PART 2

The Value of Research on Teacher Education: Informing Policy

∴

CHAPTER 6

Drawing on Large-Scale Studies to Examine the Contribution of Teacher Quality to Student Learning

A Critical Reflection on the Literature

Charalambos Y. Charalambous

Abstract

Over the past three decades, significant research efforts have been invested in understanding how teacher quality – namely different teacher characteristics and qualifications – contribute to student learning, on the grounds that such empirical evidence could inform teacher education programs and initiatives. Responding to calls to focus on large-scale studies in teacher education, this chapter first examines what large-scale studies suggest about the contribution of teacher preparation and qualifications, teaching experience, teacher knowledge, and professional development to student learning. Critically reflecting on these findings, the chapter then discusses four lessons that can inform future studies: the importance of exploring more comprehensively the effect of teacher quality factors; the need for in-depth explorations of seemingly inconsistent results; the necessity of combining teach*er* quality with teach*ing* quality investigations; and the critical role that smaller-scale (qualitative) studies can play in better understanding the phenomena of interest. The chapter concludes by considering the implications of these four lessons for teacher education research and discussing how large-scale studies can help enhance its value.

Keywords

large-scale studies – professional development – teacher knowledge – teacher preparation and qualifications – teacher quality – teaching experience

1 Introduction

Although teacher education research comprises a relatively newly developed field (Grossman & McDonald, 2008), it has often been under attack for being underdeveloped, small scale, undertheorised, fragmentary, and somewhat parochial (Mayer, 2021; Mayer & Oancea, 2021). Critics of this field often lament that it has failed to reveal how teacher education can uplift teacher quality, which in turn, is assumed to lead to better teaching quality, and eventually to raise student learning (cf. Cochran-Smith, 2021; Tatto & Pippin, 2017). As accountability pressures increase, and as voices about the importance of "fixing" teacher education become more intense, calls have repeatedly been made to conduct large-scale studies and longitudinal studies that can track the impact of different teacher education programs on teacher quality (cf. Grossman & McDonald, 2008; Mayer & Oancea, 2021).

Unsurprisingly, during the last 15 years, some large-scale studies that aimed to examine the outcomes of teacher education on prospective and practicing teachers have been undertaken. For example, starting in 2008 and being conducted every five years, the *Teaching and Learning International Study* (TALIS) aims to investigate the perceptions of teachers worldwide about their preparation and job satisfaction (OECD, 2009, 2014, 2019). Similarly, focusing on mathematics, the *Teacher Education and Development Study in Mathematics* (TEDS-M, Tatto et al., 2012) explored the mathematics preparation of teachers at the primary and lower secondary levels in 17 participating countries. Another large-scale study conducted in Australia, *Studying the Effectiveness of Teacher Education* (cited in Mayer, 2021), examined the effectiveness of teacher education for early career teachers in Australia by following over 5000 graduates from teacher education programs into their early years of teaching. Taken together, these studies have provided important insights into teachers' perceptions of their preparation and their sense of preparedness. Their importance notwithstanding, these studies do not meet one of the key concerns echoed by accountability critiques: no actual links are made between the aspects studied and their contribution to student learning outcomes. Aiming to address this limitation, this chapter focuses on and reviews large-scale studies that examined the contribution of different aspects of teacher quality to student learning.

The remainder of this chapter is organised into three sections. The first section provides an overview of the results of such large-scale studies. Critically reflecting on these findings, in the next section we discuss four lessons learned that pertain to how the limitations of such large-scale studies can be turned into affordances in future research. In the last section, we consider the

implications of these four lessons for teacher education research and conclude by discussing how the value of teacher education research can be enhanced through large-scale studies.

2 Using Large-Scale Studies to Examine the Contribution of Teacher Quality to Student Learning

2.1 *Defining Terms*

Teacher quality captures "teacher qualifications and characteristics (inputs) that are assumed to influence teaching and student outcomes" (Nilsen et al., 2016, p. 5). Our focus on teach*er* as opposed to teach*ing* quality in this chapter hinges on that the first term directly corresponds to aspects of teacher education, which are the focus of this book, whereas the second term probes on teachers' practice, without necessarily exploring the inputs contributing to the quality of teachers' work (Nilsen et al., 2016).

Large-scale studies have been influential in examining how different teacher-quality attributes contribute to student learning. As Kyriakides and Charalambous (2014) point out, such studies have provided the means to start understanding how different factors – in our case relating to teacher characteristics and qualifications – support different types of student learning (cognitive, metacognitive, affective, and psychomotor). In addition, such studies become even more pivotal when conducted at an international level. As Nielsen and colleagues (2016) explain, by "using the world as a global educational laboratory" international large-scale studies "may contribute toward an international understanding of teacher quality [...] and establish their importance for student learning outcomes across and within countries and over time" (p. 2).

But what exactly are "large-scale" studies? Attempting to bring some consensus on this issue, Middleton and colleagues (2015) list four criteria (by drawing on a list of criteria originally proposed by Anderson and Postlethwaite, 2007): (a) sample size, (b) purpose of the research, (c) generalisability of results, and (d) type and complexity of data analysis. In terms of sample size, they propose that studies should include sample sizes of at least 2500 participants for national studies or at least 7500 participants for international studies. In terms of their purpose, large-scale studies should aspire to examine and understand systems and phenomena more broadly rather than focusing on very specific aspects of them. Generalisability of results captures the goal of producing findings that are transferable beyond the local context examined, whereas the type and complexity of data analysis stipulates that more advanced and complex methods are utilised to analyse the data collected.

