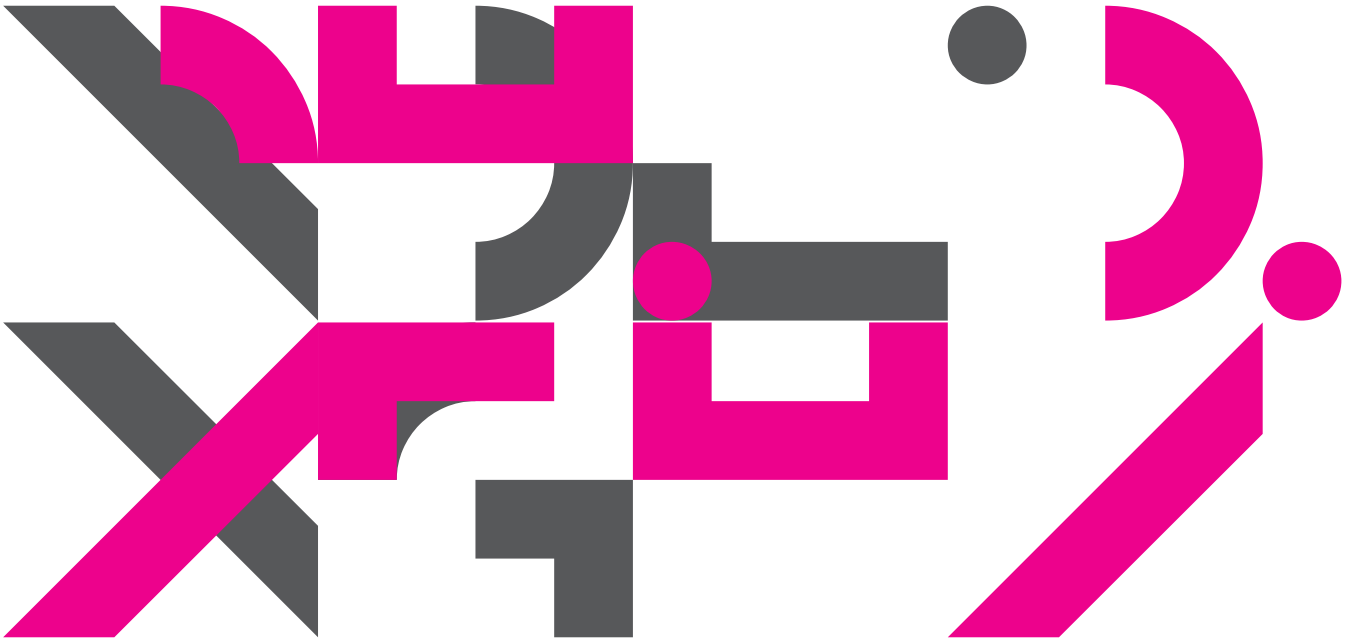

PH.D. THESIS

ANDERS ASTRUP CHRISTENSEN

INTERNATIONALLY COMPARATIVE PERSPECTIVES ON PROFESSIONAL LEARNING COMMUNITIES

AARHUS
UNIVERSITY
DANISH SCHOOL OF EDUCATION

Internationally Comparative Perspectives on Professional Learning Communities

(Professionelle læringsfællesskaber i et internationalt perspektiv [Danish title])

Ph.D. thesis handed in at the Danish School of Education (DPU), Aarhus University.

Main supervisor: Felix Weiss.

Co-supervisors: Sedat Gümüş & Kristoffer Laigaard Nielbo.

The Ph.D.-scholarship was funded by the Carlsberg Foundation under grant CF19-0751.

Note on Published Chapters:

The original version of this thesis, as submitted for defense, included full drafts of Chapters 2, 3, and 4. Chapter 2 had already been published at the time, while Chapters 3 and 4 were later accepted for publication in peer-reviewed journals. In this final version of the thesis, the full text of all three chapters has been replaced with pages containing direct links to the published articles

ISBN: 978-87-7507-571-3

DOI: https://doi.org/10.7146/aul.547

Preface

Around the world, the professional development of teachers is subject to intense debate among scholars, policy makers and practitioners, which reflects the perceived importance of teacher professional development towards meeting the continuously changing demands of schools and their outputs in terms of students' knowledge and skills. What are the best conditions for teachers to effectively collaborate? What are the effects of structured collaboration on the quality of their practices, on students learning, and what does it mean for the teachers themselves? This debate has also influenced the educational debates and initiatives in Denmark, not least following the major school reform in 2014, which, among other things, led to an increased focus on teacher professionalism. The ambitions of this reform led to substantial investments in the development of teachers' skills and practices through collaboration, not least by implementation of Professional Learning Communities (PLCs) in Danish schools[1]. There was a rather weak connection between these practical implementations and evidence from empirical education research. For example, in the materials used in the implementation of PLCs, and from institutions offering courses and consultation on how to implement them, PLCs were often presented alongside strong claims of their positive effects on student achievement and desirable outcomes for teachers. These claims could sometimes be found without supportive evidence, and when references were provided, these were most frequently textbooks or a few studies conducted outside Denmark, or even the Nordic countries which Denmark is most frequently compared to.

[1] The concept has been applied using different labels, such as "Professionelle Læringsfællesskaber" and "Stærkere Læringsfællesskaber" (Danmarks Evalueringsinstitut, 2018; Krøger, 2021; Thorborg & Qvortrup, 2024).

Engaging teachers in PLCs requires a significant amount of time, especially in the implementation phase, but also in the long term as every teacher in the school continuously spends time collaborating with their colleagues on clarifying what the students should learn, how they should be taught, and evaluating the outcomes. Across schools, grade levels, subjects and teachers, this entails a significant time commitment. At the time I applied for funding for this PhD project at the beginning of 2020, not much research had been conducted into the success of the efforts in implementing PLCs in Denmark. This raised a critical question: How can we know whether these substantial efforts produce the proclaimed effects?

Looking further into the body of research on PLCs and the extent to which it was applied globally, I found that the empirical evidence supporting these strong claims was questionable. Additionally, the potential influence of the context on the expected outcomes of PLCs for students and teachers seemed to be overlooked or neglected in terms of how the concept was being applied across the globe, referring to the same literature and research, which predominantly comes from the United States and the United Kingdom. The lack of knowledge on PLCs and their expected effects in general, and especially in relation to the differences in context, did not only pertain to the case of Denmark but was identified as a research gap in the international literature on a widely applied concept in education.

While the initial interest came from the specific case of the Danish context, the thesis aims to contribute to international and comparative research on PLCs, including how to measure the concept, the role of the contexts in which the PLCs operate, and the expected outcomes for teachers engaged in them.

Aarhus, May 2024.

Contents

Summary

Professional Learning Communities (PLCs) have gained popularity in recent decades as a way to structure teachers' professional development through an ongoing collaboration with colleagues within their schools, with the ultimate aim of improving student learning. PLCs have been applied globally, with expectations of increased student learning and positive outcomes for teachers such as increased job satisfaction and self-efficacy. However, the initial research supporting these claims is predominantly from a small number of western countries, and more recent studies have produced inconsistent results. Additionally, there are vast differences in the measures of PLCs applied, which makes it challenging to compare findings across studies. Significant investments have been made to restructure schools into PLCs internationally, but evidence of the expected effects of the PLCs is relatively sparse, especially when it comes to the potential role of the context in which they are enacted, both at the school level and across different countries. The substantial investments of time committed to PLCs across countries, schools, grade levels and subjects, raises an important question of whether the many hours spent on PLCs does lead to the expected outcomes.

