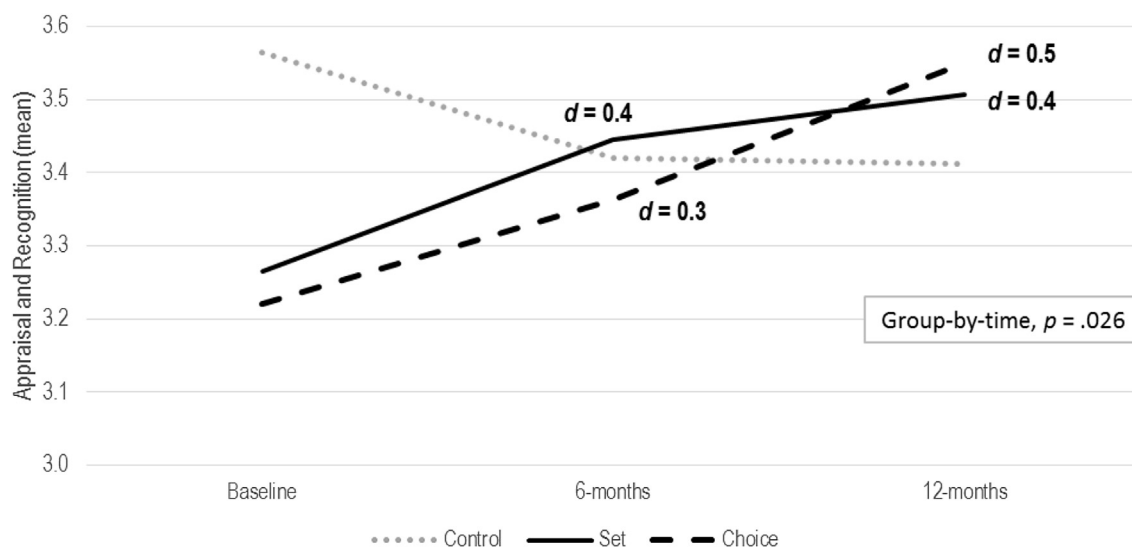**Fig. 5.** Teachers' perceptions of staff morale (5-item scale).

Fig. 6. Teachers' perceptions of appraisal and recognition (5-item sale).

studies have focused on improvements to a particular aspect of teaching, such as teacher-student interactions (De Roos, Van de Heijden & Gorter, 2010; Domitrovich et al., 2009; Pianta et al., 2008), classroom management (Akalin & Sucuoglu, 2015), or formative assessment (Furtak et al., 2016). Alternatively, studies have reported impacts on specific subject-based teaching such as science process skills (Cotabish, Dailey, Hughes, & Robinson, 2011), science pedagogy (Stevenson, Stevenson, & Cooner, 2015), facilitation of mathematical and scientific thinking (Van Whittaker, Kinzie, Williford, & DeCoster, 2016), effective mathematics teaching (Bruce & Ross, 2008), strategies targeting phonological awareness (Cunningham, Etter, Platas, Wheeler, & Campbell, 2015), or language and literacy more broadly (Landry, Swank, Anthony, & Assel, 2011; Neuman & Cunningham, 2009).

Our study's focus on the overall quality of teaching also goes beyond documenting increases in teachers' use of specific instructional practices as a result of PD (Desimone, Porter, Garet, Suk Yoon, & Birman, 2002; Kennedy, 2016), especially when supported by mentoring or coaching. These studies show that training in particular techniques enables teachers to reproduce those techniques more frequently or more consistently. Our intervention is applicable across subject and content areas and across grade levels, with potential transferability of teachers' learning to a broader range of teaching contexts—not just "this topic," "this subject," or "this set of skills," but "teaching in general." A specific content focus or specific targeted practice has been considered a core component of effective PD (Desimone, 2009). We suggest that this view may be derived more from what was measured/able to be measured than what is possible to measure. As a result, few studies address the *quality* of practices and when they do, they either tend to be very specific, such as quality of classroom interactions (Pianta et al., 2008), or quality is implied in the targeted focus.

Exceptions are one experimental study reported by Antoniou, Kyriakides, and Creemers (2011) who contrasted their "dynamic integrated approach" (pp. 16–19) to teacher professional development with a more conventional "holistic/reflective approach" (pp. 15–16). The dynamic integrated approach addressed eight factors describing instruction—orientation, structuring, questioning, teaching-modelling, applications, management of time, teacher role in setting the learning environment, and classroom assessment—which were measured using the five dimensions of frequency, focus, stage, quality, and differentiation. These researchers found that teaching skills improved more over the course of one

year for those exposed to the dynamic, integrated form of PD than for those encouraged to reflect on their teaching practices and beliefs in a more general way, with both groups participating in nine PD sessions. Similarly, Van de Pol, Volman, Oort, and Beishuizen (2014), who defined teaching quality as teachers' capacity to effectively scaffold learning using the four steps of diagnostic strategies, checking diagnoses, giving contingent support, and checking student learning, found that the teachers in the intervention group increased the quality of their teaching as defined in this way. These studies indicate the importance of specificity in guiding teachers' improvement efforts.