Although other classifications might be possible, in this chapter we adopt this classification because there seems to be a great variety in what scholars call large-scale studies, especially in terms of their sample size. In fact, a first exploration of studies we collected for the purposes of this chapter which were identified by their authors as being "large-scale" revealed a notable variation in terms of their sample size (ranging from 1043 to 100,288 student participants). Therefore, to bring some consistency in the corpus of studies examined, it was necessary to adopt certain criteria which would, in consequence, exclude a significant number of studies. Nevertheless, given the intention of this chapter – not to provide a comprehensive or systematic review of large-scale studies on the characteristics and qualifications examined but rather to make a point about the affordances and the limitations of large-scale studies exploring teacher quality – we thought that the merits of adopting such a classification outweighed the limitations. Furthermore, although meta-analyses cannot be considered as large-scale studies, they were included in this chapter for two reasons. First, meta-analyses themselves often draw on large-scale studies. Hence, instead of going back to the original studies, it was preferred to utilise the meta-analyses themselves. And second, by summarising several studies (which amount to significantly larger sample sizes than those discussed above), meta-analyses better lend themselves to identifying patterns and deviations thereof, which were critical for the arguments advanced in this chapter.

2.2 *Sampling Teacher Characteristics and Qualifications*

In one of the earliest meta-analyses of studies published between 1960 and 1976, Begle (1979) examined the contribution of several teacher characteristics and qualifications on students' mathematics learning. Given the inconsistent results yielded from this meta-analysis, Begle urged researchers to channel their efforts to other directions. He lamented (pp. 54–55, emphasis added):

> We are no nearer any answers to questions about teacher effectiveness than our predecessors were some generations ago. What is worse, no promising lines of further research have been opened up. Evidently our attempts to improve mathematics education would not profit from further studies of teachers and their characteristics. *Our efforts should be pointed in other directions*.

Nonetheless, contrary to Begle's plea, scholars in the following years have continued examining a gamut of teacher characteristics and qualifications. From this wide array, for the purposes of this chapter, we focus on four such characteristics and qualifications that are related to teacher education: (a)

teacher preparation and qualifications, (b) teaching experience, (c) professional development, and (d) teacher knowledge. The first aspect directly captures indicators of teacher education; the second aspect can be considered as compensation for the limitations in initial teacher education; the third aspect captures teachers' ongoing education, and the last aspect can be thought as gauging a by-product of teachers' initial or ongoing education.

2.3 *Large-Scale Studies on the Effect of Selected Teacher Characteristics and Qualifications: A Brief Overview*

2.3.1 Teacher Preparation and Qualifications

From the indicators examined in this category, we focus on the following three: (a) college quality, (b) certification, and (c) coursework and degrees obtained.

Research findings seem to converge on the first two indicators. In particular, meta-analyses suggest a positive correlation between college ratings and student learning since students of teachers who attended more competitive colleges were found to perform better than students of teachers who attended less competitive colleges (Coenen et al., 2018; Wayne & Youngs, 2003). Subject-matter certification also appears to matter, particularly with respect to subject matters like mathematics and language arts; however, it has not been positively related to student learning for other subject matters (Coenen et al., 2018; Wayne & Youngs, 2003).

Study findings, however, seem to diverge when it comes to coursework and the degrees obtained. Whereas some studies (Çakır & Bichelmeyer, 2016; Canales & Maldonado, 2018) showed no effects on student learning, others (Goldhaber & Brewer, 1997; Harris & Sass, 2011; Toropova et al., 2019) reported positive effects, especially for middle-school mathematics. Harris and Sass (2011) also showed that the timing of acquiring an advanced degree appears to matter. Furthermore, mixed results have also been yielded from meta-analyses. For example, in Greenwald et al. (1996), 15% of the studies reported positive significant effects, 13% negative significant effects, and 72% no significant effects. No significant effects were also found in two other meta-analyses (Coenen et al., 2018; Wayne & Youngs, 2003) for degrees or coursework in different subject matters, with the exception of mathematics and, to some extent, science. Hence, it seems that subject-matter specificity appears to moderate the effect of coursework and degrees obtained on student learning.

2.3.2 Teaching Experience

Research findings appear to be equally inconclusive regarding the contribution of teaching experience to student learning. Whereas a number of studies (e.g., Çakır & Bichelmeyer, 2016; Jung et al., 2014; Wenglinsky, 2002) reported

no statistically significant findings, other studies reported positive effects, either for the total number of years of experience (e.g., Goldhaber & Brewer, 1997) or for teaching experience in the school currently working (e.g., Canales & Maldonado, 2018). Additionally, other studies found positive effects only for experience at particular grade levels [e.g., Harris and Sass (2011) found positive effects only for elementary and middle-school, but not for high-school] or empirically corroborated curvilinear effects of teaching experience, yet with different peak levels [e.g., in Kukla-Acevedo (2009) the peak was in 14 years, whereas in Toropova et al. (2019) it was in 19 years].

Meta-analyses do not appear to shed more light on the effect of teaching experience on student learning either. For instance, in Greenwald et al. (1996) only 29% of the studies reported positive significant effects, whereas in 3% of the studies negative significant effects were reported; in most of the studies (68%), no significant effects were yielded. Other meta-analyses that were published during the last two decades (Coenen et al., 2018; Wayne & Youngs, 2003) found positive effects more consistently but in the latter one results were difficult to interpret because of different confounding factors. In the former one, findings diverged as some studies reported a positive effect only for the first three to five years and others a continuous beneficial effect of teaching experience, even up to 27 years.

In sum, although research findings provide some warrants that teaching experience appears to matter for student learning, the mixed study findings suggest that more work is needed to understand in what particular ways and under what specific conditions teaching experience does so.

