



Fresh evidence on the relationship between years of experience and teaching quality

Jennifer Gore¹ · Brooke Rosser¹ · Felicia Jaremus¹ · Andrew Miller¹ · Jess Harris¹

Received: 25 July 2022 / Accepted: 1 February 2023 / Published online: 3 March 2023
© The Author(s) 2023

Abstract

It is commonly assumed that experienced teachers are more proficient than beginners. However, evidence supporting this premise is complicated by diverging research traditions and mixed results. We explore the fundamental relationship between years of experience and teaching quality using a comprehensive pedagogical model. Our analysis of 990 lessons, taught by 512 primary teachers in New South Wales during 2014–15 and 2019–21, found no significant differences in pedagogy across the experience range (< 1–24+ years). We canvass two possible explanations: that initial teacher education (ITE) performs better than is typically assumed; and/or that experience, including ongoing participation in many forms of professional development (PD), has minimal impact on pedagogical quality. The important lesson from this study, however, is that the continual positioning of beginning teachers and ITE as deficient is unwarranted and, instead, we should focus on providing teachers with access to high-impact PD throughout their careers.

Keywords Teacher experience · Teaching quality · Classroom observation · Quality Teaching Model · Beginning teachers · Experienced teachers

✉ Brooke Rosser
brooke.rosser@newcastle.edu.au

Jennifer Gore
jenny.gore@newcastle.edu.au

Felicia Jaremus
felicia.jaremus@newcastle.edu.au

Andrew Miller
andrew.miller@newcastle.edu.au

Jess Harris
jess.harris@newcastle.edu.au

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

Introduction

Politicians and media commentators consistently bemoan the quality of teachers in the face of declining or stagnating performance on international assessments (Churchward & Willis, 2019; Dinham, 2015). In this context, beginning teachers and initial teacher education (ITE) have been subjected to an unrelenting procession of reviews and reform efforts (Tatto et al., 2018). Over the past 40 years Australia has seen more than 100 inquiries into ITE (Louden, 2008), with the latest commissioned by the Education Minister in 2021 (Paul et al., 2021; Tudge, 2021). As a result, ITE in Australia has undergone numerous reforms including greater prescription of teacher education course content, new teacher accreditation schemes, new minimum literacy and numeracy standards, and new ‘classroom readiness’ assessments for graduate teachers (Barnes & Cross, 2018; Rowe & Skourdombis, 2019; Simpson et al., 2021; Teacher Education Ministerial Advisory Group [TEMAG], 2014). Similar reforms have occurred around the globe, with countries such as England, France, Germany, Norway, Austria, and the United States instituting regulatory and policy changes to improve the quality of new teachers (Furlong, 2013; Mayer, 2021; Page, 2015; Simpson et al., 2021; Tatto et al., 2018).

In this article, we ask to what extent is the focus on ITE justified? While initiatives designed to improve the quality of graduate teachers have intensified, we are concerned about the absence of strong evidence documenting how, or indeed if, teaching quality varies by years of experience (Churchward & Willis, 2019; Graham et al., 2020; Mockler, 2018). The methodological challenges involved in measuring teaching quality mean that few robust large-scale studies have been conducted to provide such evidence (Hill et al., 2015). One of the only such studies conducted in Australia, involving classroom observations of 80 teachers, indicates that quality may not vary significantly with experience, finding no difference between teachers with 0–3 years’ experience and those with 5+ years’ experience (Graham et al., 2020).

It is often assumed, without robust evidence, that declining student outcomes stem from declining teacher quality and, further, that to improve student achievement, nations must, by necessity, raise the quality of new teachers (Churchward & Willis, 2019; Mockler, 2018; Tatto et al., 2018). Such assumptions imply significant problems with those enrolled in teaching degrees, with recent graduates, and/or with the ITE programs in which they participate. If new teachers (or their preparation) are to blame for stagnating student achievement, one might expect beginning teachers to deliver ‘poorer’ quality lessons than their more experienced colleagues. It is this concern about how quality of teaching changes with experience that our research interrogates.

We have provided fresh evidence on this question by analysing the quality of 990 lessons from a sample of 512 Australian teachers, ranging from those in their first year to those with more than 24 years’ experience. We used a comprehensive model of pedagogy called the Quality Teaching (QT) Model to address our research question: What is the relationship between teachers’ years

of experience and the quality of their teaching? Following, we begin by reviewing three distinct research traditions that contribute insights on the relationship between teachers' experience and teaching quality. Next, we describe our research methods and results and, finally, canvass different explanations for our key and somewhat surprising finding that beginning teachers deliver instruction that is of commensurate quality to that of their experienced colleagues.

Background to the study

While large-scale studies of classroom practice using standardised instruments for assessing quality have been a relatively recent addition to the literature (Graham et al., 2020; Hill et al., 2015), decades of research from diverging research traditions provide insights into the value and effects of teaching experience. We will discuss three categories. The first (which we refer to as Category 1), largely based in the United States (US), tests associations between teacher characteristics (including years of experience) and student achievement on standardised tests (e.g., Harris & Sass, 2011; Kini & Podolsky, 2016; Ladd & Sorensen, 2017; Papay & Kraft, 2015; Rockoff, 2004). Here, experience is defined as years spent teaching in classrooms post-graduation. As Graham et al. (2020) observe, these studies do not directly measure teaching practice and tend to use narrow measures of student outcomes (i.e., standardised test results in a few subjects).

Category 2 studies focus on differences in the cognition, behaviour, and performance of expert and novice teachers (e.g., Borko & Livingston, 1989; Gudmundsdottir & Shulman, 1987; Hattie & Yates, 2014; Leinhardt, 1989). Research in this category is US-centric but includes studies from European and Asian education systems. Often conducted in controlled 'laboratory' settings (Hattie & Yates, 2014; Tsui, 2005, 2009), Category 2 studies decontextualise teachers' work by assessing how they perform on specific tasks and make comparisons that are not necessarily about teachers' years of experience.

Category 3 studies (with which our own work is most closely aligned) measure differences in teaching quality using direct observations of classroom practice. These studies, largely based in the US, use a variety of pedagogical frameworks, such as the Classroom Assessment Scoring System (CLASS) (Pianta et al., 2008), Danielson's (2007) Framework for Teaching (FfT), and subject-specific frameworks such as the Mathematical Quality of Instruction (MQI) instrument (Hill, 2005) and the Protocol for Language Arts Teaching Observation (PLATO) (Grossman et al., 2014). While such studies typically do not make years of experience their key focus, the researchers often include experience categories in their statistical models. Our study extends this third group by focussing specifically on the role of teacher experience and contributing much needed insight into teaching quality beyond US contexts. Key findings generated by these three research agendas are outlined below, in turn.

Category 1. Studies of student achievement as proxy for quality

Studies investigating the relationship between teacher experience and student achievement have generated mixed insights (Graham et al., 2020). When teacher characteristics such as years of experience were first put into models predicting student achievement scores, they were often shown to be weak (or non-significant) predictors (Hanushek & Rivkin, 2006, 2012; Nye et al., 2004; Rivkin et al., 2005). Kini and Podolsky (2016) argue that recent studies showing a stronger association between teacher experience and student achievement are able to do so because of increased availability of data to match students with individual teachers and more advanced research methods. The association has typically been found in the first 3 to 5 years of teaching, with a sizeable number of studies now reporting ‘rapid’ gains in effectiveness during teachers’ first few years on the job (Araujo et al., 2016; Harris & Sass, 2011; Henry et al., 2012; Kini & Podolsky, 2016; Ladd & Sorensen, 2017; Papay & Kraft, 2015; Rice, 2010, 2013; Rockoff, 2004). However, not all studies have found such effects (Hill et al., 2015).