In this thesis, I approach these gaps in the literature by employing data from the Teaching and Learning International Study (TALIS) 2018, to develop an internationally comparable measure of PLCs, which provides the basis for an explorative study of the school contextual factors related to PLCs, and a cross-national analysis of the relationship between PLCs and teachers' job satisfaction, self-efficacy, and clarity of instruction. The main elements of the thesis are three articles, which will be outlined in the following, and an introductory chapter. The introductory chapter situates the three articles within the general field of research and elaborates and discusses some key aspects within and across the three

articles as well as their limitations. Finally, the results are summarized and discussed with reference to existing research and potential directions for further research are outlined.

The first article, 'A Global Measure of Professional Learning Communities', presents the development of a measure of PLCs using data from the TALIS 2018 study. A measure of PLCs at the school level is created using responses from individual teachers within the schools. Using multilevel confirmatory factor analysis to model a shared cluster construct, the measure developed includes a second-order construct of PLCs, and three subdimensions: Collaborative practice with focus on student learning (CPL), Shared vision and responsibilities (SVR), and Supportive conditions (SC). These subdimensions are suitable for comparisons across 42 countries, align with the criteria and estimation techniques of the official TALIS scales, and the measures are available for use by other researchers interested in further studying PLCs within the TALIS 2018 data, either for international comparative analysis or within specific countries. The measures developed in this article are applied in the following two articles.

The second article, 'Exploring School Factors Related to Professional Learning Communities: A Machine Learning Approach Using Cross-National Data', was co-authored by Kristoffer Laigaard Nielbo and Sedat Gümüş. Few studies have focused on the potential influences of school contextual factors on well-functioning PLCs. Most of the relevant studies focus on various leadership models, and across these studies, the specific models used are emphasized in seemingly incompatible ways. Equipped with a measure of PLC which is derived from teachers' responses, we took advantage of the vast information available from the principal questionnaire of TALIS 2018, which covers many different aspects of the individual school such as autonomy, various resources, school-, staff- and student characteristics and principals' values and behaviors. We took an explorative approach to assess which of the many school factors were the most important predictors of the level of

PLC within the school, using the pooled international data from 42 countries. To do so, we compared the predictive performance of five different machine learning algorithms and interpreted the best-performing model in terms of variable importance and average local effects. The results support some of the factors mentioned in existing literature and provide insights that may help refine existing theories and inform further hypotheses to test.

The third article, which was co-authored by John Jerrim, is called 'Professional Learning Communities and Teacher Outcomes. A Cross-National Analysis'. Previous studies into the outcomes for teachers engaged in PLCs have produced inconsistent results, but have done so on the basis of very different measures of PLCs and in various contexts. This raises the question of whether these inconsistent results are due to differences in measures, contextual influences, or something else entirely. We approach this question by analyzing and comparing the relationship between PLCs and teachers' job satisfaction, self-efficacy and clarity of instruction across 40 countries, using internationally comparable measures, which is unique to this study. We analyze the relationship between these teacher outcomes and the overall PLC measure as well as the three subdimensions separately. Additionally, we assess whether teachers' gender and experience moderate these relationships. We find that there is a quite robust and positive relationship between PLCs and teachers' job satisfaction, but that the relationship with teachers' self-efficacy and clarity of instruction is mostly weak, often insignificant and varies more across countries. Our results indicate that the context does, at least to some extent, influence the relationship between engagement in PLCs and expected positive outcomes for teachers, especially for teachers' self-efficacy and clarity of instruction.

Resumé (Danish summary)

Professionelle Læringsfællesskaber (PLF), på engelsk 'professional learning communities', har de seneste årtier været blandt de mest populære måder at strukturere lærernes faglige udvikling. Kort fortalt er PLF en måde systematisere et løbende samarbejde blandt lærere inden for deres skoler, med det formål at forbedre elevernes læring. Med dette formål for øje, samarbejder lærerne løbende om at planlægge, evaluere og forbedre deres undervisningspraksis, samt deler ansvaret for at eleverne lærer det, som er forventet. PLF er blevet implementeret i mange lande på tværs af kloden, med forventninger om øget læringsudbytte for eleverne og gavnlige effekter på lærerne, for eksempel øget jobtilfredshed og self-efficacy[2]. Den forskning, der primært citeres for at understøtte disse forventninger, kommer i overvejende grad fra England og USA, og en række nyere studier har givet blandede resultater. Desuden er der markante forskelle på, hvordan PLF er blevet operationaliseret i den empiriske forskning, hvilket gør det vanskeligt at sammenligne resultater på tværs af studier. På trods af store investeringer i at implementere PLF internationalt er det empiriske grundlag for de forventede effekter af PLF begrænset, særligt i forhold til den rolle, som konteksten kan have for de forventede effekter, både i forhold til den enkelte skole og i relation til forskelle på tværs af lande eller uddannelsessystemer. Dette rejser et vigtigt spørgsmål om hvorvidt den store mængde af tid lærere bruger i deres respektive PLF på tværs af lande, skoler, klassetrin og fag fører til de forventede resultater.

Denne afhandling har til formål at bidrage med mere viden om effekterne af PLF på tværs af kontekster ved at anvende data fra Teaching and Learning International Study (TALIS) 2018 til at udvikle et internationalt sammenligneligt mål for PLF. Dette mål

[2] Self-efficacy oversættes i nogle sammenhænge til *mestringsforventning* på dansk.

for PLF danner grundlag for en eksplorativ undersøgelse af, hvordan kontekstuelle faktorer knyttet til skolerne relaterer sig til forekomsten af PLF. Ligeledes anvendes PLF målet i en tværnational analyse af forholdet mellem PLF og lærernes jobtilfredshed, self-efficacy og deres undervisningspraksis. Hovedelementerne i afhandlingen er tre artikler, som vil blive skitseret i det følgende, samt et indledende kapitel. Det indledende kapitel placerer de tre artikler inden for det generelle forskningsområde og uddyber og diskuterer nogle vigtige aspekter inden for og på tværs af de tre artikler, samt artiklernes begrænsninger. Endelig opsummeres og diskuteres resultaterne i relation til eksisterende forskning og i forhold til potentielle retninger for yderligere forskning.

Den første artikel, 'A Global Measure of Professional Learning Communities', omhandler udviklingen af et mål for PLF, som skabes på baggrund af data fra TALIS 2018-studiet. Ved hjælp af faktoranalyse (multilevel confirmatory factor analysis) skaber jeg et mål for PLF på skoleniveau ved baseret på lærernes besvarelser af et omfattende spørgeskema. Målet består af en andenordens-faktor af PLF samt tre underdimensioner: Samarbejdspraksis med fokus på elevers læring (CPL), Fælles vision og ansvar (SVR) og Støttende rammer (SC). Det overordnede mål for PLF samt de tre underdimensioner er egnede til sammenligninger på tværs af 42 lande og matcher kriterierne og estimeringsteknikkerne for de officielle TALIS-skalaer. Disse mål er tilgængelige for andre forskere, som har interesse for PLF, enten i international sammenlignende analyse eller inden for specifikke lande. De mål, der er udviklet i denne artikel, anvendes i analyserne i de følgende to artikler.