2.3.3 Professional Development

A similar picture of mixed findings is also depicted in previous studies regarding the effect of professional development on student learning. Whereas no studies have been found to report negative effects, findings again range with some studies (e.g., Jacob & Lefgren, 2004; Loyalka et al., 2019) reporting no effects and others positive, albeit small effects (e.g., Wallace, 2009). Yet, there are nuances even in studies that paint a favourable picture of the contribution of professional development to student learning.

Specifically, it was shown that this effect is only present for particular grade levels and particular subject matters. For instance, Harris and Sass (2011) reported positive effects only for middle and high school mathematics teachers. Additionally, other studies showed that the content of professional development matters. For example, Wenglinsky (2002) found positive effects of professional development programs focusing on higher-order thinking skills and diversity. Desimone and colleagues (2010) went further by listing a set of

quality content characteristics that effective professional development programs had; those included a specific focus on content, active learning, coherence, duration, and the inclusion of collaborative activities. Similarly, in a recent meta-analysis of STEM programs, Lynch et al. (2019) listed a number of content and format features of effective professional development programs based on their meta-analysis findings. Content-wise, those included: (a) supporting teachers on the use of curriculum materials, (b) focusing on improving teachers' content and pedagogical content knowledge, and (c) understanding how students learn. Format-wise, they consisted of teacher meetings to troubleshoot and discuss implementation, as well as summer workshops.

In conclusion, despite inconclusive findings, there seems to be a pattern of positive findings especially when adding the content of the professional development programs into the equation. Nevertheless, future studies are needed to further explore how exactly specific aspects of professional development programs contribute to their effectiveness in terms of supporting student learning.

2.3.4 Teacher Knowledge

An increasingly growing body of studies has focused on the effects of teacher knowledge on student learning, examining different aspects of teacher knowledge, including content knowledge (CK) and pedagogical content knowledge (PCK). To a great extent, these studies support that teacher knowledge has a positive effect on student learning as different types of teacher knowledge have been examined. For example, positive associations between students' performance and teachers' CK were identified (e.g., Metzler & Woessman, 2012). Moreover, studies have shown teachers' PCK to be a better predictor of student learning compared to teachers' CK (e.g., Baumert et al., 2010; Campbell et al., 2014; see also the meta-analysis by Depaepe et al., 2013).[1] Hence, these positive findings are important for teacher education, especially when taking into consideration that PCK can be developed during initial or ongoing teacher education.

2.4 *Summarising Existing Evidence*

Summarising the findings reported above and seeing the glass half empty, one could wonder whether we have really made any progress or whether we are more or less where Begle had arrived more than 40 years ago. In fact, a cynical critic might even argue that the inconclusiveness in the findings reported actually vindicates Begle's admonition that scholarly attention should be directed to other, more productive, paths. Yet, seeing the glass half full, one could argue that we are now aware of certain teacher characteristics for which there seems to be a positive pattern of contribution (e.g., teacher knowledge

and professional development under certain conditions). What is more, even the inconclusiveness in the findings should not be taken as an indication of failure but should rather be seen as the underlining of a need for more work to untangle the nuances in the phenomena examined. Scheerens (2015) talks about taking such a stance when urging scholars to not be afraid to look at "the dark side of the moon" – at factors that at first hand do not seem to "work". In light of this recommendation, in the next section, we focus on four lessons that can be learned from the successes and failures in understanding how teacher quality characteristics contribute to student learning.

3 Four Lessons Learned from the Results of Large-Scale Studies Focusing on Teacher Quality

3.1 *Seeing the Bigger Picture*

Several of the studies reviewed above have considered the focal teacher characteristics and qualifications mostly in isolation, without any of them bringing different characteristics together. It is thus necessary to explore teacher quality characteristics more comprehensively as some of the disparate findings reported above can be attributed to omitted variable bias (cf. Hill et al., 2019).

At least five studies attempted to examine the contribution of teacher quality characteristics more comprehensively (Boonen et al., 2014; Campbell et al., 2014; Grubb, 2008; Hill et al., 2019; Palardy & Rumberger, 2008). In addition to the four categories of teacher qualifications and characteristics explored above, these studies also examined teacher attitudes and beliefs. For example, by focusing on teacher preparation and experience as well as teacher attitudes, Palardy and Rumberger (2008) found that whereas preparation and experience did not explain student learning, teacher attitudes, and in particular, teachers' efficacy beliefs did. Grubb (2008) reported positive relationships between student outcomes in mathematics in the NELS:88 data and a variety of teacher background and preparation characteristics (e.g., experience, teaching in-field, education track), along with teacher efficacy beliefs. In Boonen and colleagues' (2014) work, teaching experience and job satisfaction – a background characteristic and attitude respectively – predicted Flemish students' mathematics outcomes. Lastly, Campbell et al. (2014) found that teacher knowledge was positively associated with student outcomes in contrast to special education certification which was negatively associated with them. Teacher attitudes and beliefs largely had no effects outside interactions with knowledge itself.

Building on a wider set of teacher qualifications and characteristics from different categories (teacher preparation, teaching experience, teacher knowledge, and teacher attitudes and beliefs), Hill and colleagues (2019) have more recently examined the contribution of different teacher quality aspects on student learning. They did so by using two different types of student tests: (a) standardised state mathematics tests primarily focusing on basic skills and problem solving, and (b) a cognitively demanding test capturing students' mathematical reasoning and higher levels of thinking. By bringing together different teacher quality characteristics, these scholars were able to examine their effect net of that of other variables. For instance, this exploration showed coursework (both content and methods courses) as well as teacher knowledge to be positively related to student learning on both types of tests; teaching experience to be related only to student learning as measured by the state mathematics tests, and teacher beliefs and attitudes to have no significant effect on student learning. Interestingly, when bringing these different variables together, the two sets of variables that were directly related to teacher education – teacher coursework and knowledge – were found to be important contributors to different types of student learning whereas other variables not directly related to teacher education were not. As such, this study showcases the promise of more comprehensive explorations when examining the effect of teacher education on student learning. However, given that the effects obtained even from this study were small to moderate at best, more comprehensive studies that examine even a wider set of teacher quality aspects are warranted.