The picture is consistently less clear after the first 3 to 5 years. While some studies show small but significant improvement in teachers’ effectiveness well into their careers (Harris & Sass, 2011; Kraft & Papay, 2014; Ladd & Sorensen, 2017; Papay & Kraft, 2015), others indicate the ‘value’ added to student achievement scores plateaus or even declines after 3 to 5 years (Hanushek & Rivkin, 2006, 2012; Henry et al., 2012; Rice, 2010, 2013; Rockoff, 2004). Importantly, these findings vary by schooling context. For example, Kraft and Papay (2014) demonstrated that after 5 and 10 years of experience, teachers in the most supportive schools outperform their counterparts in the least supportive schools by 20% and 38%, respectively. In addition, the type of experience matters. Huang and Moon (2009) found total years of experience was not a significant predictor of student achievement, but years of experience teaching a particular grade level was. Despite conflicting evidence, the prevailing view is “for most teachers, experience increases effectiveness” (Kini & Podolsky, 2016, p. 1).

The validity of these studies, however, has been challenged. First, the results on standardised tests themselves can be distorted by factors such as content type, student socioeconomic status, and gender (Leder & Forgasz, 2018), casting doubt that student test results are a reliable or valid measure of teacher effectiveness. Second, the value-added models (VAMs)¹ on which studies of teacher effectiveness tend to rely have been critiqued as incomplete, volatile, and inconsistent (Amrein-Beardsley & Close, 2019; Darling-Hammond et al., 2012; Hallinger et al., 2014; Reynolds et al., 2014). While VAMs have become increasingly sophisticated and now include ‘controls’ for a range of factors, Rockoff and Speroni (2010) argue results are still “biased if some teachers are persistently given students that are difficult to teach” (p.

¹ Value-added models (VAMs) are statistical models that use a variety of measures to predict student performance on standardised tests (e.g., past test results and demographic information). A teacher’s ‘value added’ is the difference between the statistical model’s predictions and their students’ actual test performance (see Opper, 2019).

261) while others have greater choice over which schools they teach in (Hanushek & Rivkin, 2006). Notably, a measure of teaching or pedagogy is rarely included in VAMs, so what experienced teachers actually 'do' to achieve higher outcomes, when such a relationship is found, remains a mystery (Hill et al., 2015; Ingvarson & Rowe, 2008).

Category 2. Studies of differences between expert and novice teachers

Research focussed on differences in cognition, behaviour, and functioning between expert and novice teachers overwhelmingly documents the superiority of expert teachers (Hattie & Yates, 2014; Tsui, 2009). Novice teachers have been found to struggle to effectively plan and deliver coherent lessons and teaching units and to select developmentally appropriate content and teaching strategies (e.g., Borko & Livingston, 1989; Gudmundsdottir & Shulman, 1987; Leinhardt, 1989; Westerman, 1991). During lessons, novice teachers often fail to activate students' prior knowledge, to improvise when things go awry, and to notice and interpret classroom patterns (e.g., Berliner, 1988; Borko & Livingston, 1989; Hattie & Yates, 2014; Leinhardt, 1989; Westerman, 1991). The difficulties novice teachers face in reflecting on teaching and interpreting student behaviour (Kim & Klassen, 2018) can also lead to greater attention on student discipline than student learning and thinking (Huang & Li, 2012; McIntyre et al., 2017; Wolff et al., 2017), compared to expert teachers.

Despite this documented list of novice inadequacies, we find it risky to generalise findings from particular expertise studies to populations beyond the research context for at least three reasons. First, there is no consensus on how to define an expert teacher (Berliner, 2001; Tsui, 2009), substantially influencing results across the literature base. Second, while the term 'novice' pertains to those with little practical experience in a particular domain, an 'expert' is not simply someone who has accumulated more years of experience (Berliner, 2001; Hattie & Yates, 2014; Johnson, 2005). Researchers select 'expert' teachers based on a number of other characteristics, such as recommendations from school leaders, the attainment of state- and national-level teaching awards, and student achievement scores (Tsui, 2009). Third, expertise research is often cross-sectional and carried out in controlled settings away from teachers' classrooms in response to controlled stimuli, such as classroom vignettes or video excerpts (Hattie & Yates, 2014; Tsui, 2005, 2009).

However, it is well accepted that the contexts of the school and classroom, as well as the resources, goals, and orientations of teachers, are important contributors to teacher expertise (Berliner, 2001; Schoenfeld, 2011), as are opportunities to engage in 'deliberate' goal-directed practice with feedback, support, and encouragement from peers and knowledgeable others (Berliner, 2001; Hattie & Yates, 2014). As such, it is difficult to make valid assessments of teacher expertise through the deployment of controlled stimuli alone.

In response to these concerns, a small but growing number of recent studies have reconceptualised teacher expertise as a process of development, situated in schools and classrooms (Tsui, 2005, 2009). While these 'more naturalistic' studies, which take the form of in-depth case studies and longitudinal analyses, offer far greater

ecological validity than those conducted in ‘laboratory’ settings, the insights they generate still cannot be used to make generalisations about the quality of teaching delivered by broader populations of beginning and experienced teachers.

Category 3. Studies using observational frameworks to assess teaching quality

The use of observation frameworks to study pedagogy is an emerging field of research (Hill et al., 2015) in which the relationship between teacher experience and teaching quality has rarely been the focus (Graham et al., 2020). We identified 11 studies that used direct observational measures of teaching quality in primary/elementary, middle, or high school classrooms. When teacher experience has been addressed it has often been: (1) investigated as part of a host of other teacher background characteristics (e.g., Bryant et al., 1991; Guo et al., 2012; Hill et al., 2015; Mihaly & McCaffrey, 2015; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 2002, 2005; Stuhlman & Pianta, 2009); (2) represented by a few blunt categories such as greater or fewer than 5 years (e.g., Cortina et al., 2015; Graham et al., 2020; Hill et al., 2015; Mihaly & McCaffrey, 2015); or (3) largely overlooked with limited or no discussion of the results relating to teacher experience (e.g., Bryant et al., 1991; Gitomer et al., 2014; Mihaly & McCaffrey, 2015; Pianta et al., 2002).

Nevertheless, these limited investigations typically report few significant pedagogical differences between teachers of different experience levels across grades using a variety of observation tools including CLASS (e.g., Cortina et al., 2015; Gitomer et al., 2014; Graham et al., 2020), MQI (e.g., Hill et al., 2015), FfT, and PLATO (e.g., Mihaly & McCaffrey, 2015). Indeed, akin to some Category 1 studies, teacher characteristics including experience have been found to explain very little (if any) of the variability in teaching quality (see Gitomer et al., 2014; Hill et al., 2015; Mihaly & McCaffrey, 2015; NICHD ECCRN 2002, 2005). Arguably, this lack of difference is surprising given current governmental anxieties about and efforts to improve the quality of new teachers and ITE both in Australia and overseas.

When studies have found significant pedagogical differences by experience, the differences have been isolated to specific aspects of instruction and/or are difficult to interpret. For example, Guo et al. (2012) found a small but significant *negative* relationship between years of experience and the amount of time spent on ‘academic’ activities, while earlier research in first and third grade classrooms found the opposite trend (NICHD ECCRN, 2002, 2005). Graham et al. (2020) reported that ‘transitioning’ teachers with 4–5 years’ experience had worse scores for the Negative Climate and Instructional Learning Formats dimensions of CLASS, while Hill et al. (2015) found teachers with more than 2 years’ experience outperformed novices on the Classroom Organisation domain.

Similarly, using data from the Measures of Effective Teaching (MET) study, the largest known study of classroom practice to date (Bill & Melinda Gates Foundation, 2012), Mihaly and McCaffrey (2015) found no *systematic* differences in classroom observation scores by teachers’ years of experience across subjects (English and Math),

grades (4–8), or observation frameworks (CLASS, FfT, and PLATO). However, puzzlingly, English language teachers with 3 years' experience scored significantly higher on observations using CLASS and FfT than those with 4 or more years' experience, but not when the subject- or domain-specific instrument PLATO was used. Conversely, there were no significant differences on CLASS and FfT scores across experience categories for Math teachers.