Den anden artikel, 'Exploring School Factors Related to Professional Learning Communities: A Machine Learning Approach Using Cross-National Data', er skrevet i samarbejde med Kristoffer Laigaard Nielbo og Sedat Gümüş. Relativt få studier har fokuseret på indflydelsen af skolekarakteristika i relation til velfungerende PLF. De fleste relevante studier har fokuseret på indflydelsen af forskellige ledelsesmodeller, og på tværs af disse

studier bliver de specifikke modeller fremhævet som nødvendige på tilsyneladende uforenelige måder. Da der var få studier tilgængelige, valgte vi at gå eksplorativt til værks ved at udforske sammenhængene mellem PLF og de mange forskellige skolefaktorer, som indgår i skoleleder-spørgeskemaet i TALIS 2018. Dette spørgeskema belyser mange forskellige aspekter af skolerne, såsom autonomi, forskellige ressourcer, skole-, personale- og elevkarakteristika, og skolelederens værdier og adfærd. Vi ønskede at udforske hvilke af de mange skolefaktorer der bedst forudsiger forekomsten af PLF på skolen. Vi anvendte maskinlæring (machine learning) som en metode til at udforske det kombinerede internationale data bestående af mere end 8000 skoler fra 42 lande. Først sammenlignede vi den prædiktive ydeevne af fem forskellige maskinlæringsalgoritmer, for at identificere den bedste model. Denne model fortolkede vi i forhold til, hvilke variabler der havde den største forklaringskraft, og ved at analysere, hvordan disse variabler specifikt relaterede sig til forekomsten af PLF. Resultaterne understøtter nogle af de faktorer, der nævnes i den eksisterende litteratur, og giver indsigter, der kan hjælpe med at nuancere eksisterende teorier. Ligeledes identificerede vi på baggrund af resultaterne konkrete forslag til nye hypoteser, som kan danne grundlag for videre forskning.

Den tredje artikel, som er skrevet sammen med John Jerrim, har titlen ’Professional Learning Communities and Teacher Outcomes. A Cross-National Analysis’. A cross-national analysis.’ Tidligere undersøgelser af effekten af PLF på forskellige positive effekter for lærere har produceret blandede resultater, men har også anvendt meget forskellige operationaliseringer af PLF og er ligeledes udført i forskellige kontekster. Dette rejser et spørgsmål om, hvorvidt disse blandede resultater skyldes forskelle i de måleinstrumenter, der er anvendt, kontekstuelle indflydelser eller noget helt andet. Vi forsøger at komme nærmere et svar på dette spørgsmål ved at analysere og sammenligne forholdet mellem PLF og lærernes jobtilfredshed, self-efficacy og ’klarhed i undervisningen’ (clarity of instruction) på

tværs af 40 lande. Vi analyserer forholdet mellem disse lærerresultater og det samlede PLF-mål, såvel som de tre underdimensioner af PLF separat. Derudover undersøger vi, om lærernes køn og undervisningserfaring modererer disse forhold. Vores resultater viser blandt andet, at der er en temmelig robust og positiv sammenhæng mellem forekomsten af PLF og lærernes jobtilfredshed, mens forholdene til lærernes self-efficacy og undervisningspraksis oftest er svage og ikke altid statistisk signifikante, samt at disse forhold varierer mere på tværs af landene. Vores resultater tyder på, at konteksten i hvert fald i nogen grad har betydning for forholdet mellem lærernes engagement i PLF og de forventede, positive resultater for lærere, især når det gælder deres self-efficacy og deres undervisningspraksis.

Acknowledgements

First and foremost, it has been a pleasure working with Sedat Gümüş and Felix Weiss, who have been my main supervisors, and who were there to provide valuable advice and feedback every time I was in doubt. And to my co-supervisor Kristoffer Laigaard Nielbo for supervising and guiding the approach to applying machine learning and supporting my development of the skills to implement these methods.

I would also like to thank all my great colleagues at Campus Aarhus for frequent informal meetings around the coffee machine and the possibility to engage in both social activities and professional discussions. Special gratitude goes to my peer group, Astrid, Elisabeth, and Charlotte. Our many meetings and the possibilities to discuss issues, concerns, and progress with you have been invaluable, and I am certain that the last three years would have been both boring and lonely without you!

I am thankful for the opportunity to visit and collaborate with John Jerrim at the Social Research Institute, UCL London, and to the group of PhD students that I shared an office with during the spring of 2023, for giving me a warm welcome. Likewise, I am grateful to Sedat Gümüş, for the chance to visit and present my work at the Education University of Hong Kong.

Finally, I am grateful for the support of my family through ups and downs over the last three years. To my wife Stine and daughter Marie for supporting me in working odd hours, attending conferences abroad, and not least for the great time we shared visiting UCL London together. To Rita, the strongest person I know, for keeping the family together through the tragedies we have faced over the last couple of years. I dedicate this thesis to my brother and best friend, Morten, who we lost way too early.

Chapter 1

Introduction

Written by Anders Astrup Christensen

The reference style is APA 7th.

Abstract

This is an article-based PhD dissertation, including three articles that are presented in chapters 2 to 4. This introductory chapter provides an overview of the general field of research and recent developments that the thesis is positioned within, the overall problem statement and the objectives and contributions of each of the three articles. Additionally, this chapter provides a more thorough discussion and explanation of some choices made within each of the articles, as well as a discussion of the strengths and limitations of the approach taken. Finally, the results throughout the three research articles are summarized, leading to a discussion of how the combined results of this thesis contribute to the existing research, and potential directions for further research on professional learning communities.

Teacher professional development

A school can be defined narrowly as ‘An establishment or institution for the formal education of children or young people’ (Oxford English Dictionary, 2024). What such formal education entails, and how it should be achieved is an ongoing historical and philosophical debate, shaped by external developments over time, covering aspects such as qualification, socialization, and subjectification (Biesta, 2009). What I suppose is easier to agree upon is that teachers are crucial actors within schools, whatever the intention of schooling may be. This applies in terms of what specifically it is that students learn about, how it is structured and approached, and the depth of the understanding or skills that students acquire from their interaction with the teacher and their classmates. This section will outline how and why teachers’ skills and practices are considered important, and how the means for developing such skills have changed in recent times, which has led to the international popularity of Professional Learning Communities.