3.2 *Exploring (In)Consistency*

Inconclusive results have often been regarded as problematic and have been associated with the "noise" that exists in a system when studying a phenomenon. Under this assumption, attempts are made to minimise the noise in order to better detect "the signal". However, several scholars (e.g., Hall et al., 2020; Scheerens, 2015; Scheerens & Blömeke, 2016) increasingly propose adopting a different stance; one that regards such inconclusive results as part of the signal rather than the noise. In this context, they argue that attempts should be made to better understand what contributes to the inconsistencies identified upon examining a particular phenomenon.

About 40 years ago, Shulman (1986) advocated the inclusion of content back in the equation of teaching understanding. Since then, scholars have increasingly attended to the demands that teaching particular subject matters imposes on teachers (see, for example, the emphasis placed on identifying

different teaching practices for different subject matters: in mathematics, see, for example, Mitchell et al., 2014; in history, see Fogo, 2014; in science, see Windschitl et al., 2012; for a more general discussion, see Grossman, 2018). Scholars within the field of educational effectiveness (e.g., Charalambous et al., 2014; Muijs et al., 2016; Scheerens, 2016) have also underlined the importance of moving beyond generic practices to also explore differences in effectiveness across subject matters – a need that has been empirically corroborated by recent evidence (e.g., Charalambous et al., 2019; Cohen, 2018). The overview of studies presented in the previous section underlines the necessity to introduce subject matter as a moderator in the explorations undertaken, since, notable differences were identified across subject matters regarding the contribution of several indicators of teacher quality to student learning.

In addition to considering the role of the content, Hall and colleagues (2020) urge for resisting the "context-stripping" tendency that appears to have permeated research on educational effectiveness during the past decades. Thus, the role of context also needs to be carefully examined as a potential contributor to the results obtained in studies exploring teacher effects. This implies that scholars need to invest in examining the consistency of teacher effects across different educational systems and countries [especially systems and countries that differ in their teacher education policies, see more on those differences in Tatto and Pippin (2017) as well as Brown (2017)]; across different student populations, in terms of their age and other background characteristics; as well as across different levels of schooling (pre-primary, primary, secondary-general, secondary-vocational, and tertiary education).

Research attempts can also be invested in considering different types of student outcomes as in most of the studies examined, scholars have focused on cognitive outcomes. Given that different types of student outcomes can lead to different conclusions about the contribution of teacher quality indicators to student learning (cf. Cappella et al., 2016; Lindorff et al., 2020; Reynolds et al., 2016; Scheerens & Blömeke, 2016) future studies need to broaden the type of outcomes examined to also incorporate non-cognitive outcomes. The results of Hill et al.'s (2019) study recounted above, that showed differences even across dissimilar cognitive outcomes, also underline the value of expanding the examination of both cognitive and non-cognitive types of outcomes, as well as differentiating even within the same type of outcomes.

Two relatively recent large-scale studies (Blömeke et al., 2016; Blömeke & Olsen, 2019) suffice for stressing the importance of emphasising and understanding the inconsistencies found in teacher quality effectiveness research. In the first study, the authors examined five teacher characteristics (teaching experience, teacher education degree, major focus in studies, professional

development, and sense of preparedness) across 47 participating countries. Their results showed significant variation across countries, leading the authors to conclude that no empirical support can be provided for a universal model. Similar results were obtained in the second study that examined the effect of four teacher quality indicators (i.e., teacher education major, teaching experience, professional development, and sense of preparedness) on two student outcomes (achievement, and motivation) across five countries, by recruiting students from two grades (Gr. 4 and Gr. 8) and focusing on two subject matters (mathematics and science). The study showed little consistency, if at all, both across and even more critically within countries, with respect to the outcomes, the subject matters, and the student populations examined. In conjunction, both these studies suggest that inconsistencies are to be expected when introducing different moderating factors (e.g., educational system/country, subject matter, student population, type of outcomes) into the equation. More than expecting such inconsistencies, scholars (e.g., Scheerens, 2016; Lindorff et al., 2020) advocate their systematic exploration and understanding, shifting in consequence the question "what works?" into also investigating under what conditions, how, and for which particular reasons different teacher quality factors might contribute to student learning.

3.3 *Adding Teaching Quality to the Equation*

The studies reviewed in Section 2 have mostly considered teacher quality but did not open up the black box of teaching in order to understand how different teacher characteristics and qualifications can contribute to student learning through the improvement of teaching quality. Hence, it is encouraging that over the past two decades scholars within the field of educational effectiveness have not only stressed the importance of bringing together teacher and teaching quality factors – in an attempt to better understand what contributes to student learning – but have also proposed different theoretical frameworks and models for doing so (see, for example, Blömeke et al., 2015; Creemers & Kyriakides, 2008; Nilsen et al., 2016; Scheerens, 2016).