Despite such findings, it would be premature to infer that experience is irrelevant. These studies tend to classify all teachers with at least 5 years in the profession as 'experienced,' thereby obscuring possible differences across the career span. Furthermore, most studies examine teaching quality in US classrooms using pedagogical models developed for that context, such as CLASS (e.g., Cortina et al., 2015; Gitomer et al., 2014; Graham et al., 2020; Hill et al., 2015; Mihaly & McCaffrey, 2015) and its predecessors (e.g., Guo et al., 2012; Pianta et al., 2002; NICHD ECCRN, 2002, 2005; Stulhman & Pianta, 2009). As a result, it is unclear whether findings would be consistent across teaching populations using alternative frameworks.

In sum, our review demonstrates that diverging research traditions, each with their own strengths and limitations, provide complicated answers to the question of whether concern about the quality of new teachers and ITE is warranted. Each tradition shines a different light on the question. Studies on teacher experience and student achievement (Category 1) show that student test scores improve in the first few years of a teacher's career, with mixed findings after this point. However, they provide limited insight into how the actions of teachers might drive these trends. Studies of expert and novice teachers (Category 2) document the superiority of experts, but acknowledge that experience and expertise are not synonymous. Observational studies of pedagogy (Category 3), which have primarily been conducted in US classrooms, demonstrate few differences overall between beginning and experienced teachers yet tend to rely on blunt comparisons among a few experience categories.

In this paper, we favour the third *in situ* approach because it enables investigation of differences in pedagogy by years of experience, but we seek to address two gaps. First, we extend our vista to teaching across the career span, from beginning teachers to those with more than 24 years' experience, thus conducting a more fine-grained analysis of how experience matters beyond the 2-, 3- or 5-year marks commonly used in Category 3 studies. Second, we contribute much needed evidence of teaching quality beyond the US using a comprehensive pedagogical model developed for use in Australian schools, known as the Quality Teaching (QT) Model. In previous work, the QT Model has been used to examine: the relationship between teaching quality and school socioeconomic status (Gore et al., 2022); improvement in teaching quality following participation in Quality Teaching Rounds (QTR) professional development (PD) (Gore et al., 2017); and, associated improvement in student achievement (Gore et al., 2021). By employing the QT Model and examining teaching in Australia across the career span, we seek to provide fresh insights into the relationship between teaching quality and experience.

Methods

To address the research question—What is the relationship between teachers' years of experience and the quality of their teaching?—we drew on classroom observational data derived from two randomised controlled trials (RCTs) conducted in NSW government schools during 2014–15 and 2019–21. Both trials were designed to assess the efficacy of QTR—an approach to teacher PD that involves teachers working in professional learning communities to observe and analyse each other's lessons using the QT Model.² Participating teachers reported their years of experience in surveys and had lessons observed by the research team before they were randomly allocated to treatment groups. Although the RCTs were designed for a different purpose, the baseline (i.e., pre-intervention) data have enabled us to explore associations between teaching quality and years of experience for a relatively large sample. The trials received university and education department ethics approvals before recruitment commenced. Below, we include only those details pertinent to our research question. Readers interested in details of the trials might like to access earlier publications of the study protocols and outcomes (see Gore et al., 2015, 2017, 2021; Miller et al., 2019).

The Quality Teaching Model

The QT Model is a comprehensive model of pedagogy derived from an extensive research synthesis of classroom factors that positively impact student learning (Ladwig & King, 2003). Applicable to any developmental stage or curriculum area, the QT Model has its roots in research undertaken on Authentic Pedagogy (Newmann, 1996) and Productive Pedagogy (Lingard et al., 2001). For almost two decades, the model has been endorsed by the NSW Department of Education as a model of teaching quality for government schools (NSW Department of Education and Training, 2006; Quality Teaching Academy, 2020), signalling its enduring resonance with teachers and school leaders.

The QT Model has three dimensions, each consisting of six elements (18 elements in total) that focus teachers' attention on principles underpinning the quality of teaching as manifest in classroom practice: (1) Intellectual Quality, (2) Quality Learning Environment, and (3) Significance (see Table 1). Each element is accompanied by a 1-to-5 coding scale and associated descriptors that distinguish quality at a high level of specificity (see online Appendix for an elaboration of one of the elements). Together, the 18 elements comprise a holistic model of pedagogy addressing lesson content and the intellectual demands placed on students, the environment within which learning occurs, and the relevance of lesson content to students' lives beyond the classroom.

² Note that our research adopts a post-positivist stance. We are of the view that it is possible to produce meaningful measurements of pedagogical quality using robust observation frameworks and rigorous research designs (e.g., RCT designs). At the same time, however, we recognise that such measurements are necessarily value-laden and partial.

Table 1 The Quality Teaching Model

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

The dimensions and elements of the NSW Quality Teaching Model of pedagogy. Reprinted from *Quality Teaching in NSW Public Schools: A Classroom Practice Guide*, 2nd edition (p. 10), by NSW Department of Education and Training, 2006, Ryde, Australia: Author. Copyright © 2006 by the State of NSW, Department of Education and Training, Professional Learning and Leadership Development Directorate. Reprinted with permission

While there is no consensus on what should be included in a pedagogical framework (Coe et al., 2014; Martinez et al., 2016), research using the QT Model has demonstrated several broad positive impacts on the profession (see Gore & Bowe, 2015; Gore et al., 2017, 2021; Gore & Rickards, 2021; Gore & Rosser, 2022), lending validity to its use as a tool for teachers and making its use in this study an important contribution to the literature. For example, the aforementioned RCTs demonstrated that the QT Model, when used as the core component of QTR PD, is associated with improved student achievement in mathematics, improved teaching quality, improved teacher morale, and improved teacher perceptions of appraisal and recognition (Gore et al., 2017, 2021).

Data sources

To address the relationship between experience and quality, we drew on pre-intervention data only, given that post-intervention data (after teachers participate in QTR) would confound the results. A total of 990 baseline observations of whole lessons taught by 512 teachers in 260 schools were conducted. The schools were representative of schools in Australia, with an average Index of Community Socio-Educational Advantage³ (ICSEA) of 1005 (standard deviation=83), consistent with the national mean of 1000 and standard deviation of 100. All observations used in this analysis were of primary school lessons, the majority (79%) of which were conducted in Grade 3 or 4 classrooms (age 8–10 years). The observed lessons were mostly in the key learning areas (KLAs) of English (52.8%) and Mathematics (28.4%), with a range of other KLAs such as Human Society and its Environment (HSIE) (7%), Physical Development, Health and Physical Education (PDHPE) (3.9%), Creative Arts (3.7%), and Science (3.5%) represented. All teachers

³ ICSEA is a standardised measure of school-level advantage in Australia. It includes parental education and occupation, school location, and proportion of Indigenous students. The mean ICSEA score is 1000 and the standard deviation is 100.

participating in the research were employed on at least a 12-month contract, given that the RCTs sought to measure change over the course of a school year for teachers who had their own classes and were not in casual employment.

Years of experience

In an online questionnaire, teachers reported demographic information, including their years of teaching experience, using the following categories: less than 1 year, 1–3 years, 4–6 years, 7–9 years, 10–12 years, 13–15 years, 16–18 years, 19–21 years, 22–24 years, and more than 24 years. While much of the literature uses the starting category of 0–3 years (e.g., Graham et al., 2020), we were keen to contribute fresh insights about the first year of teaching and so used <1 and 1–3 years' experience for our main analyses. Given concerns about the quality of ITE and readiness of beginning teachers, it is useful to make this distinction. The sample included teachers across the entire range. While the majority of lesson observations were taught by teachers with between 1 and 15 years' experience, at least 34 observations occurred in every experience category (Table 2).

Sample

Where possible each teacher was observed twice during data collection, which was the case for 92% of the teachers. To scrutinise the sample for potential bias, we investigated if there were any systematic patterns or differences in QT scores for those with only one observation (Table 2). The proportion of single observations by experience category ranged from 0 to 16% and, when expressed as an effect size (Cohen's *d*), the mean differences in QT scores between teachers with one or two observations ranged from -0.89 to 0.49 . This suggests that there is no systematic variation across the experience categories for those with only one observation. Similarly, the distribution of teachers in each experience category by ICSEA indicates no clear pattern. Using cut points representing half of one standard deviation away from the national mean ICSEA value of 1000 (Table 2), we found nothing to suggest our results by years of experience would be biased due to over-representation from any specific experience category or ICSEA category.