Decades of research on ‘educational effectiveness’ and ‘teacher effects’ has confirmed the importance of the classroom level in explaining variation in student achievement, and that much of this variation can be attributed to individual teachers and what they do in their classrooms (Jackson et al., 2014; Muijs et al., 2014; Scheerens, 2016). Traditionally, practical approaches to teacher professional development (TPD) have focused on teachers learning about specific, transferable skills or best practices, through courses, workshops, or programs. The expectation was that teachers would implement these skills and practices in their respective classrooms, and that learning outcomes would thereby improve. Such initiatives have generally been found to be insufficient because they are too isolated from the school and classroom realities to have the expected impact, and because they fail to take teachers’ existing beliefs, attitudes and knowledge into account (Buczynski & Hansen, 2010; Garet et al., 2001; Little, 1993). As a response to the realization that schools were unable to adapt in a timely way to the rapidly changing demands of students’ knowledge and skills, the focus shifted towards developing teachers’ professionalism and skills within schools, by restructuring schools into learning organizations. As a result, the focus of TPD shifted from one of instrumental transmission of ideas and effective practices, towards a focus on continuous development of teachers’ professionalism within the context of their schools, inspired by developments of effective organizations and organizational learning in business and industry (Kools & Stoll, 2016; Supovitz & Turner, 2000). Influential international actors in education policy, such as the Organisation for Economic Co-operation and Development (OECD), stress the importance of continuously developing education systems, including the professionalism of teachers, to keep up with rapid changes in societies. In a report focused on international perspectives on teaching excellence, OECD suggested that:

> "Today's teachers need to prepare students for jobs that have not yet been created, to use technologies that have not yet been invented, and to solve social problems that haven't arisen before … Given the rapid changes in education, the potentially long careers that many teachers have, and the need for updating skills, teachers' development must be viewed in terms of lifelong learning, with initial teacher education conceived as providing the foundation for ongoing learning, rather than producing ready-made professionals" (Schleicher, 2016, p. 9).

Similarly, the United Nations Educational, Scientific and Cultural Organization (UNESCO) recognizes teachers as key agents of change towards establishing flexible and adaptive education systems globally, and specifically emphasizes the importance of teachers' continuous development as professionals (UNESCO, 2022b). PLCs, which this thesis revolves around, are among the most popular approaches to restructuring schools into learning organizations.

Professional Learning Communities

Across books and research articles, there are many definitions of PLCs available, and although they share common characteristics, many researchers agree that the concept is lacking clear definition/operationalization, which causes confusion (Doğan & Adams, 2018; Kruse & Johnson, 2016; Watson, 2014). There is broad agreement that the first systematic description of the concept should be attributed to Shirley Hord in the late 1990s, and more specifically the publication 'Professional Learning Communities: Communities of Continuous Inquiry and Improvement' (Hargreaves, 2019; Harris et al., 2017; Hord, 1997). Since then, PLCs have received much attention in school research internationally. However, disagreement has also developed around the term, as it has been used to describe many

different things, leading to a risk of the concept losing its the original meaning (DuFour, 2004). In relation to this, Hairon et al. (2017) argue that, despite the large number of studies on PLCs, there are still gaps in the theoretical foundation of the concept, which may be partly due to the fact that each of the three words that make up the concept (professional, learning and community) are contested to some degree in terms of how they should be understood, which adds to the complexity of the concepts and makes it difficult to agree on a universal definition. This means that researchers face a choice between many different sources and definitions of PLCs when they describe the concept and how they intend to study or operationalize it. Drawing on examples from some of the most influential researchers in the PLC literature, definitions or conceptualizations include one-sentence definitions, such as: '*a group of people sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way; operating as an enterprise'* (Stoll et al., 2006, p. 223)[3], or *'An effective professional learning community has the capacity to promote and sustain the learning of all professionals in the school community with the collective purpose of enhancing pupil learning'* (Bolam et al., 2005, p. 145). Others introduce the concept of PLCs through more elaborate descriptions of the processes and goals of effective PLCs; see for example the 'Three Big Ideas That Drive the PLC Process' by DuFour and Marzano (2011), which is outlined in Chapter 4, or the three elements comprising PLCs proposed by Hargreaves and Fullan (2012), which are presented in Chapter 2.

In empirical research, PLCs are most frequently defined through important characteristics or dimensions, which are also used as the basis for operationalizing and measuring them.

[3] In the source cited, there are citations within the quote to support the argument, which has been removed in his citation for simplicity.

Although there is variation in the specific labels and descriptions used, as well as the number of characteristics (Hord et al., 2008; Moosa et al., 2022; Vangrieken et al., 2017), many include the following:

- Supportive and shared leadership
- Shared values, vision, and goals
- Collective learning and application
- Shared individual practice
- Supportive conditions

The descriptions of, and the items used to operationalize these dimensions and characteristics differ considerably across research articles. Chapter 2 includes a section outlining some PLC measures with examples that illustrate the vast differences in the characteristics/dimensions that are applied in quantitative research on PLCs. In a review of PLC instruments, Lee et al. (2022) conclude that only three dimensions are featured throughout all of the eleven applied instruments that they include in their analysis:

> 1) shared vision, values, and/or norms among school staff as a "community," 2) focus on both improved student "learning" as an aim and teachers' continuous "learning" as an engine of PLCs, and (3) "collaboration" as a way to form and sustain PLCs' (p. 277).

In addition to the confusion about how to define and, as a consequence, measure PLCs, there is disagreement about who the legitimate members of the PLC are: is it solely the teachers, does it include the leadership staff, or all staff in schools (Zhang et al., 2020). Can students, parents, school boards and the surrounding community be considered members of the PLC, and does it involve collaboration between schools in the district, or even with knowledge bases in other countries? According to some of the influential contributors to the PLC

literature, these questions do not have a clear answer, although they do emphasize that within-school PLC and the direct focus on improving student learning is most essential (Stoll & Louis, 2007).

Preceding and related concepts

In educational research there are other concepts with varying degrees of overlap with the idea and practice of PLCs, including: Professional Community, Teacher Community, Communities of Practice, Professional Development Community, Critical Friends Groups and Teacher Networks (Curry, 2008; Hadar & Brody, 2013; Lomos et al., 2011; Louis et al., 1996; Pennell & Firestone, 1996; Vangrieken et al., 2017). Some of the central ideas of PLCs, and of the aforementioned related concepts, can be traced to earlier influential theories and developments in education and in the field of organizational learning. The following presents an overview of these foundations.

'Team teaching', which gained popularity in the 1960s and 1970s, emphasizes the importance of instructional teamwork by teachers with the aim of increasing school efficiency. The idea that it is beneficial to de-privatize classroom teaching and collaborate in planning as well as in practice thereby by far precedes the concept of PLCs (Buckley, 2000; Dean & Witherspoon, 1962). The inquiry into teacher practice by teachers as a community, which is an important part of the PLC process, has much in common with the concept of action research. In the literature regarding teachers in schools as researchers, Lawrence Stenhouse stands out as a central figure, advocating the process of systematic inquiry performed by the insiders of a community, and arguing that teachers should systematically, critically and cyclically investigate their own practice, in order to produce context-specific knowledge which enables them to learn together (Nigel & John, 2012; Stenhouse, 1975). Donald A. Schön's book *The Reflective Practitioner* describes how practitioners, as they gain experience, produce tacit knowledge; knowledge that teachers have acquired, but which they

are not able to talk about. Schön maintains that it is important for the development of practitioners (for example teachers) that they reflect upon their practice with other colleagues. The reflective practitioner gains insights through this reflection on their practice, with their colleagues, with a potential to develop and improve practice (Carlgren et al., 1994; Schön, 1992). The ambition of developing context-specific knowledge about practice in schools is thus related to both action research and the idea of the reflective practitioner.

In Hord's initial description of PLCs, Peter Senge's portrayal of 'The Learning Organisation' is presented as an important inspiration. Senge's thoughts on shared visions, collective engagement and problem solving originate in the business sector but were adapted to the education sector under the label learning communities. The focus on continuously learning together to expand capacity is emphasized by Hord (1997), who is inspired by the idea of a learning organization *'where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together'* (Senge, 2006, p. 3).