As a form of PD, QTR also differs from many PD interventions that rely on ongoing support following the PD intervention (typically for six to twelve months). After the initial workshop (which was attended by half of the teachers), teachers participating in Quality Teaching Rounds supported themselves. It is also noteworthy that a relatively short intervention produced sustained effects six months later, for teaching in general and in a new school year. In studies with a reported effect on student achievement, teachers were engaged, on average, in more than 80 h of PD (Supovitz & Turner, 2000). By contrast, teachers averaged 42 h of PD in the Set intervention and only 24 h in the Choice intervention (including the time in the two-day workshop for half of the participants).

3.3.2. Meaningful effects

As with all statistical results, the practical significance of our findings is worth further consideration. The effects on quality of teaching (0.4–0.5) are considered moderate in statistical terms. However, in educational research where effect sizes are often reported in relation to the impact of interventions on aspects of student achievement, effect sizes of 0.3 are typically considered meaningful (Coe, 2002; Sanders & Ní Chonaire, 2015).

Fewer studies report effect sizes in relation to changes in teaching practices or teaching quality. We stress the significance of the reported effects with a primary outcome measure of teaching quality that is as multi-faceted and translatable to all teaching contexts as the Quality Teaching measure. Our findings are more complex than students mastering new cognitive or linguistic tasks or teachers reproducing a specific set of skills. Rather, our results signal a change in the overall quality of teaching—where the measure of quality includes assessment of the treatment of knowledge, the level of understanding produced among students,

expectations conveyed to students, support for learning in the classroom environment, and significance of the learning activities. To achieve effects of 0.4–0.5 with such an instrument for both intervention groups, with no change in the control group, while using strict CONSORT protocols including independent randomisation and blinded observers, is a noteworthy outcome. This is especially so with such a short-term intervention primarily involving teachers themselves rather than external consultants or expertise, after the initial workshop. Hence the main costs in running the intervention related to releasing teachers to participate in Quality Teaching Rounds. Demonstrably, the return on investment was high.

3.4. Why the approach works

As with other RCTs, this study is useful in providing evidence of the effectiveness of an intervention; in this case, an approach to PD making a difference to the quality of teaching practice. Equally important is understanding *how* and *why* the approach works (Connolly, 2009)—questions that are informed by the qualitative evidence gathered. As noted by Lewis, Perry, and Murata (2006), "[i]nnovations often fail when educators focus only on the surface features of the innovation rather than the underlying mechanism[s] that will enable it to work" (p. 5). We argue that Quality Teaching Rounds works both because of the specific nature of the QT framework and because of the way that the framework is used to guide teachers in the observation, analysis, discussion, and refinement of each other's teaching. Specifically, we contend that three key "mechanisms" of Quality Teaching Rounds are likely explanations for the effect on quality of teaching practice for teachers in the intervention groups; namely that Quality Teaching Rounds impact on: (1) the knowledge base for teaching; (2) the power relations among collaborating teachers; and, (3) teaching culture in schools through new professional relationships among teachers. Each of these mechanisms is addressed briefly, illustrated with excerpts from the interviews with participating teachers.

3.4.1. A mechanism for structuring the knowledge base for teaching

The QT framework does particular kinds of work. It provides a structure for thinking about the practice of teaching that organises the many discrete skills and practices into three key ideas—the dimensions of Intellectual Quality, Quality Learning Environment, and Significance—elaborated through the 18 elements of the model. In so doing, it reduces the multiplicity of teaching's demands without denying its complexity. The dimensions are comprehensive in focus, addressing the intellectual demands of the work, the learning support provided by and expectations of teachers, and ways of connecting school learning to broader issues in students' lives. The QT framework articulates joint concerns for academic excellence, equity (especially for those who have been most disadvantaged by schooling), and curriculum relevance (Gore, 2007), without pitching these goals in opposition to each other as has sometimes been the case historically (Black, 2002).

In structuring the knowledge base for teaching, the QT framework also provides a level of specificity and guidance for teachers that is rare, especially for such a comprehensive framework, while doing so in ways that are adaptable to local contexts (subjects, topics, lessons, classes). While some commentators react negatively to any system involving numbers, it is the way in which the QT framework resists simplistic adding of numbers and rating of individuals, insists on using the approach only for PD purposes, and focuses on teaching rather than teachers, that guards against performativity (Ball, 2003). At the same time, it enables enough calibration/specificity to guide teachers and give them confidence in

their own and each other's analyses (Bowe, 2016). High levels of buy-in and commitment to using the framework have been found (Gore & Bowe, 2015), features of PD which are essential for sustaining long-term growth in teaching quality.