Studies that have explored both teacher and teaching quality effects corroborate the need of adding teaching quality into the equation. This becomes quintessential for at least two reasons. First, teacher characteristics and qualifications might have an indirect effect on student learning through teaching quality. This was suggested, for example, in Blömeke et al.'s (2016) study, which, in addition to the teacher characteristics and qualifications discussed above, also considered teaching quality. In that study, whereas the direct impact of professional development on student performance was not significant in most of the 47 countries examined, professional development turned out to be the

strongest predictor of teaching quality in 23 of the countries. This implied that professional development could have an effect on student learning largely because of its contribution to teaching quality. Second, introducing teaching quality to the equation can help better link teacher to teaching quality aspects. For instance, studies (albeit smaller-scale, qualitative ones) have shown that teachers with different levels of knowledge (e.g., Hill et al., 2008; Santagata & Lee, 2021), beliefs and attitudes (e.g., Sleep & Eskelson, 2012) or combinations thereof (e.g., Charalambous, 2015; Zhang, 2022) differ in the quality of their lessons, thus pointing to an important link between certain teacher characteristics and aspects of their teaching. In sum, by incorporating teaching quality into the picture, scholars might be in a better place to understand how teacher education can contribute to teachers' work, and in turn, to student learning.

3.4 *Capitalising on Complementarity*

Large-scale studies – be they national or international – are important for understanding *what* teacher qualifications and characteristics contribute to student learning. Yet, as suggested by the review of the studies in the previous section, they have their own limitations in shedding light on *how* and *why* teacher education can support student learning. This is because, due to their design, such studies cannot provide answers on how teacher and teaching quality can contribute to student learning, how teacher education might support (or not) changes in teaching quality and through that, changes in student learning, and *why* (in)consistent results emerge.

Due to these limitations, it is argued that there is significant benefit in combining large-scale with small-scale studies. Indeed, for years, an overemphasis on small-scale studies has been accused of producing results that oftentimes were very particular to the context in which they were generated and could hardly inform broader educational policies. For example, almost thirty years ago, Cooney (1994) bemoaned the limitations of small-scale studies – and particularly case-studies – in teacher education, urging scholars to "move beyond collecting interesting stories" to start seeing "how those stories begin to tell a larger story" (p. 627). However, moving away from small-scale studies completely runs the risk of pushing the pendulum to the exact opposite end (large-scale studies only). Hence, our argument here is that we need to strike a balance between both types of studies, capitalising on the benefits of both, since such complementarity can help better understand the complex phenomena of teaching and student learning.

Small-scale (qualitative) studies have a number of affordances upon which future scholarly work can capitalise. First, they can unravel the mechanisms of how teachers learn and how this learning might affect their teaching, and

consequently student learning. Such explorations might actually challenge what is considered by some scholars (e.g., Cochran-Smith, 2021; Mayer, 2021) as a simplistic chain of associations (teacher learning → improved teaching quality → improvements in student learning), thus helping better understand complexities and nuances in these associations. Second, small-scale studies can assist in discerning the conditions under which teacher and consequently student learning can be supported. This resonates with the scholarly calls listed above of achieving a deeper understanding of the conditions under which certain factors/variables work. Third, and equally important, small-scale studies can help explain the inconsistent findings often emerging from large-scale studies.

To illustrate the promise of these studies, we next briefly share two such studies from our work, one conducted with pre-service teachers and another conducted with in-service teachers. Neither study is meant to be presented as exemplary; rather, both of them illustrate how doing research in and on teacher education can yield important insights into supporting teacher learning.

The first study (Charalambous et al., 2018) focuses on three pre-service teachers who were followed during their practicum and were also supported in their work by being engaged in guided analyses of teaching practice – theirs and that of their colleagues – in a video-club setting (see more on such settings in Sherin & van Es, 2009). The analysis of these pre-service teachers' lessons during their practicum documented differences in the learning trajectories both in planning and enacting of their lessons. More than suggesting that teachers benefit in different degrees when exposed to particular interventions, these differences challenge a tendency to consider teacher learning on the average – an inherent feature of large-scale studies. As such, these findings illustrate the need to better understand why such differences occur and how pre-service teachers' characteristics along with the characteristics of the intervention interact, yielding these different learning trajectories.

Utilising the same idea of video-clubs, the second study (Charalambous et al., 2023) examined in-service teachers' experimentation with ambitious teaching. The study documented how practicing teachers can be scaffolded to materialise such ambitious teaching visions in their practice through the use of certain praxis tools – namely tools that can help them materialise complex theoretical ideas in their practice. By portraying the changes that five practicing teachers introduced in their teaching and the challenges they encountered while trying to teach ambitiously, this study provided an account of what "typical" practicing teachers can achieve in their daily practice. Therefore, such studies can provide what Lampert et al. (2011) called "images and narratives for ambitious teaching that portray how one can be a mere mortal and yet

capably meet its routine demands" (p. 1394). Such images and narratives are important as they help illustrate that types of more demanding teaching are feasible even for "typical"/average teachers. At the same time, by documenting the challenges that these teachers face in their practice, such studies can provide important insights for teacher educators in terms of how they can better support practicing teachers in responding to more complex types of teaching. We argue that large-scale studies do not easily lend themselves to producing such important insights and lessons, whereas smaller-scale, qualitative studies can better support such explorations.

4 Looking Forward

What was discussed in the previous two sections has important implications for teacher education research. In this last section, we discuss three such implications and conclude by discussing how the value of teacher education research can be enhanced through large-scale studies.

The first implication relates to the value of continuing explorations on the contribution of certain promising teacher characteristics to teaching quality and student learning through large-scale studies. Although studies of this type have been conducted in abundance in the previous decades, we argue that there is merit in continuing this line of work (but also adapting and complementing it as will be discussed next). For example, one of the most promising characteristics yielded from prior research relates to teacher knowledge. Yet, several questions remain unaddressed – or are partly addressed – when it comes to its contribution. For instance, what types of teacher knowledge have larger effects and why? How consistent are the results across different contents and contexts? If inconsistencies arise, what might account for them? Equally critical, if teacher knowledge is so important, how do these types of knowledge develop during initial and ongoing teacher education? As already argued, addressing all these questions only through large-scale studies is impossible given that large-scale studies can, at best, help us address only what works and under which conditions. Hence, there is a need to complement such large-scale studies with small-scale (qualitative studies) that can help unravel how and why things work.