Quality of teaching measure

Dimension level scores and an overall QT score were obtained by researchers observing and coding whole lessons using the QT Model. Dimension level scores were calculated using the mean of the six elements for each of the Intellectual Quality, Quality Learning Environment, and Significance dimensions (range 1–5). The total QT score was calculated using the mean of the 18 elements (range 1–5).

In total, 64 raters were involved in data collection across the two RCTs. They all received two days' intensive training and subsequently completed independent scoring against pre-rated (20-min) lesson extracts. To ensure reliable scoring, no rater was sent into the field for data collection unless they achieved above 90% exact scoring. To further investigate inter-rater reliability among the large pool of raters,

Table 2 Number of observations by teachers' years of experience

Years of teaching	Teachers, <i>n</i>	Observations <i>n</i> (% of total)	Teachers with a single observation		ICSEA	
			<i>n</i> (% of category)	Effect size ^a , <i>d</i>	<950	
					<i>n</i> (% of total)	<i>n</i> (% of total)
<1	18	36 (4)	0 (0)	–	8 (3)	1050+ <i>n</i> (% of total)
1–3	106	195 (20)	17 (16)	0.00	44 (17)	10 (2)
4–6	118	230 (23)	8 (7)	0.49	62 (25)	79 (18)
7–9	71	138 (14)	5 (7)	0.33	38 (15)	90 (21)
10–12	44	88 (9)	0 (0)	–	28 (11)	68 (16)
13–15	34	66 (7)	3 (9)	–0.89	8 (3)	38 (9)
16–18	30	57 (6)	3 (10)	–0.49	8 (3)	44 (10)
19–21	35	69 (7)	2 (6)	–0.41	25 (10)	29 (7)
22–24	17	34 (3)	0 (0)	0.27	6 (2)	34 (8)
>24	39	77 (8)	1 (3)	0.01	25 (10)	14 (3)
Total	512	990 (100)	39 (8)	–	252 (100)	28 (6)
						434 (100)
						284 (100)

^a 1 vs 2 lessons; – denotes data not available

317 lessons were double-coded (~32% of total observations) and the intraclass correlation coefficient ($ICC_{(1)—one-way\ random\ effects}$) of the QT score was calculated. The ICC for a single measure (single-rater score used for analysis) was 0.848 (95% CI 0.814–0.876), indicating good reliability at the lesson level.

The two observations of the same teacher at each time point (which account for 960 of the 990 observations) were investigated for consistency at the teacher level, using an intraclass correlation coefficient ($ICC_{(3)—two-way\ mixed\ effects}$). The ICC (average measures) for the two observations displayed moderate reliability at 0.603 (95% CI 0.524–0.669), indicating some variability between the two lessons at the teacher level. However, the raw change in mean QT score (overall) between repeated observations was -0.009 (95% CI -0.069 – 0.050), equating to a negligible difference of 0.33%. This indicates that, while there was some variability between the repeated measures, the magnitude of the variability renders it largely inconsequential.

Analysis

Data were analysed using IBM PASW Statistics 27 (SPSS Inc. Chicago, IL) software with alpha levels set at $p < 0.05$. Linear mixed models were fitted, treating years of experience as the explanatory variable (categorical) and teaching quality as the dependent variable (continuous). Teaching quality was calculated at the dimension level using the mean of the six elements in the dimension. The mean of all 18 elements was used as a total measure of teaching quality given that the elements combine to form a holistic model of pedagogy. To account for the hierarchical nature of the data (teachers within schools and multiple lessons per teacher), random intercepts for school, and teacher within school, were included in the model. The school ICSEA value was also included as a covariate. To ensure the correct p -value when comparing the different experience categories, pairwise comparisons (Sidak contrasts) were used to assess differences between categories in relation to the reference category of less than 1 year. Effect sizes were calculated using Cohen's $d = (\text{reference group mean} - \text{comparison group mean}) / \text{pooled standard deviation (reference and comparison groups)}$. Ninety-five per cent confidence intervals (95% CIs) of the effect size were computed using the `compute.es` function (Del Re, 2013) in R version 3.4.4 (R Core Team, 2022). This function computes the confidence intervals using the variance in d derived by the Hedges and Olkin (1985) formula. For comparison to previous research in the field, we also conducted a second analysis using a combined 0–3 years category as the reference category in pairwise comparisons.

Results

Figure 1 illustrates the mean QT scores for each experience category, with lesson scores, group means, and 95% confidence intervals depicted. Table 3 provides the means, standard deviations, and 95% confidence intervals for the overall QT score, as well as for each QT dimension, by experience category. The overlapping

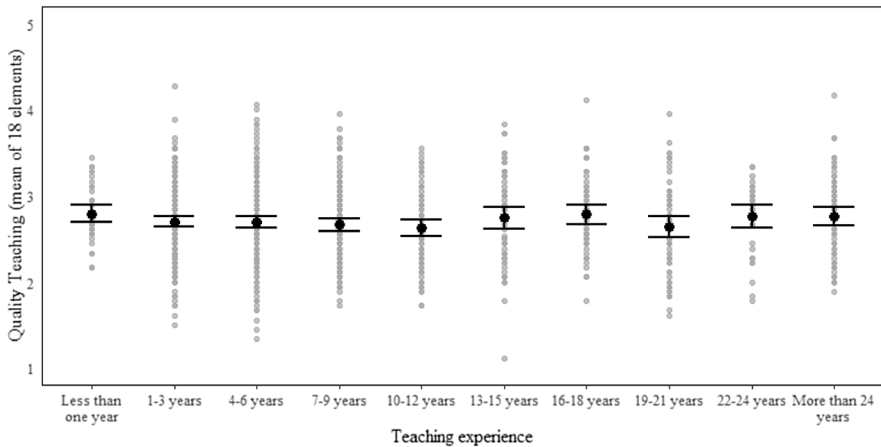


Fig. 1 Quality Teaching (total) by experience category—all observations

confidence intervals across the experience categories, visible in Fig. 1 and outlined in Table 3, highlight the similarity in the average dimension and QT scores.

The test of fixed effects for QT scores formally demonstrated no significant differences between experience categories at the $p < 0.05$ level ($F(9, 486) = 0.569$, $p = 0.823$). There were also no significant differences in the Intellectual Quality, Quality Learning Environment, and Significance dimensions by experience category. The pairwise comparisons also demonstrated no significant differences between the reference category (<1 year) and all other categories (Table 3) at the $p < 0.05$ level. In short, years of teaching experience did not explain a significant proportion of the variance in the quality of teaching.

When analysed using 0–3 years as a combined reference category to mimic comparisons used in previous literature, we also found no significant difference ($F(8, 489) = 0.535$, $p = 0.830$) for the QT score. Likewise, there were no significant differences identified in the dimension level analysis.

Discussion

This study investigated how pedagogical quality, as measured by the QT Model, varies by years of teaching experience in classrooms in NSW, Australia. We sought to test this relationship with a robust and comprehensive model of pedagogy and with a wide range of teaching experience levels. Our analysis of nearly 1000 lessons found no significant differences between experience categories across the range, from teachers in their first year to those teaching for more than 24 years. No significant differences were found for overall QT score or among the dimensions of Intellectual Quality, Quality Learning Environment, and Significance by experience category. This somewhat counterintuitive finding makes an important empirical contribution, calling into question relentless critiques of the adequacy of beginning teachers.