The QT framework's structuring of the work of teaching (or instruction) makes "good teaching" more accessible and in so doing helps teachers feel more confident about their work and helps them makes sense of the complex and critical activity of teaching, as illustrated:

> It gives you … a clearer lens to look through or a framework to base your teaching on … it's nice to have a little bit of a framework to … pin your teaching on, to know that you're doing the right thing really.… It's given us greater confidence in what we're doing because we can see that the programs we run … are addressing the Quality Teaching model, and it gives you greater confidence in your teaching and what you're doing. (Molly, primary teacher of 5 years)

This teacher highlights the role of the framework in guiding teachers toward the "right thing," thus enhancing their confidence in teaching. Given that teaching is such a complex activity and the knowledge needed for teaching is so wide-ranging, the articulation of and guidance for good teaching provided by the Quality Teaching framework is likely to be a central factor in the increased quality found for the intervention groups.

Another teacher articulated how the shared knowledge base provided by the Quality Teaching framework impacted on the quality of conversations among teachers:

> I really like the way that you actually have the [Quality] Teaching framework there as the basis …. It's not just a matter of "that lesson worked well." It's not a free-for-all discussion. It is actually guided by the framework which is really good because you've got that common language and everybody kind of understands where you're coming from, and I think I really valued that. (Jade, primary teacher of 3 years)

This teacher signals how the common language helped build understanding among teachers, a feature of Quality Teaching Rounds that contributes to the positive experience of collaboration. The framework also provided teachers with a sense of being safe, despite opening their classrooms to colleagues in ways that typically increase teachers' sense of vulnerability, as clearly articulated by one of the beginning teachers:

> Look, I think what makes it safe is having that [Quality Teaching] Classroom Practice Guide because you can really depersonalise it and say, "Well, going on the language here, this is what the coding needs to be, because that is the evidence we have to be able to back it up." So, because you're working in that framework with that language as your guide it does make it really safe. (Tilly, primary teacher of 1–3 years)

Access to a defined knowledge base for teaching enabled teachers to have discussions that felt "safe" and "depersonalised." This aspect of Quality Teaching Rounds, together with increasing confidence in their own and each other's teaching, helps to explain the growth in morale and sense of recognition for their work among the teachers in the intervention groups.

3.4.2. A mechanism for enhancing collaboration by flattening power hierarchies

One of the major barriers to successful collaboration among teachers is the underlying power dynamics that authorise some teachers to dominate in group discussions or produce a level of politeness that gets in the way of critical analysis (Hargreaves,

1994). The organisation of Quality Teaching Rounds whereby all PLC members take their turn to host a Round, subjecting their lesson to observation and analysis, acts to "level the playing field." Here, a beginning teacher indicated how participation enabled her to speak among more experienced peers:

> I think I'm … more willing to interact with the [other teachers]—I don't feel as intimidated. As a new teacher you sort of sit back and go, "I don't really know what they're talking about" but QT has given me a lot more confidence so I know I can speak a bit more because I'm more confident as a teacher. I'm not afraid to speak in front of people, even if I'm wrong. (Kate, secondary teacher of 2 years)

The process of Quality Teaching Rounds is not about "giving" a teacher feedback or "coaching" from a position of greater expertise, but about a group of teachers discussing each other's teaching using a lesson taught by each teacher as the basis for their discussions. Everyone in a PLC, whether a beginning teacher, a highly experienced teacher, or a principal, participated at all stages of the process—being observed, coding the lesson (observers and teacher), and taking turns to speak in the analytical discussions:

> One thing that I've really liked about this is that there's just been the four of us, so you're almost—not forced, but you have to speak up. You do get the opportunity to share your thoughts and to speak. I know when we were coding, you know, we'd have to justify why we gave … that particular code.… the best thing about this is that it's been in a small [group]. I've gotten to really know three other teachers and their practices throughout this time. I feel like I've been heard. (Victoria, primary teacher of 2 years)

As a commitment when doing Quality Teaching Rounds, it is critical that PLC members agree to uphold confidentiality and use lesson evidence for the purposes of the PLC discussions and not for evaluative or rating purposes. As a result of feeling safe to do this critical analytical work, Quality Teaching Rounds enabled teachers to speak and be heard, whether relatively inexperienced, as in the quotes above, or highly experienced as follows:

> This is like learning by doing or refreshing … and then thinking again …. Just having people talk to you and … taking you back to step one. Because when you see stuff then it all starts flooding back to you and it was … you know, learning from experience. (Tony, secondary teacher of more than 24 years)

The processes within Quality Teaching Rounds ensured that no matter the level of experience or position of authority within a school, professional learning was the likely outcome.