Lastly, an important related concept is that of communities of practice or situated learning in communities of practice, as presented by Jean Lave and Etienne Wenger (Lave, 1991; Wenger, 1998). Lave and Wenger highlight the social aspect of learning and posit that learning and knowledge cannot be separated from the community of practice in which they take place. While there are many similarities between the concepts of PLCs and communities of practice, the literature on PLCs is generally more explicit in its focus on increasing student learning as the main purpose of collaboration. There is also a stronger emphasis on the organizational level in PLCs in terms of creating a school-wide culture of collaboration, and on the role of leadership in creating and sustaining this collaborative culture (Blankenship & Ruona, 2007; DuFour et al., 2016).

Research on the effects of professional learning communities

As explained in the introduction, my initial interest in PLCs comes from the lack of evidence and planned studies into the expected outcomes of PLCs in the Danish context, at a time when significant investments have been made into restructuring schools into PLCs. Throughout the literature, alongside claims that PLCs should be implemented because of strong/convincing/numerous sources underlining their positive effects on student learning and teacher, it is just as easy to find researchers arguing that we still need to know more about the effects of PLCs. In the seminal work of Shirley Hord (1997), she reviews various literature and research findings that collectively relate to large numbers of positive outcomes for both teachers and students. Hord integrates these theories and connections from research into PLCs, and summarizes that: "*The collection of research studies cited in this review clearly identifies the power of the organized professional learning community that makes possible the advancement of student achievement"* (p. 25). Then, about 20 years later, Hord et al. (2008) stated that:

> "Interestingly, while the PLC has been touted as a significant and effective school improvement strategy or structure, it has been characterized in endless ways, depending on who defines it … One explanation is that while there has been much talk about the importance of PLCs, little attention has been given to the research studies that have investigated what it is and what outcomes it can produce" (p. 8).

Having read many studies and reviews related to PLCs during my work on this thesis, although without having done a systematic review, it is my impression that the majority of them find some positive relationships between various dimensions of PLCs on outcomes for both students and teachers, but often not for all PLC dimensions included in the respective

studies. What seems less certain is whether the available studies, with their inherent differences in conceptualization, warrant the strong claims that are sometimes used to advocate their importance/effects for students and teachers, much like Hord acknowledges in the quote above. Chapter 4 gives an example of how previous studies on the links between PLCs and teachers' job satisfaction, self-efficacy and instructional practices yield inconsistent results, and lack details about how specifically PLCs affect teachers and their practices.

Problem statement and research objectives

International or cross-national research related to PLCs is very limited, and, to my knowledge, there are no quantitative studies that compare any effects of PLCs for students or teachers across contexts. The few studies that include data from multiple contexts focus on qualitative comparisons between countries, on differences in the prevalence of PLCs, or on differences in how teachers engage in them, or similar concepts (Lomos, 2017; Vieluf et al., 2012; Webb et al., 2009). Arguably, there are a number of reasons why cross-national research into teacher collaboration and the associated outcomes can be important: most of the theories and concepts have been developed in the western world, and may not be easily transferrable to other contexts; the effects of inputs and processes related to teachers' professional development may be moderated by the characteristics of the education systems in different countries; philosophies of education, pedagogical traditions and teaching practices differ across countries, as do levels of tracking and teacher autonomy (Guo & Wang, 2021; OECD, 2020; Vieluf et al., 2012). The available empirical research on PLCs, and especially the empirical studies on the relationship with student learning, is dominated by studies from the United States and the United Kingdom (Bolam et al., 2005; Stoll & Louis, 2007), but more recently there has been a growing interest internationally (Bellibas et al., 2016; Qiao et al., 2018; Sleegers et al., 2013; Valckx et al., 2020). The potential influence of

such cross-national differences on the expected effects of PLCs seems to have received little to no attention, as PLCs have been implemented across the globe during recent decades, drawing on the same body of literature.

Hence, a number of important questions related to PLCs remain unanswered and need further investigation. Hairon et al. (2017) proposed a research agenda to advance the knowledge on PLCs, which they argue has substantial gaps despite a large number of studies on the subject being conducted over the course of the last thirty years. After reviewing the current state of the PLC literature, Hairon et al. highlight three major gaps, which they understand as interdependent, and which should be investigated in future studies in order to advance our knowledge of PLCs: 1) The construct of professional learning communities, 2) Conditions-contexts of professional learning communities, and 3) Causalities of professional learning communities.

Firstly, on the basis of the vast differences in definitions and measures applied in PLC research (similar to what is outlined in the previous section), Hairon et al. (2017) argue that there is a need for dialectical work between theorization and operationalization. Such work should focus on establishing an agreement on what the substantial essence of the construct is, and on how that can be operationalized. Secondly, Hairon et al. argue that there is 'much room' for studies investigating the relationship between PLCs and the context in which they are enacted. Context is broadly defined as covering aspects within the school (leadership, management, structures, resources), and outside the school (district/system, societal and national culture and policies). They maintain that further research is needed to understand how PLCs may be differently enacted in different settings:

> 'a PLC is shaped by the context of the school, the school is shaped by the context of the district (or county), the district is shaped by the context of the

education system, and the education system is shaped by the context of the national system' (pp. 77-78).

Finally, what the authors discussion of the causalities of PLCs pertains to research on outcomes or effects of PLCs for teachers and students. They argue that the available research on these matters was not withstanding the 'international claims of positive effects of PLCs', and consequently that future studies need to invest more in testing the hypothesis that PLCs have a positive impact on student learning through their impact on the development of teachers' knowledge, skills, practices, beliefs, self-efficacy and commitment (Hairon et al., 2017).

Research objectives and structure of this thesis

From the literature outlined above, it seems that the gap I found in the Danish context can be identified as a larger issue with significance for other countries and the general field of research on PLCs. With the ambition to contribute to this lack of international research on the subject, my idea was to use data from international large-scale studies to contribute to comparative research on PLCs and their relationship with other educational aspects, such as the context in which they function, leadership factors and outcomes for teachers and students. As such, this thesis interacts with each of the three gaps formulated by Hairon et al. (2017) throughout the three articles, although it does not meet their request for PLC research to employ mixed methods, experimental and longitudinal approaches. The following section outlines the ambition behind each of the articles in addressing these gaps in the research, the arguments that explain some of the choices made, and a discussion of their inherent strengths and limitations. The ambitions of this thesis can be broadly summarized as:

1) To develop a measure of PLCs that enables international and cross-national research.
2) To explore the school contextual factors related to PLCs using international data.

3) To investigate the relationship between PLCs and various teacher outcomes across different national contexts.

Chapter 2 contains the research article which takes on the challenge of the first ambition. Having realized that PLCs are measured and defined in an abundance of ways already, I was reluctant to suggest yet another measure and thereby add to the complexity and confusion that already existed. I did, however, come to the realization that using existing international studies was the most feasible way to contribute to cross-national research on PLCs within the financial and time constraints of a PhD project. At the same time, the publicly available data from such studies, and the wide variety of educational topics they cover, enables other researchers with an interest in PLCs to use such a measure in their research in relation to a variety of educational aspects in their own context or across contexts.