Table 3 Pairwise comparisons

Years of teaching	Intellectual Quality		Quality Learning Environment		Significance		Quality Teaching	
	Mean (SD)	Effect size <i>d</i> (95% CI)	Mean (SD)	Effect size <i>d</i> (95% CI)	Mean (SD)	Effect size <i>d</i> (95% CI)	Mean (SD)	Effect size <i>d</i> (95% CI)
<1	2.88 (0.59)	Reference	3.00 (0.40)	Reference	2.48 (0.48)	Reference	2.79 (0.34)	Reference
1–3	2.76 (0.69)	-0.17 (-0.75–0.40)	3.00 (0.56)	-0.01 (-0.60–0.58)	2.33 (0.51)	-0.26 (-0.79–0.26)	2.70 (0.46)	-0.19 (-0.76–0.39)
4–6	2.75 (0.70)	-0.13 (-0.70–0.44)	2.98 (0.57)	0.02 (-0.57–0.60)	2.34 (0.55)	-0.22 (-0.74–0.29)	2.69 (0.49)	-0.14 (-0.71–0.43)
7–9	2.72 (0.67)	-0.13 (-0.73–0.47)	2.91 (0.58)	-0.07 (-0.69–0.54)	2.36 (0.51)	-0.17 (-0.71–0.37)	2.66 (0.47)	-0.16 (-0.76–0.44)
10–12	2.62 (0.63)	-0.24 (-0.88–0.40)	2.95 (0.56)	-0.02 (-0.67–0.63)	2.30 (0.53)	-0.27 (-0.85–0.30)	2.62 (0.46)	-0.24 (-0.87–0.40)
13–15	2.80 (0.71)	-0.04 (-0.70–0.62)	3.02 (0.66)	0.04 (-0.64–0.72)	2.43 (0.51)	-0.07 (-0.67–0.53)	2.75 (0.52)	-0.04 (-0.70–0.62)
16–18	2.80 (0.66)	-0.12 (-0.81–0.56)	3.09 (0.55)	0.14 (-0.56–0.83)	2.44 (0.55)	-0.05 (-0.66–0.57)	2.78 (0.46)	-0.02 (-0.70–0.66)
19–21	2.65 (0.78)	-0.17 (-0.84–0.49)	2.95 (0.62)	0.01 (-0.67–0.69)	2.31 (0.56)	-0.19 (-0.79–0.41)	2.64 (0.54)	-0.16 (-0.82–0.51)
22–24	2.70 (0.60)	-0.22 (-0.98–0.54)	3.19 (0.50)	0.29 (-0.49–1.07)	2.39 (0.50)	-0.15 (-0.84–0.54)	2.76 (0.41)	-0.05 (-0.81–0.71)
>24	2.81 (0.65)	0.05 (-0.60–0.69)	2.89 (0.63)	-0.10 (-0.76–0.56)	2.58 (0.66)	0.25 (-0.33–0.83)	2.76 (0.48)	0.07 (-0.58–0.71)

SD = Standard deviation; CI = Confidence interval

Empirically, our study adds to the handful of international (Category 3) studies that use observation tools to examine teaching quality by years of experience. Overall, these studies also show small, non-significant differences between beginning and more experienced teachers. Significant differences found have been isolated to specific aspects of instruction and are inconsistent across samples.

Taken as a whole, evidence is accumulating that newly qualified teachers, on average, demonstrate a level of teaching quality commensurate to that of experienced teachers in a variety of contexts. In Australia, this pattern of evidence has now been found in two states, using two different measurement frameworks, the QT Model in NSW and CLASS in Queensland (for CLASS, see Graham et al., 2020), which suggests the result is not simply due to the sample or selection of observation tool. Our larger sample size, broader career span, and closer examination of the first year of teaching than most prior studies make important contributions to this body of literature.

Our finding that graduate teachers (on average) demonstrate pedagogical quality that is equivalent to their experienced colleagues is somewhat at odds with (Category 1) studies that report rapid gains in student achievement during the first few years of teaching (Kini & Podolsky, 2016; Ladd & Sorensen, 2017). It also is at odds with (Category 2) studies that document weaknesses in the cognition, behaviour, and functioning of novice teachers. Methodological differences in what is being measured and the extent to which context is considered might account for these different findings.

The fact that students are not randomly allocated to teachers, nor teachers randomly allocated to schools, must also be considered when explaining different findings between categories of studies. Early-career teachers are more often employed in hard-to-staff and disadvantaged schools (Luschei & Jeong, 2018; McKenzie et al., 2014; Rice, 2010, 2013). While we found no significant association between years of experience and school disadvantage (ICSEA) in our sample, this may be due to the small number of teachers who participated in each school. The complex relationships among school advantage, teaching experience, teaching quality, and student outcomes all warrant further investigation with larger samples of teachers from participating schools.

Nonetheless, we suggest two explanations that, if valid, have significant implications for education in Australia and beyond. First, ITE programs could be performing relatively well and better than is typically assumed in policy and the media (Mockler, 2022). That is, new graduates may enter the profession ‘classroom ready’ (TEMAG, 2014) and capable of demonstrating levels of pedagogical skill commensurate with their experienced colleagues. It is possible that improvements to ITE programs by higher education institutions, including lengthening the required days of professional experience and the use of standardised capstone assessments of teaching practice, have considerably advanced the teaching capacity of graduates compared to earlier cohorts. Such generational or ‘cohort effects’ are well-documented in the expertise (Category 2) literature whereby average performers today, across many skill domains, outperform the ‘experts’ of generations past (Hattie & Yates, 2014).

This is not to say that graduate teachers do not face difficulties. Indeed, the high rates of beginning teacher attrition (Dadvand & Dawborn-Gundlach, 2021) attest to the challenges of effective induction and adequate support. Nonetheless, our evidence suggests that those entering the teaching profession today might be better equipped to deliver higher quality *instruction* than their predecessors were at the outset of their careers. If this is the case, resources spent on the continuous procession of reviews and reforms in ITE might be more effective if directed elsewhere.

A second possible explanation is that on-the-job experience is insufficient to improve the quality of teaching over teachers' careers. Indeed, the three categories of literature we have reviewed converge in finding that experience alone does not guarantee continual improvement. However, teachers gain more than classroom experience over the course of their careers, including participation in countless hours of PD much of which is designed to enhance practice. In Australia, almost 80% of teachers have been teaching for more than 5 years (McKenzie et al., 2014) and 99% of teachers participate in various forms of in-service training each year (Organisation for Economic Co-operation and Development [OECD], 2019). Specifically, in NSW, full-time teachers must document a minimum of 100 h of PD every 5 years to maintain their accreditation (New South Wales Education Standards Authority, 2021).

Our finding of no difference by years of experience raises important questions about the role of PD in strengthening the quality of teaching, especially given that many of the most rigorously evaluated PD interventions produce little change in student outcomes (Sims et al., 2021). The lack of difference we found in the quality of teaching across career stages could suggest that years of participation in PD has not translated into improvements in the quality of teaching (as measured by the QT Model), at least not above the level demonstrated by new members of the teaching profession. At the same time, importantly, some studies show that it is possible to enhance teachers' classroom practice through PD (Garrett et al., 2019; Gore et al., 2017). To improve the quality of teaching across the career span, we need to ensure that all teachers participate in high-impact forms of PD with demonstrated positive effects on pedagogy.

While we have presented alternative explanations for our findings, both might have merit. Beginning teachers might be better prepared to enter the classroom than were previous cohorts, and all teachers might require better support to continue to improve their pedagogy throughout their careers. Of course, any explanation remains speculative without further investigation, including qualitative research to understand more deeply the relationship between years of experience and teaching quality. For now, we posit that: (1) new graduates can produce teaching of equivalent quality to their more experienced colleagues; and (2) years of experience do not ensure superior quality instruction. Therefore, we argue for policy that does not assume the inadequacy of beginning teachers or ITE and, instead, recognises the need for investment in high-impact forms of PD at all stages of teachers' careers.

To be clear, we do not wish to imply that participation in PD is not worthwhile. It is plausible that PD has many positive effects that are not directly observable in teachers' pedagogy as measured by the QT Model; for example, improving their content knowledge, morale, or self-efficacy—clearly worthy outcomes. Nor are we

suggesting that years of experience are irrelevant. Experienced teachers make valuable contributions to school improvement through such activities as leadership, mentoring, and coaching.