The second implication relates to the importance of continuing the investigations of not so promising characteristics and others for which extant studies have yielded inconclusive and mixed results (e.g., professional development). Future studies should, however, not be geared toward replicating the results of prior research, but largely to help better understand under what conditions the

effects of these characteristics might be strengthened or even optimised. For example, when it comes to professional development – a key feature of teacher education – future scholarly attempts could be directed toward addressing questions such as, what characteristics render professional development programs more or less effective? Why is this the case? In addition, smaller scale studies can also help examine the mechanisms through which professional development can inform teaching quality and through that student learning.

The third implication relates to using more comprehensive designs in how we examine teacher and teaching effects. The recent OECD (2020) TALIS video-study *Global Teaching InSights* provides one such example of a large-scale study conducted at an international level. Focusing on eight different countries, in this study, scholars explored different teacher characteristics, aspects of teaching quality, and types of student learning. Such studies, especially when conducted at an international level, can help explore different moderation and mediation effects, thus further enhancing existing knowledge about the contribution of teacher and teaching characteristics to different types of student learning. However, when such comprehensive designs are difficult to run at an international level, national research agencies and centres might complement (international) large-scale studies by adding the missing pieces of the complex chain of associations that links teachers, teaching, and student learning. This is something that, for instance, scholars of the COACTIV project did in the past (see Baumert et al., 2010), when complementing the PISA study with research components that better allowed the concurrent studying of teacher characteristics (such as teacher knowledge), different aspects of teaching quality, and student learning.

At the beginning of this chapter, we referred to critiques often voiced regarding the field of teacher education research – critiques that, to some extent, might be nurtured by the difficulties of this body of research to directly inform policy decisions. Coming full circle, we conclude by discussing how large-scale studies can help enhance the value of this field of research. We see three ways in which this can be done.

First, although generalisations are often particularly desirable for policy making, nowadays it seems to be increasingly understood that generalizable patterns might be neither feasible nor productive to derive, especially when it comes to studying complex phenomena like teaching and learning. By exploring different moderators to the association between teacher characteristics, teaching quality characteristics, and student learning, large-scale studies can produce results that, albeit of limited generalisability than those policymakers might be longing for, take into consideration different contextual factors. Doing so is important not only for further developing teacher education research but

also for producing less inconclusive results – which can also be considered a significant limitation of teacher education research.

Second, neither large-scale studies nor small-scale studies alone can help us understand comprehensively the complex chain of associations between teacher characteristics, teaching, and student learning. As argued above, a productive combination of large-scale and small-scale (qualitative) studies can help us delve deeper into exploring these associations. This complementarity can contribute toward understanding not only what works and under what conditions, but also how and why things work. We maintain that addressing these different types of questions can help uplift the status of teacher education research since complex educational phenomena require a more comprehensive and holistic approach in studying them. Otherwise, according to the well-known Indian fable of the seven blind men studying an elephant, teacher education researchers run the risk of producing fragmentary knowledge pieces which can hardly move the field really forward.

Finally, another way to elevate the status of teacher education research by producing more consistent and applicable findings, that can inform teacher educators as well as policy makers, lies in combining explorations of teacher effects with teaching quality effects. It is encouraging that during the past years such investigations are observed more frequently than in the past (see, for example, Blömeke et al., 2016; OECD, 2020). We argue, however, that they need to be intensified and that researchers have to experiment with different ways of measuring teacher and teaching quality effects (e.g., teacher reports, principal reports, student reports, classroom observations, teacher logs) to better capture and study these effects. Such explorations are envisioned to produce more nuanced results that better lend themselves to informing the design of different teacher education programs as well as decision-making at different levels.

In conclusion, large-scale studies can uplift the status of teacher education research through their contributions, but only if they are critically examined on what they can help us achieve and how. Despite their shortcomings, when they are used in combination with other types of studies their limitations can be turned into affordances. Thus, upon reflecting on Begle's (1979) and Scheerens' (2015) admonitions, it can be argued that there are still very productive research paths on teacher and teaching quality when considering the whole of the moon, looking both at its bright and dark sides.

Note

1 More recently, positive associations were also yielded between student learning and teachers' general pedagogical knowledge (GPK) (e.g., König et al., 2021). We do not report on these

studies above, although they resonate with the findings of CK and PCK, because they do not meet the sample-size criterion of 2500 student participants.

References

Anderson, L. W., & Postlethwaite, T. N. (2007). *Program evaluation: Large-scale and small-scale studies*. *International Academy of Education*. www.unesco.org/iiep/PDF/Edpol8.pdf

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Begle, E. G. (1979). *Critical variables in mathematics education: Findings from a survey of the empirical literature*. Mathematical Association of America and the Council of Teachers of Mathematics.

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (Eds.). (2015). *Assessment of competencies in higher education*. Hogrefe.

Blömeke, S., Olsen, R. V., & Suhl, U. (2016) Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (IEA Research for Education, Vol. 2, pp. 21–50). Springer. https://doi.org/10.1007/978-3-319-41252-8_2

Blömeke, S., & Olsen, R. V. (2019). Consistency of results regarding teacher effects across subjects, school levels, outcomes and countries. *Teaching and Teacher Education, 77,* 170–182. https://doi.org/10.1016/j.tate.2018.09.018

Boonen, T., Van Damme, J., & Onghena, P. (2014). Teacher effects on student achievement in first grade: Which aspects matter most? *School Effectiveness and School Improvement, 25*(1), 126–152. https://doi.org/10.1080/09243453.2013.778297

Brown, G. T. L. (2017). What we know and what we don't know about teacher education. In D. J. Clandinin & J. Huku (Eds.), *The Sage international handbook of research on teacher education* (pp. 123–138). Sage.