3.4.3. A mechanism for enhancing professional relationships to build a culture of learning among teachers

Quality Teaching Rounds typically involve teachers in PLCs from different subject areas and/or different grade levels working together. This is quite different from professional development that emphasises subject-specific knowledge as the highest priority. The Quality Teaching Rounds approach treats teaching holistically, recognising that what teachers know will always be mediated by what they do in the classroom. In addition to addressing the knowledge to be taught and how to engage students in powerful learning experiences, QTR gives at least equal attention to teachers' knowledge *of* teaching (that is, teachers' broad pedagogical knowledge) in the quest for improving teaching quality. As a comprehensive pedagogical framework to guide reflection, analysis, and improvement, QT offers an approach that enables teachers

to work across subject and grade level divides to learn from each other, thus providing a broader conceptualisation of what teachers need to know if they are to succeed in the classroom.

Working across the typical boundaries between disciplines and grade levels (common staffroom divides) fostered new relationships and new respect within schools that helped build a culture of support and collaboration:

> Doing the Rounds was the best thing that I've been able to do. Building that relationship with the other teachers … within different faculties that I didn't really have a relationship with before, and being able to go into that faculty and get assistance for certain things especially across-[discipline] ideas. It strengthened my relationship so that this year I organised a medieval day and I could get the food tech teachers—they catered. Then the drama teachers helped, and it was really good across [disciplines]. Yes, it strengthened relationships in terms of going to get assistance from other faculties and other areas and didn't make it so segregated. (Christina, secondary teacher of 1 year)

Some participants highlighted that Quality Teaching Rounds enabled new relationships with colleagues:

> And it's good to work with people that I don't normally work with.… And the fact that we aren't all on the same stage just brings us that little bit [closer]—like, I've got to know one lady …. I didn't even realise she was a teacher here before, particularly in the middle of last year—I thought she was just a parent that didn't leave. I didn't even realise! It's pretty sad … the school is so big I didn't realise …. I've actually been involved in her lessons and got to know her through this, and her passion; yeah, it's been good. (Tessa, primary teacher of 11 years)

Other teachers highlighted how Quality Teaching Rounds enabled new understandings of their colleagues, and enhanced levels of trust and respect:

> They did not like me, and I did not like them, and it was only on hearsay and reputation alone, and they did not know me from a bar of soap, I did not know them from a bar of soap. But when I was in the room with them and working with them, I respected them and I learned to trust them and I learned who they really were. (Karen, secondary teacher of 22 years)

These enriched professional relationships among teachers contributed to building a culture of learning within schools:

> I think in terms of impact on myself and my colleagues and the kids, I think really this [QTR] has been the biggest winner to be honest [because] … breaking down the barriers going into other peoples' classrooms to share, that collegiate feeling. The kids, probably giving them a more engaging set of activities and the way that I present the work in the classroom, just more thought goes into that, and I think as a whole school initiative, you know, everyone's involved so everyone seems to be on board and we have that common goal to work towards. (Michelle, secondary teacher of 27 years)

We argue that the improvement in quality of teaching for our intervention groups was not simply the result of teachers working collaboratively with a pedagogical framework to guide their practice and analysis. The particular qualities of the QT framework and ways of using it in the context of Quality Teaching Rounds, as outlined in the three mechanisms addressed above, are critical to the effects produced.

4. Conclusion

The findings from this study highlight how robust pedagogical frameworks such as QT might be used to guide the substance of preservice and/or inservice teacher development and enhance collaborative processes for such professional development. Moreover, our results broaden the international evidence base for improving teaching. Although our approach might be considered part of the most recent practice-based turn in teacher education (Zeichner, 2012), we believe the QT framework and its use in Quality Teaching Rounds does different kinds of work. QT is not just about practices of and for teaching but includes concerns for social justice, connecting school learning to broader social issues, and the treatment of knowledge. It thus brings together key commitments of different traditions within teacher education (as argued by Gore, 2001), going beyond the social efficiency and competence-based tradition that characterises much of the earlier work in this field. Moreover, the Quality Teaching framework is not simply practice-based in the sense of practising discrete teaching skills, but instead is utilised in the context of full lessons. It develops inquiry habits and productive ways of collaborating with colleagues in the assessment and refinement of teaching. In this way, the approach strengthens rather than diminishes the intellectual and professional aspects of teaching.