More broadly, it is important to remember teaching is a contextual endeavour, with the structural inequalities that pervade society likely to continue affecting teaching quality, irrespective of the quality of PD provided (Gore et al., 2022). What students bring to school (including family background or conditions of poverty) remains predictive of academic performance even when statistical models control for teacher effects (Hanushek, 2016; Hattie, 2003, 2009; Konstantopoulos, 2006; Konstantopoulos & Borman, 2011; OECD, 2005). However, the impact of societal factors remains obscured when current debates focus so heavily on teachers. As Graham et al. (2020) argue, the “narrow focus on ITE and the graduates it produces may mean that the true nature and breadth of the problems impacting school education remain undetected and unresolved, while others are magnified beyond their actual or practical significance” (p. 2).

Limitations

Several limitations of this research should be noted. In terms of study design, our results pertain only to primary school teachers (mostly Years 3 and 4) in the NSW government school sector. Additional research is needed beyond these parameters to assess whether the findings hold more broadly. Second, only a select number of teachers from each school participated in the RCTs. Given research ethics requirements, teachers were asked to opt-in and participant places at each school were limited. Hence, our results may reflect a more motivated sample of teachers unrepresentative of whole schools. It is also possible that only very confident beginning teachers opted-in or were asked to be involved—although this could be true of teachers across the entire range of experience levels. Third, given that our analysis is cross-sectional, no information is provided to demonstrate changes to pedagogical quality over time. Nor is it possible to determine how cohort effects influence the results.

There are innumerable methodological challenges associated with ‘measuring’ something as complicated as the quality of teaching. While the QT Model offers a holistic model of teaching quality, we acknowledge that the work of teaching is broader than the pedagogical quality of lessons measured by the model. There are complex skills, knowledges, and dispositions that go beyond what can be directly observed in classrooms, highlighting the importance of considering a wide range of measures when evaluating teachers and teaching.

Finally, we acknowledge that ratings of lessons require a degree of subjective judgement. However, this limitation is mitigated by the careful elaboration of descriptors for each point on the 1-to-5 rating scale (see online Appendix), extensive training of observers, and strong ICCs. While inter-rater reliability for the QT score is considered ‘good’ (ICC = 0.848), the ‘moderate’ reliability at the teacher level (ICC = 0.603) indicates variability between lessons taught by the same teacher (i.e., 60% of teaching quality, as measured by the QT Model, is

captured using two observations). Increasing the number of observations might reduce variability in QT score and increase the explanatory power of the statistical models presented. Without such data, appropriate caution should be used in generalising these results.

Conclusion

This study used a standardised pedagogical observation instrument, the Quality Teaching Model, to assess the quality of teaching produced by teachers across a broad range of experience levels. Despite continued government and media focus questioning the quality of new teachers and ITE, we found no evidence to indicate new teachers were inadequate, despite less on-the-job experience. Our findings suggest that ITE programs are producing graduates whose classroom practice is on par with those across the career span and that experience, including participation in formal and informal PD, does not necessarily produce higher quality pedagogy. In response, we urge policy makers to: (1) acknowledge the good work being done by beginning teachers and ITE programs; and (2) ensure that teachers have access to demonstrably high-impact PD over the entire course of their careers.

To be clear, we are not suggesting that beginning teachers do not face immense challenges upon entry to the profession. Nor do we wish to imply that experienced teachers are not valuable or that PD in general is not worthwhile. Rather, we argue that policy efforts to raise the quality of teaching must focus on the provision of PD with evidence of positive effects on pedagogy at *all* career stages. At the same time, we acknowledge that high-impact PD alone cannot compensate for the stratifying effects on student outcomes of increasing inequity, school resourcing disparities, and school socioeconomic segregation (Bonnor & Shepherd, 2017). Schools should be resourced properly and teachers supported well, especially in difficult contexts where delivering quality teaching is harder (Gore et al., 2022). Nonetheless, the fresh evidence we have provided—showing no relationship between quality and experience—raises important questions about assumptions held and provisions made to support ongoing improvement in teaching across all contexts and career stages.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13384-023-00612-0>.

Acknowledgements We wish to thank all teachers and principals who took part in this research, as well as all project staff. We are especially grateful to the project manager of both RCTs, Wendy Taggart.

Author contributions The authors confirm contributions to the paper as follows: Study conception and design: JG, AM, JH; Data analysis: AM, FJ, JG; Draft manuscript preparation: BR, FJ, JG. All authors reviewed the analysis, refined the manuscript and approved the final version.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This research was funded by the Paul Ramsay Foundation, the NSW Department of Education and the Australian Research Council (Grant No. DP180100285).

Data availability Not applicable.

Declarations

Competing interests The authors report no potential conflict of interest.

Ethical approval The studies were approved by the University of Newcastle's Human Research Ethics Committee (Approval Nos. H-2014-0123, H-2018-0340) and the NSW Department of Education State Education Research Applications Process (SERAP) (Approval Nos. 2014103, 2018458).

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

References

- Amrein-Beardsley, A., & Close, K. (2019). Teacher-level value-added models on trial: Empirical and pragmatic issues of concern across five court cases. *Educational Policy*, 26, 5–28. <https://doi.org/10.1177/0895904819843593>
- Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in Kindergarten. *The Quarterly Journal of Economics*, 131(3), 1415–1454. <https://doi.org/10.2307/26372667>
- Barnes, M., & Cross, R. (2018). 'Quality' at a cost: The politics of teacher education policy in Australia. *Critical Studies in Education*, 62(4), 455–470. <https://doi.org/10.1080/17508487.2018.1558410>
- Berliner, D. C. (1988). Implication of studies on expertise in pedagogy for teacher education and evaluation. In J. Pfliederer (Ed.), *New directions for teacher assessment. Proceedings of the 1988 ETS Invitational Congress* (pp. 39–68). Educational Testing Service.
- Berliner, D. C. (2001). Learning about learning from expert teachers. *International Journal of Educational Research*, 35, 463–482.
- Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Retrieved from <https://usprogram.gatesfoundation.org/news-and-insights/usp-resource-center/resources/gathering-feedback-on-teaching-combining-high-quality-observations-with-student-surveys-and-achievement-gains--report>
- Bonnor, C., & Shepherd, B. (2017). *Losing the game: State of our schools in 2017*. Retrieved from <https://cpd.org.au/wp-content/uploads/2017/06/FINAL-Losing-the-Game-June-21-with-dedication.pdf>
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26, 473–498. <https://doi.org/10.3102/00028312026004473>
- Bryant, D. M., Clifford, R. M., & Peisner, E. S. (1991). Best practices for beginners: Developmental appropriateness in Kindergarten. *American Educational Research Journal*, 28(4), 783–803.
- Churchward, P., & Willis, J. (2019). The pursuit of teacher quality: Identifying some of the multiple discourses of quality that impact the work of teacher educators. *Asia-Pacific Journal of Teacher Education*, 47(3), 251–264. <https://doi.org/10.1080/1359866X.2018.1555792>
- Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. Retrieved from <https://www.suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-REPORT.pdf>.
- Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking

- with expert and novice teachers. *International Journal of Science and Mathematics Education*, 13(2), 389–403. <https://doi.org/10.1007/s10763-014-9610-5>
- Dadvand, B., & Dawborn-Gundlach, M. (2021). *The challenge to retain second-career teachers*. Retrieved from <https://pursuit.unimelb.edu.au/articles/the-challenge-to-retain-second-career-teachers>
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Association for Supervision and Curriculum Development.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8–15. <https://doi.org/10.1177/003172171209300603>
- Del Re, A. C. (2013). *Compute.es: Compute effect sizes* [Computer Software]. Retrieved from <https://cran.rproject.org/package=compute.es>.
- Dinham, S. (2015). The worst of both worlds: How the US and UK are influencing education in Australia. *Education Policy Analysis Archives*, 23, 1–20. <https://doi.org/10.14507/epaa.v23.1865>
- Furlong, J. (2013). Globalisation, neoliberalism, and the reform of teacher education in England. *The Educational Forum*, 77(1), 28–50. <https://doi.org/10.1080/00131725.2013.739017>
- Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education*, 43(1), 106–137. <https://doi.org/10.3102/0091732X19830634>
- Gitomer, D., Bell, C., Qi, Y., McCaffrey, D., Hamre, B., & Pianta, R. C. (2014). The instructional challenge in improving teaching quality: Lessons from a classroom observation protocol. *Teachers College Record*, 116(6), 1–32.
- 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., Jaremus, F., & Miller, A. (2022). Do disadvantaged schools have poorer teachers? Rethinking assumptions about the relationship between teaching quality and school-level advantage. *The Australian Educational Researcher*, 49, 635–656. <https://doi.org/10.1007/s13384-021-00460-w>
- Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education*, 68, 99–113. <https://doi.org/10.1016/j.tate.2017.08.007>
- Gore, J. M., Miller, A., Fray, L., Harris, J., & Prieto, E. (2021). Improving student achievement through professional development: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education*, 101, 103297. <https://doi.org/10.1016/j.tate.2021.103297>
- Gore, J., & Rickards, B. (2021). Rejuvenating experienced teachers through Quality Teaching Rounds professional development. *Journal of Educational Change*, 22, 335–354. <https://doi.org/10.1007/s10833-020-09386-z>
- Gore, J., & Rosser, B. (2022). Beyond content-focused professional development: Powerful professional learning through genuine learning communities across grades and subjects. *Professional Development in Education*, 48(2), 218–232. <https://doi.org/10.1080/19415257.2020.1725904>
- Gore, J., 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: Rational and study protocol for a cluster randomised controlled trial. *International Journal of Educational Research*, 74, 82–95. <https://doi.org/10.1016/j.ijer.2015.08.002>
- Grossman, P., Cohen, J., & Brown, L. (2014). Understanding instructional quality in English Language Arts: Variations in the relationship between PLATO and value-added by content and context. In K. Kerr, R. Pianta, & T. Kane (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 303–331). Jossey-Bass.
- Gudmundsdottir, S., & Shulman, L. (1987). Pedagogical content knowledge in social studies. *Scandinavian Journal of Educational Research*, 31, 59–70. <https://doi.org/10.1080/0031383870310201>
- Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal*, 113, 3–24. <https://doi.org/10.1086/665816>
- Graham, L. J., White, S. L. J., Cologon, K., & Pianta, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education*, 96, 1–10. <https://doi.org/10.1016/j.tate.2020.103190>
- Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26(1), 5–28. <https://doi.org/10.1007/s11092-013-9179-5>

- Hanushek, E. A. (2016). What matters for student achievement: Updating Coleman on the influence of families and schools. *Education next*, 16(2), 18–26.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 1051–1078). Elsevier.
- Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, 4, 131–157. <https://doi.org/10.1146/annurev-economics-080511-111001>
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95, 798–812. <https://doi.org/10.1016/j.jpubeco.2010.11.009>
- Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* [Paper presentation]. Australian Council for Educational Research 2003 Research Conference, “Building teacher quality: What does the research tell us?,” Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hattie, J., & Yates, G. C. R. (2014). *Visible learning and the science of how we learn*. Routledge.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press. Retrieved from <https://idostatistics.com/hedges-olkin-1985-statistical-methods-for-metaanalysis/>
- Henry, G. T., Fortner, C. K., & Bastian, K. C. (2012). The effects of experience and attrition for novice high-school science and mathematics teachers. *Science*, 335(6072), 1118–1121. <https://doi.org/10.1126/science.1215343>
- Hill, H. (2005). Content across communities: Validating measures of elementary mathematics instruction. *Educational Policy*, 19(3), 447–475. <https://doi.org/10.1177/0895904805276142>
- Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching: Examining personal and institutional predictors of high-quality instruction. *AERA Open*, 1(4), 1–23. <https://doi.org/10.1177/23232858415617703>
- Huang, R., & Li, Y. (2012). What matters most: A comparison of expert and novice teachers’ noticing of mathematics classroom events. *School Science and Mathematics*, 112(7), 420–432. <https://doi.org/10.1111/j.1949-8594.2012.00161.x>
- Huang, F. L., & Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing school. *Educational Assessment, Evaluation and Accountability*, 21, 209–234. <https://doi.org/10.1007/s11092-009-9074-2>
- Ingvarson, L., & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education*, 52(1), 5–35.
- Johnson, K. (2005). The ‘general’ study of expertise. In K. Johnson (Ed.), *Expertise in second language learning and teaching* (pp. 11–33). Palgrave Macmillan.
- Kim, L. E., & Klassen, R. M. (2018). Teachers’ cognitive processing of complex school-based scenarios: Differences across experience levels. *Teaching and Teacher Education*, 73, 215–226. <https://doi.org/10.1016/j.tate.2018.04.006>
- Kini, T., & Podolsky, A. (2016). *Does teaching experience increase teacher effectiveness? A review of the research*. Learning Policy Institute.
- Konstantopoulos, S. (2006). Trends of school effects on student achievement. *Teachers College Record*, 108(12), 2550–2581.
- Konstantopoulos, S., & Borman, G. D. (2011). Family background and school effects on student achievement: A multilevel analysis of the Coleman data. *Teachers College Record*, 113(1), 97–132. <https://doi.org/10.1177/016146811111300101>
- Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476–500. <https://doi.org/10.3102/0162373713519496>
- Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, 12(2), 241–279. https://doi.org/10.1162/EDFPa_00194
- Ladwig, J. G., & King, M. B. (2003). *Quality teaching in NSW public schools: An annotated bibliography*. NSW Department of Education and Training Professional Support and Curriculum Directorate.
- Leder, G. C., & Forgasz, H. J. (2018). Measuring who counts: Gender and mathematics assessment. *ZDM Mathematics Education*, 50(4), 687–697. <https://doi.org/10.1007/s11858-018-0939-z>