Çakır, H., & Bichelmeyer, B. A. (2016). Effects of teacher professional characteristics on student achievement: An investigation in blended learning environment with standards-based curriculum. *Interactive Learning Environments, 24*(1), 20–32. https://doi.org/10.1080/10494820.2013.817437

Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., DePiper, J. N., Frank, T. J., Griffin, M. J., & Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal of Research in Mathematics, 45*(4), 419–459. https://doi.org/10.5951/jresematheduc.45.4.0419

Canales, A., & Maldonado, L. (2018). Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics. *International Journal of Educational Development, 60*, 33–50. https://doi.org/10.1016/j.ijedudev.2017.09.009

Charalambous, C. Y. (2015). Working at the intersection of teacher knowledge, teacher beliefs, and teaching practice: A multiple-case study. *Journal of Mathematics Teacher Education, 18*(5), 427–445. https://doi.org/10.1007/s10857-015-9318-7

Charalambous, C. Y., Agathangelou, S., Kasapi, E., & Christofidou, E. (2023). Learning to teach ambitiously: a multiple case-study of practicing teachers' experimentation with enablers and extenders. *Journal of Mathematics Teacher Education, 26*, 363–394. https://doi.org/10.1007/s10857-022-09532-9

Charalambous, C. Y., Komitis, A., Papacharalambous, M., & Stefanou, A. (2014). Using generic and content-specific teaching practices in teacher evaluation: An exploratory study of teachers' perceptions. *Teaching and Teacher Education, 41*(1), 22–33. https://doi.org/10.1016/j.tate.2014.03.001

Charalambous, C. Y., Kyriakides, E., Kyriakides, L., & Tsangaridou, N. (2019). Are teachers consistently effective across subject matters? Revisiting the issue of differential teacher effectiveness. *School Effectiveness and School Improvement, 30*(4), 353–379. https://doi.org/10.1080/09243453.2019.1618877

Charalambous, C. Y., Philippou, S., & Olympiou, G. (2018). Reconsidering the use of video clubs for student-teachers' learning during field placement: Lessons drawn from a longitudinal multiple case study. *Teaching and Teacher Education, 74*, 49–61. https://doi.org/10.1016/j.tate.2018.04.002

Cochran-Smith, M. (2021). Rethinking teacher education: The trouble with accountability. *Oxford Review of Education, 47*(1), 8–24. https://doi.org/10.1080/03054985.2020.1842181

Coenen, J., Cornelisz, I., Groot, W., Maassen van den Brink, H., & Van Klaveren, C. (2018). Teacher characteristics and their effects on student test scores: A systematic review. *Journal of Economic Surveys, 32*(3), 848–877. https://doi.org/10.1111/joes.12210

Cohen, J. (2018). Practices that cross disciplines? Revisiting explicit instruction in elementary mathematics and English language arts. *Teaching and Teacher Education, 69*, 324–335. https://doi.org/10.1016/j.tate.2017.10.021

Cooney, T. J. (1994). Research and teaching education: In search of common ground. *Journal of Research in Mathematics Education, 25*(6), 608–636. https://doi.org/10.5951/jresematheduc.25.6.0608

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12–25. https://doi.org/10.1016/j.tate.2013.03.001

Desimone, L. M., Smith, T. M., & Frisvold, D. E. (2010). Survey measures of classroom instruction comparing student and teacher reports. *Educational Policy*, *24*(2), 267–329. https://doi.org/10.1177/0895904808330173

Fogo, B. (2014). Core practices for teaching history: The results of a Delphi panel survey. *Theory & Research in Social Education, 42*(2), 151–196. https://doi.org/10.1080/00933104.2014.902781

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, *32*(3), 505–523. https://doi.org/10.2307/146181

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361–396. https://doi.org/10.3102/00346543066003361

Grossman, P. (Ed.). (2018). *Teaching core practices in teacher education.* Harvard Education Press.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184–205. https://doi.org/10.3102/0002831207312906

Grubb, W. N. (2008). Multiple resources, multiple outcomes: Testing the "improved" school finance with NELS88. *American Educational Research Journal, 45*(1), 104–144. https://doi.org/10.3102/0002831207308636

Hall, J., Sammons, P., & Lindorff, A. (2020). Continuing towards international perspectives in educational effectiveness research. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research* (pp. 383–406). Springer.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality, and student achievement. *Journal of Public Economics, 95*(7–8), 789–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Hill, H. C., Blunk, M. Charalambous, C. Y., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. https://doi.org/10.1080/07370000802177235

Hill, H. C., Charalambous, C. Y., & Chin, M. (2019). Teacher characteristics and student learning in mathematics: A comprehensive assessment. *Educational Policy, 33*(7), 1103–1134. https://doi.org/10.1177/0895904818755468

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226–244. https://doi.org/10.1162/003465304323023778

Jung, E., Brown, E. T., & Karp, K. S. (2014). Role of teacher characteristics and school resources in early mathematics learning. *Learning Environments Research, 17*(2), 209–228. https://doi.org/10.1007/s10984-014-9159-9

König, J., Blömeke, S., Jentsch, A., Schlesinger, L., née Nehls, C. F., Musekamp, F., & Kaiser, G. (2021). The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educational Studies in Mathematics, 107*(1), 189–212. https://doi.org/10.1007/s10649-020-10021-0

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review, 28*(1), 49–57. https://doi.org/10.1016/j.econedurev.2007.10.007

Kyriakides, L., & Charalambous, C. Y. (2014). Educational effectiveness research and international comparative studies: Looking back and looking forward. In R. Strietholt, W. Bos, J.-E. Gustafsson, & M. Rosén (Eds.), *Educational policy evaluation through international comparative assessments* (pp. 33–50). Waxmann.