With demonstrated effects of our intervention on the quality of teaching, and on teacher morale and recognition, this study makes a significant contribution to the field of teacher professional development and provides an exciting foundation for further studies into such matters as the longer term sustainability of impact, impact on teachers at different career stages, the impact of ongoing or longer term participation in Quality Teaching Rounds, the impact on student outcomes of various kinds (Ladwig, 2010), and impact across national contexts. Granted limitations posed by complex and intransigent social, political, and economic conditions, improving teaching remains a key policy goal that is widely considered as fundamental to improving student outcomes. This study's demonstration of improvement in the quality of teaching for a representative group of teachers in a diverse range of schools and teaching contexts signals its potential value across school and community settings.

Funding

Appendix

Changes in secondary outcome variables (using survey data) for teacher by treatment group from baseline to 6 months and 12 months.

Outcomes	Group	Baseline		6-months			12-months			Within group[a] p	Group-by-time p	Post hoc analyses[b]
		Mean	95% CI	Mean	95% CI	d	Mean	95% CI	d			
Quality Teaching	Control	3.82	(3.69, 3.96)	3.90	(3.77, 4.04)	—	3.84	(3.70, 3.98)	—	0.207	0.157	
	QTR-Set	3.91	(3.78, 4.05)	3.95	(3.81, 4.08)	—	4.01	(3.87, 4.15)	—	0.159		
	QTR-Choice	3.83	(3.69, 3.96)	3.95	(3.81, 4.08)	—	3.82	(3.69, 3.96)	—	**0.025**		
Intellectual Quality	Control	3.89	(3.72, 4.05)	3.97	(3.80, 4.13)	—	3.94	(3.77, 4.11)	—	0.387	0.439	
	QTR-Set	4.03	(3.87, 4.20)	4.07	(3.90, 4.23)	—	4.15	(3.98, 4.32)	—	0.193		
	QTR-Choice	3.92	(3.76, 4.08)	4.09	(3.92, 4.26)	—	4.03	(3.86, 4.20)	—	**0.025**		
Quality Learning Environment	Control	3.67	(3.53, 3.80)	3.72	(3.59, 3.86)	—	3.82	(3.68, 3.96)	—	**0.010**	0.190	
	QTR-Set	3.84	(3.70, 3.98)	3.93	(3.79, 4.07)	—	4.03	(3.89, 4.17)	—	**0.001**		
	QTR-Choice	3.78	(3.64, 3.92)	3.89	(3.75, 4.03)	—	3.84	(3.69, 3.98)	—	0.114		
Significance	Control	3.79	(3.66, 3.93)	3.90	(3.76, 4.04)	—	3.77	(3.63, 3.91)	—	**0.032**	0.153	
	QTR-Set	3.78	(3.64, 3.92)	3.79	(3.65, 3.93)	—	3.84	(3.70, 3.98)	—	0.560		
	QTR-Choice	3.68	(3.54, 3.81)	3.77	(3.63, 3.91)	—	3.77	(3.63, 3.92)	—	0.147		
Teacher to Teacher Trust	Control	5.40	(4.94, 5.86)	4.31	(3.85, 4.77)	—	4.38	(3.92, 4.85)	—	**0.000**	0.302	
	QTR-Set	5.43	(4.97, 5.89)	4.58	(4.12, 5.05)	—	4.73	(4.26, 5.20)	—	**0.000**		
	QTR-Choice	5.45	(4.99, 5.91)	4.35	(3.88, 4.81)	—	4.43	(3.96, 4.89)	—	**0.000**		
Teacher Responsibility	Control	4.28	(3.99, 4.58)	4.61	(4.31, 4.91)	—	4.03	(3.72, 4.33)	—	**0.000**	0.664	
	QTR-Set	4.42	(4.12, 4.72)	4.82	(4.52, 5.12)	—	4.15	(3.85, 4.46)	—	**0.000**		
	QTR-Choice	4.42	(4.12, 4.72)	4.82	(4.52, 5.12)	—	4.06	(3.75, 4.36)	—	**0.000**		
Morale	Control	3.93	(3.43, 4.43)	3.73	(3.23, 4.24)	—	3.68	(3.17, 4.19)	—	**0.029**	**0.014**	
	QTR-Set	3.91	(3.40, 4.41)	4.02	(3.51, 4.53)	0.4	4.12	(3.61, 4.63)	0.6	0.122		Set > Control (B-6M), p = **0.027** Set > Control (B-12M), p = **0.002**
	QTR-Choice	3.89	(3.38, 4.39)	3.74	(3.24, 4.25)	0.1	3.94	(3.43, 4.45)	0.4	0.152		Choice > Control (B-12M), p = **0.041**
Appraisal and Recognition	Control	3.56	(3.13, 4.00)	3.42	(2.98, 3.86)	—	3.41	(2.97, 3.86)	—	0.293	**0.026**	
	QTR-Set	3.26	(2.82, 3.71)	3.45	(3.00, 3.89)	0.4	3.51	(3.06, 3.96)	0.4	0.092		Set > Control (B-6M), p = **0.037** Set > Control (B-12M), p = **0.016**
	QTR-Choice	3.22	(2.78, 3.66)	3.36	(2.92, 3.81)	0.3	3.55	(3.10, 4.00)	0.5	**0.022**		Choice > Control (B-12M), p = **0.004**

Note. QTR-Set = Quality Teaching Rounds Set intervention; QTR-Choice = Quality Teaching Rounds Choice intervention; CI = confidence intervals; d = Cohen's d; 6M = 6-months; 12M = 12-months; B = baseline. Significant p-values < 0.05 highlighted in bold.