Having reviewed the available large-scale studies in terms of relevant questionnaire items that relate to the important aspects of PLCs, I was convinced that the Teaching and Learning International Study (TALIS) 2018 had the greatest potential for the development of a comprehensive measure. While other large-scale studies have the advantage of including measures of student achievement, which could advance research into the effects of PLCs further, I found that these studies only had a few PLC-related questionnaire items. A more detailed discussion of the possibilities and limitations of using cross-sectional large-scale education studies is provided in the methodology section.

Due to the many misconceptions about the concept and the label of PLC being used to describe very different things (DuFour, 2004; Hargreaves & Fullan, 2012), I found it best to rely on teachers' reports regarding how often they engage in PLC-related activities, and how they perceive the collaborative culture and support for engaging in professional development activities, etc. This approach was chosen as even if a school principal indicates

that a school is functioning as a PLC, this does not necessarily say much about the perceptions and the day-to-day interactions between the teachers, who are the core actors driving the PLC process. I argue that a better way to determine if a school is a PLC is to rely on teachers' reports about the frequency of their engagement in related activities and their perceptions about the culture and shared beliefs in the school. Having decided that the measure of PLC should be evaluated at the school level, and that the best measurement items were given at the teacher level, I found inspiration in the work of Veletić and Olsen (2021), who developed a 'shared cluster construct' of instructional leadership in TALIS 2018. The measure developed includes a second-order construct of PLCs, and three subdimensions: Collaborative practice with focus on student learning (CPL), Shared vision and responsibilities (SVR), and Supportive conditions (SC). A fundamental concern in developing any latent measure or scale is that of validity. The most straightforward aspect of validity in this case is measurement validity: The measure developed meets common standards of fit statistics and measurement invariance across the countries included, which means comparisons of estimates across countries can be justified. In this sense, I believe that the properties of the scales developed are comparable to the official scales included in the official TALIS 2018 data with metric invariance, which is the level achieved by the majority of the official scales (22 out of 31 have metric invariance) (OECD, 2019c). It is much more difficult to determine the construct validity of the measure developed; the degree to which the dimensions measure the most important aspects of PLCs, and only this. This question is discussed in the article, pointing towards the lack of information about the use of data in evaluating student learning, and differences in how the PLC characteristics are grouped together. Given the general conceptual confusion, others might identify entirely different limitations of this measure. I will not in any way claim that the perfect measure of PLCs was miraculously derived from my secondary analysis. Rather, my question would be whether

this measure is substantially worse, or might be as good as many of the other measures applied, to which the effectiveness of PLCs is currently ascribed. This question must be for others to answer, but the measure arguably covers common aspects included in such measures (Lee et al., 2022), achieved metric invariance across a wide range of countries, and can be applied in research related to the many other aspects covered by the TALIS 2018 study.

The second article of the thesis, which was co-authored by Kristoffer Laigaard Nielbo and Sedat Gümüş, presented in chapter 3, addresses the questions related to the contexts in which PLCs are applied in two ways: 1) By exploring and ranking the importance of a wide variety of contextual factors in terms of their statistical relationship with PLCs (for example leadership models and autonomy, resources, school-, staff- and student characteristics), and 2) challenging the expectations of the universality of PLCs, which is implicit in the way that PLCs have been applied globally with reference to the same literature, which is developed in Western countries, predominantly the United States. As argued by Hairon et al. (2017) and Vieluf et al. (2012), there is a need for further studies of the contexts in which PLCs are applied, which may shape their functioning at different levels. Given the limited number of studies of the relationships between such contextual factors and PLCs, leaving us with little knowledge to form hypotheses to test, and because we had a relatively large dataset with many potentially important predictors accompanying our PLC measure, we decided to conduct an explorative study. We chose this explorative approach to evaluate if the factors described in the few studies available would come out as important among the many other potentially important factors covered in the TALIS study, and to test those assumptions across the international dataset covering schools and teachers from around the globe. As mentioned in the previous section, the PLC measures were created based on the answers given by the teachers in the school, and in this analysis, we explore the relationship between these collective teacher measures and various aspects about the school as described

by the principals. It is important to note that we do not know if schools deliberately attempt to act as PLCs, but what this analysis sheds light on is which of the many potential predictors most frequently occur in the schools where teachers agree that there is a high degree of PLC-related activities and a collaborative culture. Furthermore, the computational methods we applied for this explorative analysis are suitable for detecting nonlinear relationships and interactions, with the potential to support or refine existing findings, and to generate hypotheses for further research into the contextual factors that may support or hinder the prevalence of well-functioning PLCs. The specific methodological approach in this article is discussed in detail in the following section.

The third article, which was co-authored with John Jerrim, presented in chapter 4, focuses on the expected outcomes of PLCs for the teachers engaged in them, and contributes with comparisons of these outcomes across a range of different contexts. Besides increasing student learning, the literature frequently highlights the ways in which engaging teachers in PLCs is expected to improve their job satisfaction and self-efficacy (Hord, 1997; Hord et al., 2008; Stoll et al., 2006). Within TALIS we do not have any direct measure of student achievement, but we do have teachers' reports about different aspects of their instructional practices, their sense of self-efficacy and job satisfaction. While self-reported measures of instructional practices are less ideal than direct measures of student achievement, any improvements in student learning as a result of PLCs would need to happen through changes in teachers' practices (Hudson, 2023). Within the limitations of cross-sectional data, this article contributes to the research gap related to the outcomes of PLCs for teachers (and indirectly for students) and to the context of PLCs, by comparing differences in these aspects across a variety of different countries.

Methodology

This section provides an overview of the philosophy of science, the methodologies and data sources used throughout this thesis, including a discussion of their strengths and limitations. The chapter addresses the approach taken in this thesis by elaborating and discussing the following themes: Philosophy of science; International Large-Scale Assessments in education; secondary analysis of cross-sectional data and causality; latent constructs; specific methods applied; and the approach to Open Science.

Philosophy of science

In this thesis, quantitative methods and international data were used to measure and analyze various relationships between the concept of PLCs, the contexts in which they are applied, and the teachers engaged in them. Consequently, two main questions arise within the philosophy of science; one related to the concept in question, namely that of PLCs, and another one for the specific methods applied to study PLCs. This section will briefly address how these perspectives differ, and what this means for the approach taken in this thesis.

PLCs and the inherent focus on continuous learning within a context, have ties to the constructivist (or constructionist) philosophy, emphasizing that the validity and relevance of knowledge or learning obtained is tied to the context in which it is socially generated. Learning is first and foremost considered to be a social interaction, although influenced and affected by external norms, rules, and regulations. The development of the teachers' professionalism and their practices and solutions to issues, in this sense, cannot be separated from the school context, and is perceived as a collective generation of knowledge which takes place through social interactions with the aim of improving instructional

practices and ultimately student learning (Borko, 2004; Joseph & Russell, 2011; Kools & Stoll, 2016; Vieluf et al., 2012).