- Leinhardt, G. (1989). Math lessons: A contrast of novice and expert competence. *Journal for Research in Mathematics Education*, 20, 52–75. <https://doi.org/10.5951/jresmetheduc.20.1.0052>
- Lingard, B., Ladwig, J., Mills, M., Bahr, M., Chant, D., Warry, M., Ailwood, J., Capeness, R., Christie, P., Gore, J., Hayes, D., & Luke, A. (2001). *Queensland school reform longitudinal study: Final report*. Education Queensland.
- Louden, W. (2008). 101 damnations: The persistence of criticism and the absence of evidence about teacher education in Australia. *Teachers and Teaching*, 14(4), 357–368. <https://doi.org/10.1080/13540600802037777>
- Luschei, T. F., & Jeong, D. W. (2018). Is teacher sorting a global phenomenon? Cross-national evidence on the nature and correlates of teacher quality opportunity gaps. *Educational Researcher*, 47(9), 556–576. <https://doi.org/10.3102/0013189X18794401>
- Martinez, F., Taut, S., & Schaaf, K. (2016). Classroom observation for evaluating and improving teaching: An international perspective. *Studies in Educational Evaluation*, 49, 15–29. <https://doi.org/10.1016/j.stueduc.2016.03.002>
- Mayer, D. (Ed.). (2021). *Teacher education policy and research: Global perspectives*. Springer. <https://doi.org/10.1007/978-981-16-3775-9>
- McIntyre, N. A., Mainhard, M. T., & Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze. *Learning and Instruction*, 49, 41–53. <https://doi.org/10.1016/j.learninstruc.2016.12.005>
- McKenzie, P., Weldon, P., Rowley, G., Murphy, M., & Mcmillan, J. (2014). *Staff in Australia's schools 2013: Main report on the survey*. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1021&context=tll_misc
- Mihaly, K., & McCaffrey, D. F. (2015). Grade-level variation in observational measures of teacher effectiveness. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 9–49). Wiley.
- Miller, A., Gore, J., Wallington, C., Harris, J., Prieto-Rodriguez, E., & Smith, M. (2019). Improving student outcomes through professional development: Protocol for a cluster randomised controlled trial of Quality Teaching Rounds. *International Journal of Educational Research*, 98, 146–158. <https://doi.org/10.1016/j.ijer.2019.09.002>
- Mockler, N. (2018). Early career teachers in Australia: A critical policy historiography. *Journal of Education Policy*, 33(2), 262–278. <https://doi.org/10.1080/02680939.2017.1332785>
- Mockler, N. (2022). *Constructing teacher identities: How the print media define and represent teachers and their work*. Bloomsbury Publishing.
- National Institute of Child Health and Human Development Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal*, 102(5), 367–387. <https://doi.org/10.1086/499709>
- National Institute of Child Health and Human Development Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behaviour. *The Elementary School Journal*, 105(3), 305–323. <https://doi.org/10.1086/428746>
- Newmann, F. M. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. Jossey Bass.
- New South Wales Education Standards Authority. (2021). *Maintenance of teacher accreditation policy*. Retrieved from <https://educationstandards.nsw.edu.au/wps/portal/nesa/teacher-accreditation/resources/policies-procedures/maintenance-of-teacher-accreditation-policy>
- New South Wales Department of Education and Training. (2006). *Quality teaching in NSW public schools: A classroom practice guide* (2nd ed.). Retrieved from https://app.education.nsw.gov.au/quality-teaching-rounds/Assets/Classroom_Practice_Guide_ogogVUqQeB.pdf
- Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257.
- Opper, I. M. (2019). *Value-added modeling 101: Using student test scores to help measure teaching effectiveness*. RAND Corporation.
- Organisation for Economic Co-operation and Development. (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. Retrieved from <https://www.oecd.org/education/school/attractingdevelopingandretainingeffectiveteachers-finalreportteachersmatter.htm>
- Organisation for Economic Co-operation and Development. (2019). *Country note - Australia: Results from TALIS 2018*. Retrieved from https://www.oecd.org/education/talis/TALIS2018_CN_AUS.pdf

- Page, T. M. (2015). Common pressures, same results? Recent reforms in professional standards and competences in teacher education for secondary teachers in England, France and Germany. *Journal of Education for Teaching*, 41(2), 180–202. <https://doi.org/10.1080/02607476.2015.1011900>
- Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105–119. <https://doi.org/10.1016/j.jpubeco.2015.02.008>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system™: Manual K-3*. Brookes Publishing.
- Pianta, R. C., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, 102(3), 225–238. <https://doi.org/10.1086/499701>
- Paul, L., Louden, W., Elliot, M., & Scott, D. (2021). *Next steps: Report of the Quality Initial Teacher Education Review*. Report prepared for the Australian Government. Retrieved from <https://www.dese.gov.au/quality-initial-teacher-education-review/resources/next-steps-report-quality-initial-teacher-education-review>
- Quality Teaching Academy. (2020). *Quality teaching: Classroom practice guide* (3rd ed.). NSW Department of Education.
- R Core Team. (2022). *R: A language and environment for statistical computing* (Version 3.4.4) [Computer Software]. R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (ERR): A state-of-the-art review. *School Effectiveness and School Improvement*, 25(2), 197–230. <https://doi.org/10.1080/09243453.2014.885450>
- Rice, J. K. (2010). *The Impact of teacher experience: Examining the evidence and policy implications*. National Centre for Analysis of Longitudinal Data in Education Research. Retrieved from <https://eric.ed.gov/?id=ED511988>
- Rice, J. K. (2013). Learning from experience? Evidence on the impact and distribution of teacher experience and implications for teacher policy. *Education Finance and Policy*, 8(3), 332–348. https://doi.org/10.1162/EDFP_a_00099
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. <https://doi.org/10.1257/0002828041302244>
- 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>
- Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review*, 100, 261–266. <https://doi.org/10.1257/aer.100.2.261>
- Rowe, E. E., & Skourdombis, A. (2019). Calling for ‘urgent national action to improve the quality of initial teacher education’: The reification of evidence and accountability in reform agendas. *Journal of Education Policy*, 34(1), 44–60. <https://doi.org/10.1080/02680939.2017.1410577>
- Schoenfeld, A. H. (2011). Reflections on teacher expertise. In Y. Li & G. Kaiser (Eds.), *Expertise in mathematics instruction* (pp. 327–341). Springer.
- Simpson, A., Cotton, W., & Gore, J. (2021). Teacher education/ors in Australia: Still shaping the profession despite policy intervention. In D. Mayer (Ed.), *Teacher education policy and research: Global perspectives* (pp. 11–26). Springer.
- Sims, S., Fletcher-Wood, H., O’Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J., & Anders, J. (2021). *What are the characteristics of teacher professional development that increase pupil achievement? A systematic review and meta-analysis*. Education Endowment Foundation. Retrieved from <https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/teacherprofessional-development-characteristics>
- Stuhlman, M. W., & Pianta, R. C. (2009). Profiles of educational quality in first grade. *The Elementary School Journal*, 109(4), 323–342. <https://doi.org/10.1086/593936>
- Tatto, M. T., Burn, K., Menter, I., Mutton, T., & Thompson, I. (2018). *Learning to teach in England and the United States: The evolution of policy and practice*. Routledge.
- Teacher Education Ministerial Advisory Group. (2014). *Action now, classroom ready teachers* [Report]. Retrieved from <https://www.dese.gov.au/teaching-and-school-leadership/resources/action-now-classroom-ready-teachers-report-0>
- Tsiu, A. B. M. (2005). Expertise in teaching: Perspectives and issues. In K. Johnson (Ed.), *Expertise in second language learning and teaching* (pp. 167–189). Palgrave Macmillan.

- Tsui, A. B. M. (2009). Teaching expertise: Approaches, perspectives and characterizations. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to second language teacher education* (pp. 190–197). Cambridge University Press.
- Tudge, A. (2021). *Initial teacher education review launched* [Media release]. Retrieved from <https://ministers.dese.gov.au/tudge/initial-teacher-education-review-launched>
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292–305. <https://doi.org/10.1177/002248719104200407>
- Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. A. (2017). Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, 66, 295–308. <https://doi.org/10.1016/j.tate.2017.04.015>
- Zeichner, K. (2006). Different conceptions of teacher expertise and teacher education in the USA. *Education Research and Perspectives*, 33(2), 60–79.

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

Jennifer Gore is a Laureate Professor and leader of the Teachers and Teaching Research Centre in the School of Education at the University of Newcastle. Professor Gore's research centres on enhancing schooling for teachers and students with joint focuses on quality and equity. Her development of, and research into, Quality Teaching and Quality Teaching Rounds has had significant impact for students, teachers, school leaders and systems across Australian jurisdictions and internationally.

Brooke Rosser is a research assistant and PhD candidate with the Teachers and Teaching Research Centre in the School of Education at the University of Newcastle. Brooke's research interests include teacher professional learning, teacher collaboration and professional communities. Her thesis examines the significance of teachers' social capital for improved outcomes.

Felicia Jaremus Dr Felicia Jaremus is a postdoctoral fellow with the Teachers and Teaching Research Centre in the School of Education at the University of Newcastle. She has extensive experience in mixed methods research. Her research interests include social justice, mathematics education, and qualitative research methods. Her thesis focused on senior secondary mathematics participation, with particular attention to gender.

Andrew Miller Dr Andrew Miller is a senior lecturer in the School of Education at the University of Newcastle and a chief investigator at the Teachers and Teaching Research Centre. Dr Miller's research focus is on the professional development of teachers for the improvement of teaching quality and student achievement in physical and academic settings.

Jess Harris Dr Jess Harris is an associate professor in the School of the Education at the University of Newcastle and a chief investigator at the Teachers and Teaching Research Centre. Dr Harris' research interests include the practices used by school leaders, teachers, students, and community members to improve schooling for all students both in Australia and internationally.