[a] Within group change from baseline to 12-months.

[b] Post hoc analysis results for adjusted difference between groups over time (non-significant findings not reported). Cohen's d was calculated using the following formula:

$$d = \frac{[(\text{intervention posttest mean} - \text{intervention baseline mean}) - (\text{control posttest mean} - \text{control baseline mean})]}{\text{pooled standard deviation of change}}.$$

References

Aelterman, N., Vansteenkiste, M., Van Keer, H., De Meyer, J., Van den Berghe, L., & Haerens, L. (2013). Development and evaluation of a training on need-supportive teaching in physical education: Qualitative and quantitative findings. *Teaching and Teacher Education, 29,* 64–75. http://dx.doi.org/10.1016/j.tate.2012.09.001.

Akalin, S., & Sucuoglu, B. (2015). Effects of classroom management intervention based on teacher training and performance feedback on outcomes of teacher-student dyads in inclusive classrooms. *Educational Sciences: Theory and Practice, 15*(3), 739–758.

Antoniou, P., Kyriakides, L., & Creemers, B. (2011). Investigating the effectiveness of a dynamic integrated approach to teacher professional development. *Center for Educational Policy Studies Journal, 1*(1), 12–41.

Armour, K., & Yelling, M. (2007). Effective professional development for physical education teachers: The role of informal, collaborative learning. *Journal of Teaching in Physical Education, 26,* 177–200.

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet, 355*(9209), 1064–1069. http://dx.doi.org/10.1016/s0140-6736(00)02039-0.

Australian Curriculum Assessment and Reporting Authority (ACARA). (2013). *Guide to understanding 2012 Index of community Socio-educational Advantage (ICSEA) values.* Sydney, Australia: Author.

Australian Government Department of Education and Training. (2016). *uCube.* Retrieved April 30, 2016 from http://highereducationstatistics.education.gov.au.

Avalos, B. (2011). Teacher professional development in *Teaching and Teacher Education* over ten years. *Teaching and Teacher Education, 27,* 10–20.

Ball, S. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy, 18*(2), 215–228. http://dx.doi.org/10.1080/0268093022000043065.

Black, S. (2002). Stretching students' minds. *American School Board Journal, 188*(6), 31–33.

Blank, R. K., & de las Alas, N. (2009). *Effects of teacher professional development on gains in student achievement: How meta-analysis provides scientific evidence useful to education leaders (Report).* Washington, DC: Council of Chief State School Officers.

Bowe, J. M. (2016). *Quality Teaching Rounds: Strengthening the knowledge base and collaborative processes for teacher professional development* (Unpublished doctoral dissertation). Newcastle, Australia: University of Newcastle.

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. http://dx.doi.org/10.1080/13540602.2016.1206522.

Bowles, T., Hattie, J., Dinham, S., Scull, J., & Clinton, J. (2014). Proposing a comprehensive model for identifying teaching candidates. *The Australian Educational Researcher, 41*(4), 365–380.

Bruce, C., & Ross, J. A. (2008). A model for increasing reform implementation and teacher efficacy: Teacher peer coaching in grades 3 and 6 Mathematics. *Canadian Journal of Education, 31*(2), 346–370.

Bryk, A. S., & Schneider, B. (2003). Trust in schools: A core resource for school reform. *Educational Leadership, 60*(6), 40–45.

Connolly, P. (2009). The challenges and prospects for educational effectiveness research. *Effective Education, 1*(1), 1–12. http://dx.doi.org/10.1080/19415530903105570.

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?* London, UK: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. Retrieved from http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=395&language=en-US.

Cotabish, A., Dailey, D., Hughes, G. D., & Robinson, A. (2011). The effects of a STEM professional development intervention on elementary teachers' science process skills. *Research in the Schools, 18*(2), 16–25.

Council for the Accreditation of Educator Preparation. (CAEP). (2015). *CAEP evidence guide.* Washington, DC: Author.

Coe, R. (2002). "It's the effect size, stupid": What effect size is and why it is important. Paper presented at the annual conference of the British Educational Research Association, Exeter, United Kingdom. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm.