Using quantitative methods to analyze data such as the ones available from the TALIS study is associated with a different perspective on philosophy of science. While organizations conducting these studies rarely address questions of philosophy in the study frameworks and technical reports, the very ambition of the studies, and the substantial efforts made to ensure valid and reliable data suitable for international comparisons, arguably suggests a position where an observable reality exists. For example, the great emphasis placed on testing, developing, translating and verifying items, and testing measurement invariances across country contexts, suggests that they believe that an objective reality exists, and that it can, at least to some degree, be systematically measured and compared across contexts and time (Ainley & Carstens, 2018; OECD, 2019c). This approach is aligned with (post-)positivism. Positivism is often associated with natural science, such as physics or biology, where the goal may be to uncover universalities of the world. Post-positivists share the ontological standpoint with positivists that an objective social reality exists, which researchers should strive to uncover. Post-positivism differs from positivism by being skeptical towards the possibility of fully observing and uncovering this reality, due to the complexity of the world and the potential bias or inherent values of the researcher (Joseph & Russell, 2011). Within post-positivism, there is a distinction between a realist and a constructivist approach, in which the former:

> … adheres to the notion that there is some objective reality to the social world, while acknowledging that the Schutzian analysis of social science as interpretative and therefore ultimately subjective sense-making, precludes the discovery of that reality once-and-for-all. All that can be achieved is the aspiration to knowledge through rigour, multiple data analysis and theory-building and testing (Fox, 2008, p. 5).

In the PLC literature, the collaborative practice and development of instructional quality is closely linked to the context and the progression of the students within the school, but the expected positive outcomes for students and teachers are often credited to studies using quantitative methods (Bolam et al., 2005; Lomos et al., 2011). While this may seem contradictory, the use of data to collectively monitor student learning is an important part of the PLC process. Data in this sense is broadly defined, and can cover anything from standardized tests to teachers' qualitative assessments of student products and formative assessments of student learning, but should in any case be suitable for evaluating whether the students have learned what was intended, and is as such measurable (DuFour & Marzano, 2011). I take the stance of realist post-positivism in the analyses carried out in this thesis, studying the relationships between PLCs, the contextual factors, and the outcomes for teachers who are engaged in this constructivist approach. This entails that I strive towards understanding the observable reality of these processes within schools, while acknowledging that such processes are extremely complex, and are unlikely to be explained to their full extent.

International Large-Scale Assessments in education

Since the 1960s the number of International Large-Scale Assessments (ILSAs), as well as the number of countries participating in them, has increased significantly. Today, such studies cover a variety of aspects that they seek to compare, for instance: achievement in specific school subjects such as reading, mathematics and science; civic and citizenship education; teaching and learning; computer and information literacy and adult competencies (IEA, 2017; Wagemaker, 2020a). There are a number of reasons for countries to participate in International Large-Scale Assessments (ILSAs), including the wish to monitor the progression in educational outcomes, and to be able to compare relationships within their

own countries, with the corresponding relationship in other countries, as well as tracking changes over time for the various subjects and stages of education that are covered by different studies, and the reason for participation may differ between countries (Hernández-Torrano & Courtney, 2021). The ambition of these studies is to provide policymakers and researchers with insights on factors associated with the quality of teaching and learning, which are comparable across the participating countries and over time, while acknowledging the challenge involved in ensuring comparability, validity and reliability across differences in culture, language, economic and educational development (Wagemaker, 2020b). In public and political debate, ILSA studies are often mentioned when international results are published, where attention is often on the international rankings of the studies measuring student achievement. In these rankings the participating countries are ordered by the average student achievement, and changes in such ranks, or the introduction of new participants, has been the source of intense public debate in various countries over time. For instance, what is sometimes referred to as a 'PISA shock', describes how the placement of a country in such international rankings in some cases has been disappointing or not in line with expectations, leading to intense public debate and sometimes actual educational policy reforms. In the case of rankings, the OECD's Programme for International Student Assessment (PISA) has been the source of most debate, leading to reforms in countries such as Germany, Japan, Norway, Spain and Australia, and even to placements in such rankings to be included in policy evaluations or goals, in ways which do not always align with the intention and scope of these studies (Choi & Jerrim, 2016; Rutkowski et al., 2020).

Besides the substantial interest from policymakers and in the public debate, ILSA studies have provided the basis for a large field of research using the publicly available data in secondary analysis within and across the participating countries. These studies are now widely applied across different fields of research such as sociology, economics,

management and psychology (Hernández-Torrano & Courtney, 2021; Veletić et al., 2024). These studies, and the influence they have, have also been critiqued in different ways and by various actors (Rutkowski et al., 2020). For instance, educational theorists and philosophers of education argue that the content of these studies, and the influence that their results has on policymakers, shapes the debate of the purpose of education to focus almost exclusively on what can be measured and compared across countries, as opposed to what may be valued or desirable in the individual societies: '*The danger here is that we end up valuing what is measured, rather than that we engage in measurement of what we value*' (Biesta, 2009). Other academics have investigated and discussed the methodological and technical details and difficulties these studies face in meeting their ambition of providing data that is comparable across countries (Anders et al., 2021; Zieger et al., 2022).

The Teaching and Learning International Survey

TALIS is an OECD study collecting data from representative samples of teachers and principals across a wide range of countries and economies. The study covers aspects such as school and classroom contexts, working conditions, beliefs, attitudes and motivations towards teaching and learning, and is repeated at intervals of five years to enable the study of trends over time and across countries. The study was completed in 2008, 2013 and 2018, and across these study cycles, the number of participating countries and themes covered has increased. According to OECD, the TALIS study has three main purposes:

> "The first is to describe teaching and learning conditions; the second is to identify the relationships among components of those conditions. The third is to identify and describe how teaching and learning conditions and relationships vary within and across TALIS participants and over time" (Ainley & Carstens, 2018, p. 11).

Like other ILSA studies, TALIS involves a complex sampling design, which requires researchers to perform secondary analysis of the data to apply sampling and replicate weights to obtain correct results. TALIS differs from most of the other popular ILSA studies by not measuring skills or achievement such as those often presented and debated in international rankings. Instead, it provides the teachers' and principals' views on the teaching and learning contexts across different contexts and over time. The number of research articles that conduct secondary analysis of TALIS data has grown over time, although not as fast and across fewer disciplines than those analyzing the PISA study (Hernández-Torrano & Courtney, 2021; Veletić et al., 2024).

Secondary analysis of cross-sectional data and the question of causality

The use of already existing data, often not collected by the researcher using it, and which may have been initially collected for a different purpose, is generally referred to as secondary analysis or 'analysis of secondary data'. In the case of TALIS, it is a declared goal of the study to enable the use of the collected data in analyses beyond what is presented in the international reports, and guidelines for such secondary analysis are provided (OECD, 2019a). More generally, among the benefits of using secondary data are the possibilities to study various phenomena on a large scale, with substantially reduced investment in terms of time and money, than would otherwise be possible within for example a PhD project. In addition, the fact that some official and large datasets carry a certain authority may also be reflected in secondary application of them. Drawbacks of using secondary data include the risk of the research undertaken being determined by what data is available, in opposition to researching what is actually relevant. Additionally, there is a risk that working with such data detached from the settings in which it is collected causes a lack of practical realism towards the studied phenomena (Gorard, 2003). In relation to this, the measure of PLCs developed in

chapter 2, which is then applied in the two other articles, is admittedly limited to some extent by the data available in the TALIS; for example, it is argued in the discussion of the article, that further items describing the teachers' use of data in evaluating student learning could provide the basis for a better coverage of the PLC practices. This thesis does however also benefit from the application of a large, reputable study with extensive data collected in many countries, applying validated questionnaire items on representative samples of teachers and principals, which by far exceeds what could be achieved by an independent data collection within a PhD project.