Lampert, M., Boerst, T., & Graziani, F. (2011). Organizational resources in the service of school-wide ambitious teaching practice. *Teachers College Record, 113*(7), 1361–1400. https://doi.org/10.1177/016146811111300706

Lindorff, A., Sammons, P., & Hall, J. (2020). International perspectives in educational effectiveness research: a historical overview. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research* (pp. 9–31). Springer.

Loyalka, P., Sylvia, S., Liu, C., Chu, J., & Shi, Y. (2019). Pay by design: Teacher performance pay design and the distribution of student achievement. *Journal of Labor Economics, 37*(3), 621–662. https://doi.org/10.1086/702625

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis, 41*(3), 260–293. https://doi.org/10.3102/0162373719849044

Mayer, D. (2021). The connections and disconnections between teacher education policy and research: Reframing evidence. *Oxford Review of Education, 47*(1), 120–134. https://doi.org/10.1080/03054985.2020.1842179

Mayer, D., & Oancea, A. (2021). Teacher education research, policy and practice: Finding future research directions. *Oxford Review of Education, 47*(1), 1–7. https://doi.org/10.1080/03054985.2021.1853934

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics, 99*(2), 486–496. https://doi.org/10.1016/j.jdeveco.2012.06.002

Middleton, J. A., Cai, J., & Hwang, S. (2015). Why mathematics education needs large-scale research. In J. A. Middleton, J. Cai, & S. Hwang (Eds.), *Large-scale studies in mathematics education* (pp. 1–13). Springer.

Mitchell, R., Charalambous, C. Y., & Hill, C. H. (2014). Examining the task and knowledge demands needed to teach with representations. *Journal of Mathematics Teacher Education, 17*, 37–60. https://doi.org/10.1007/s10857-013-9253-4

Muijs, D., Reynolds, D., & Kyriakides, L. (2016). The scientific properties of teacher effects/effective teaching process. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement: Research, policy, and practice* (pp. 100–123). Routledge.

Nilsen, T., Gustafsson, J.-E., & Blömeke, S. (2016). Conceptual framework and methodology of this report. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality, and student outcomes: Relationships across countries, cohorts, and time* (pp. 1–19). Springer.

OECD. (2009). *Creating effective teaching and learning environments: First results from TALIS*. OECD Publishing. https://www.oecd.org/education/school/43023606.pdf

OECD. (2014). *TALIS 2013 results: An international perspective on teaching and learning TALIS*. OECD Publishing. http://dx.doi.org/10.1787/9789264196261-en

OECD. (2019). *TALIS 2018 results (Volume I): Teachers and school leaders as lifelong learners.* OECD Publishing. https://doi.org/10.1787/1d0bc92a-en

OECD. (2020). *Global teaching InSights: A video study of teaching.* OECD Publishing. https://doi.org/10.1787/20d6f36b-en

Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111–140. https://doi.org/10.3102/0162373708317680

Reynolds, D., Chapman, C., Clarke, P., Muijs, D., Sammons, P., & Teddlie, C. (2016). Conclusions: The future of educational effectiveness and improvement research, and some suggestions and speculations. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement: Research, policy, and practice* (pp. 408–439). Routledge.

Santagata, R., & Lee, J. (2021). Mathematical knowledge for teaching and the mathematical quality of instruction: A study of novice elementary school teachers. *Journal of Mathematics Teacher Education, 24*(1), 33–60. https://doi.org/10.1007/s10857-019-09447-y

Scheerens, J. (2015). Theories on educational effectiveness and ineffectiveness. *School Effectiveness and School Improvement, 26*(1), 10–31. https://doi.org/10.1007/978-94-017-7459-8_11

Scheerens, J. (2016). *Educational effectiveness and ineffectiveness: A critical review of the knowledge base*. Springer.

Scheerens, J., & Blömeke, S. (2016). Integrating teacher education effectiveness research into educational effectiveness models. *Educational Research Review, 18*, 70–87. https://doi.org/10.1016/j.edurev.2016.03.002

Sherin, M., & Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20–37. https://doi.org/10.1177/0022487108328155

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102%2F0013189X015002004

Sleep, L., & Eskelson, S. (2012). MKT and curriculum materials are only part of the story: Insights from a lesson on fractions. *Curriculum Studies, 44*(4), 537–558. https://doi.org/10.1080/00220272.2012.716977

Tatto, M. T., & Pippin, J. (2017). The quest for quality and the rise of accountability systems in teacher education. In D. J. Clandinin & J. Huku (Eds.), *The Sage international handbook of research on teacher education* (pp. 68–89). Sage.

Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M., & Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA Teacher Education and Development Study in Mathematics* (TEDS-M). International Association for the Evaluation of Student Achievement.

Toropova, A., Johansson, S., & Myrberg, E. (2019). The role of teacher characteristics for student achievement in mathematics and student perceptions of instructional quality. *Education Inquiry, 10*(4), 275–299. https://doi.org/10.1080/20004508.2019.1591844

Wallace, M. R. (2009). Making sense of the links: Professional development, teacher practices, and student achievement. *Teachers College Record, 111*(2), 573–596. https://doi.org/10.1177/016146810911100205

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122. https://doi.org/10.3102/00346543073001089

Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*(12), 12. https://doi.org/10.14507/epaa.v10n12.2002

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education Policy, 96,* 878–903. https://doi.org/10.1002/sce.21027

Zhang, Q. (2022). Understanding Chinese mathematics teaching: How secondary mathematics teachers' beliefs and knowledge influence their teaching in mainland China. *ZDM–Mathematics Education,* 1–15. https://doi.org/10.1007/s11858-022-01336-8