Cuban, L. (2013). *Inside the black box of classroom practice: Change without reform in American education.* Cambridge, MA: Harvard Education Press.

Cunningham, A. E., Etter, K., Platas, L., Wheeler, S., & Campbell, K. (2015). Professional development in emergent literacy: A design experiment of teacher study groups. *Early Childhood Research Quarterly, 31,* 62–77.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives, 8*(1). http://dx.doi.org/10.14507/epaa.v8n1.2000.

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

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*(8), 597–604.

De Roos, S. A., Van de Heijden, M. H. R. M. A., & Gorter, R. J. (2010). Professional development schools and early childhood education: Interactive skills of students, playgroup and kindergarten teachers. *European Early Childhood Education Research Journal, 18*(1), 15–28.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.

Desimone, L. M., Porter, A. C., Garet, M. S., Suk Yoon, K., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *American Educational Research Association, 24*(2), 81–112.

Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal, 46*(2), 567–597.

Elmore, R. F. (2007). Professional networks and school improvement. *School Administrator, 64*(4), 20–25.

Evans, L. (2014). Leadership for professional development and learning: Enhancing our understanding of how teacher develop. *Cambridge Journal of Education, 44*(2), 179–198. http://dx.doi.org/10.1080/0305764X.2013.860083.

Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., et al. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science, 44*(3), 267–291.

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. http://dx.doi.org/10.3102/00028312038004915.

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher, 44*(2), 87–95. http://dx.doi.org/10.3102/0013189x15574905.

Gore, J. M. (2001). Beyond our differences: A reassembling of what matters in teacher education. *Journal of Teacher Education, 52*(2), 124–135.

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

Gore, J. M. (2014). Towards quality and equity: The case for Quality Teaching Rounds. In *Proceedings of the Australian Council for Educational Research (ACER) Research Conference, "Quality and equity: What does research tell us?"* (pp. 86–91). Melbourne, Australia: ACER.

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. http://dx.doi.org/10.1016/j.ijer.2015.05.006.

Gore, J., Barron, R. J., Holmes, K., & Smith, M. (2016). Who says we are not attracting the best and brightest? Teacher selection and the aspirations of Australian school students. *The Australian Educational Researcher, 43,* 527–549. http://dx.doi.org/10.1007/s13384-016-0221-8.

Gore, J. M., Smith, M., Bowe, J., Ellis, H., Lloyd, A., & Lubans, D. (2015). Quality Teaching Rounds as a professional development intervention for enhancing the quality of teaching: Rationale and study protocol for a cluster randomised controlled trial. *International Journal of Educational Research, 74,* 82–95. http://dx.doi.org/10.1016/j.ijer.2015.08.002.

Gore, J., Weaver, N., Lloyd, A., & Smith, M. (2016). *Teaching experience and relative socio-educational school advantage: Assessing their impact on classroom management and effective teaching* (Report prepared for the Grattan Institute). Newcastle, Australia: The University of Newcastle.

Grissom, J. A., & Youngs, P. (Eds.). (2015). *Improving teacher evaluation systems: Making the most of multiple measures.* New York, NY: Teachers College Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3), 381–391.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100,* 267–271.

Hargreaves, A. (1994). *Changing teachers, changing times: Teachers work and culture in the postmodern age.* London, UK: Cassell.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95,* 798–812.

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. *British Journal of Educational Psychology, 70*(2), 211–228. http://dx.doi.org/10.1348/000709900158065.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Hoboken, NJ: Taylor & Francis.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher, 42*(9), 476–487. http://dx.doi.org/10.3102/0013189X13512674.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives, 13*(10), 1–28.

Jackson, P. W. (1968). *Life in classrooms.* New York, NY: Holt, Rinehart and Winston.

Kennedy, A. (2005). Models of continuing professional development: A framework for analysis. *Journal of In-service Education, 31*(2), 235–250.

Kennedy, A. (2014). Understanding continuing professional development: The need

for theory to impact on policy and practice. *Professional Development in Education, 40*(5), 688–697. http://dx.doi.org/10.1080/19415257.2014.955122.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980.

Korthagen, F. (2016). *Inconvenient truths about teacher learning: Towards professional development 3.0. Teachers and Teaching: Theory and Practice.* Advance online publication. http://dx.doi.org/10.1080/13540602.2016.1211523.

Ladwig, J. G. (2010). Beyond academic outcomes. *Review of Research in Education, 34*, 113–141. http://dx.doi.org/10.3102/0091732X09353062.

Ladwig, J. G., & King, M. B. (2003). *Quality teaching in NSW public schools: An annotated bibliography.* Sydney, Australia: NSW Department of Education and Training/Professional Support and Curriculum Directorate.

Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education, 61*(1–2), 21–34. http://dx.doi.org/10.1177/0022487109347321.

Landry, S. H., Swank, P. R., Anthony, J. L., & Assel, M. A. (2011). An experimental study evaluating professional development activities within a state funded pre-kindergarten program. *Reading and Writing, 24*, 971–1010.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3–14.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education, 106*(4), 532–575. http://dx.doi.org/10.2307/1085627.

Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33*(4), 757–798.

Metzler, M. W. (2014). Teacher effectiveness research in physical education: The future isn't what it used to be. *Research Quarterly for Exercise & Sport, 85*, 14–19.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., et al. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal, 340*, c869. http://dx.doi.org/10.1136/bmj.c869.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532–566.

Newmann, F. M. (1991). Classroom thoughtfulness and students' higher order thinking: Common indicators and diverse social studies courses. *Theory and Research in Social Education, 19*, 410–433. Retrieved from http://scholarcommons.usf.edu/coedu_pub/71.

Newmann, F. M. (1996). *Authentic achievement: Restructuring schools for intellectual quality.* New York, NY: Jossey-Bass.

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*, 280–312.

NSW Department of Education and Training. (NSW DET). (2003). *Quality teaching in NSW public schools: A classroom practice guide.* Sydney, Australia: NSW Department of Education and Training/Professional Support and Curriculum Directorate.

Organisation for Economic Co-operation and Development. (OECD). (2005). *Teachers matter.* Paris, France: Author.

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40*(7), 331–337. http://dx.doi.org/10.3102/0013189x11421826.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) manual: Pre–K.* Baltimore, MD: Brookes.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252. http://dx.doi.org/10.1257/0002828041302244.

Rowe, K. (2003). The importance of teacher quality as a key determinant of students' experiences and outcomes of schooling. In *Proceedings of the Australian Council for Educational Research (ACER) Research Conference, "Building teacher quality: What does the research tell us?"* (pp. 15–23). Melbourne, Australia: ACER.

Sanders, M., & Ní Chonaire, A. (2015). *"Powered to detect small effect sizes": You keep saying that. I do not think it means what you think it means (Working Paper No. 15/337).* Retrieved from University of Bristol website: http://www.bris.ac.uk/media-library/sites/cmpo/documents/WP15337_Web_Version.pdf.

Sartain, L., Stoelinga, S. R., & Brown, E. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation (Research Report).* Retrieved from Consortium on Chicago School Research website: https://consortium.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Report%20FINAL.pdf.

Stevenson, M., Stevenson, C., & Cooner, D. (2015). Improving teacher quality for Colorado Science teachers in high need schools. *Journal of Education and Practice, 6*(3), 42–50.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(9), 963–980.

Teacher Education Ministerial Advisory Group. (TEMAG). (2014). *Action now: Classroom ready teachers.* Retrieved from https://docs.education.gov.au/system/files/doc/other/action_now_classroom_ready_teachers_accessible.pdf.

Timperley, H., & Alton-Lee, A. (2008). Reframing teacher professional learning: An alternative policy approach to strengthening valued outcomes for diverse learners. *Review of Research in Education, 32*, 328–369. http://dx.doi.org/10.3102/0091732X07308968.

Tzivinikou, S. (2015). The impact of an in-service training program on the self-efficacy of special and general education teachers. *Problems of Education in the 21st Century, 64*, 95–107.

Ullah, T., & Jundran, S. (2014). Impact of district teacher educators' mentoring support on professional development of primary school teachers. *The Journal of Educational Research, 17*(1), 112–130.

Van Whittaker, J., Kinzie, M. B., Williford, A., & DeCoster, J. (2016). Effects of MyTeachingPartner-Math/Science on teacher-child interactions in prekindergarten classrooms. *Early Education and Development, 27*(1), 110–127.

Van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2014). Teacher scaffolding in small-group work: An intervention study. *Journal of the Learning Sciences, 23*(4), 600–650.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91.

Weldon, P. R. (2015, March). *The teacher workforce in Australia: Supply, demand and data issues (Policy Insights, Issue #2).* Retrieved from Australian Council for Educational Research website: http://research.acer.edu.au/policyinsights/2/.

Weldon, P., McKenzie, P., Kleinhenz, E., & Reid, K. (2013). *Teach for Australia pathway: Evaluation report, phase 3 of 3.* Retrieved from Australian Council for Educational Research website: http://research.acer.edu.au/cgi/viewcontent.cgi?article=1013&context=teacher_education.

Wilson, R., Dalton, B., & Baumann, C. (2015, March 16). *Six ways Australia's education system is failing our kids.* The Conversation. Retrieved from: https://theconversation.com/six-ways-australias-education-system-is-failing-our-kids-32958.

Zeichner, K. (2012). The turn once again toward practice-based teacher education. *Journal of Teacher Education, 63*(5), 376–382.