Most ILSA studies, including TALIS, are repeated cross-sectional studies. This means that the study is conducted repeatedly (every five years in the case of TALIS), and for each wave, a sample of the population within various national stratifications is randomly sampled to participate. These participants answer the extensive questionnaires at one point in time, which provides a snapshot of the various themes covered in the study at the time that the study is conducted. Among the benefits of repeated cross-sectional studies is the possibility of analyzing trends over time for a given population, which may occur due to changes in policies affecting learning environments or other relevant factors, and the possibility to compare such trends across countries. Cross-sectional data also has several limitations. First and foremost, such data is not suitable for establishing causal relationships, which could be possible with longitudinal data, meaning that data is collected from the same individuals or schools at different times. With cross-sectional data alone, it is in principle impossible to determine whether a strong link between a set of variables is because one causes the other as theory or logic may suggest, whether the relationship is spurious, or whether the causality is reversed. These issues related to causality in relation to the cross-sectional data from such studies have been acknowledged since early studies done by the International Association for the Evaluation of Educational Achievement in the 1960s and

1970s, but already at that time, the language used to describe the correlations that may be found in such data trended towards implying causality:

> The study documentation indicates that although the inability of testing causal hypothesis with cross-sectional data is acknowledged, nevertheless, associations are perceived at least as an indication of a possible effect and thus, the overall goal of the study is described as to extract "malleable factors," which "have substantial effect" on the outcomes of students (Stancel-Piątak & Schwippert, 2022, p. 159, referring to the Six-Subject Study and the First International Science Study) .

Different strategies and methods have been developed and applied over time in order to enhance the possibilities of establishing causal relationships within such datasets, for example difference-in-differences, instrumental-variables or regression-discontinuity designs, but the issues with establishing causal relationships within ILSA studies remain a challenge (Angrist & Pischke, 2009; Hanushek & Woessmann, 2014; Strietholt et al., 2014).

Consequently, many researchers using such data are careful not to use causal language at all when describing their research questions and when discussing their findings. In relation to this, Hernán (2018) argues that the way that researchers, sometimes at the request of editors, refrain from acknowledging the causal ambitions of their research, for example when notions of 'impact' or 'benefit' are routinely exchanged with non-causal terms such as 'correlation', 'pattern' or 'association', is problematic. For instance, without explicit references to causal effects, the ambitions of many studies relying on cross-sectional data can only be expressed ambiguously or in a roundabout way. Moreover, Hernán emphasizes that the methodologies suitable for establishing causality differ from those used in purely correlational or associative studies: If a study is solely interested in an association, according to Hernán, then no adjustment for confounding variables is necessary, arguing that confounding a causal concept that does not apply to associations, and that there is no such

thing as a ‘spurious association’. This means that using control variables to isolate a relationship between two variables is methodologically related to causal ambitions. Since many relevant research questions cannot be answered through randomized experiments in ethically justifiable ways, and because even such randomized experiments have limitations in establishing causation, Hernán (2018) advocates that the causal ambitions of studies using observational data should be clearly stated, and that using cross-sectional data should be acknowledged as the best approach for answering causal questions in some cases.

While I do acknowledge the inherent limitations of cross-sectional data, I tend to agree with the arguments above, and would like to see a shift towards acknowledging the underlying causal ambitions of such research. However, I find that the general convention in journals publishing quantitative studies in education is still to avoid causal language entirely. As such, I have accepted the current premise, and I have tried to avoid causal language in the interpretation of the results of the analyses of this thesis, by describing associations as for example ‘statistical relationships’, and ‘strengths of associations’ as opposed to signaling that ‘x has a strong effect on y’.

Latent constructs

The analyses of this thesis involve latent traits or constructs in several ways. In the first article, teacher questionnaire items are used in chapter 2 to develop three latent PLC dimensions, which then make up the higher order construct of PLC. These latent PLC constructs are then applied in different ways in the second and third articles. The latter article also utilizes three official latent constructs or scales from the TALIS study, related to teachers’ job satisfaction, self-efficacy, and instructional practices.

Generally, scales or latent constructs involve the use of multiple items or questions which are combined in order to measure a construct which cannot be directly

observed. Various frameworks for constructing and validating such latent constructs exist, and in the case of TALIS, the official scale scores are developed using confirmatory factor analysis (CFA). Items are designed or selected to reflect specific aspects of latent traits of interest, relying on inputs from research theories within relevant fields and experts in item and scale construction, as well as experiences from previous cycles of the study when available. These items are then administered in field trials, and the resulting data is used to evaluate the latent constructs and potentially make adjustments before the main data collection. On the basis of the main data, each scale undergoes a series of checks before each construct is modeled using CFA on first the full data sample (pooled data), and then modeled separately on the data for each country and ISCED population. Further steps involve measurement invariance testing across the population, which determines which statistical analyses and comparisons can be justified, and finally computing the scale scores to be included in the official data (OECD, 2019c). In developing the measure of PLC, I tried to mimic these steps from the TALIS technical report to the greatest extent possible. Besides the obvious difference that the items included in this PLC measure were not administered with the intent and theoretical underpinning to create a latent PLC construct, the main limitation was that, based on theoretical arguments, as described in the introduction, the PLC measure is a multi-level construct, in which teachers' responses are used to create measures at the school level. None of the official constructs in the TALIS data are multi-level, and this posed a few challenges, which are described in chapter 2; for example, the fact that the measurement invariance testing had to be done as a single-level CFA model. The specific approach to modeling PLC as a multilevel construct was theoretically motivated, and is aligned with the recommendations of Stapleton et al. (2016), who proposed various ways to conceptualize latent constructs in the presence of multiple levels, depending on the level of interest. In this case, the shared cluster construct was used, with the aim to reflect a construct at the cluster

level (in this case schools), since the items used to measure the construct was at the individual level (teacher level in this case), similar to the work of (Veletić & Olsen, 2021). Besides these differences and limitations, the development of the PLC measure follows the same basic strategy as applied in the validation of the existing scales, uses the same estimation technique, meets the same criteria for various fit indices, and reached the same level of measurement invariance as the majority of teachers scales included in the official data (OECD, 2019c).

Using Machine Learning for explorative analysis

The article presented in chapter 3 applies Machine Learning (ML) as the means for conducting an explorative analysis of how school contextual factors relate to PLCs, for reasons which are presented in the introduction. During our work on this article, and when presenting and discussing the work with colleagues, we received different kinds of criticism of this approach, which I will present and discuss in the following. Firstly, several colleagues pointed out that theory is built into the questionnaire of TALIS, meaning that the specific items administered in the TALIS questionnaires therefore have a theoretical purpose. In that sense, it is already determined which relationships should be found or investigated, raising the question of why or if this purposefully collected data should be subject to exploration. It is definitely true that there are theoretical arguments behind the themes covered by TALIS and the corresponding items included as measures (Ainley & Carstens, 2018). However, I am not convinced that this means that there is no potential for an explorative study of such a dataset. I believe that having hundreds of variables that in various ways are already considered important for school organization, teaching, and learning, may well be a benefit rather than a drawback, compared to exploring items that may not be related to the field of study at all. For instance, themes such as teachers' job satisfaction can be studied in relation

to characteristics of the student body, their colleagues or leader, the school system, teachers' salaries, or autonomy, to name a few possibilities. If this critique were to be taken to the extreme, it would mean that the conceptual framework for TALIS would offer a complete guideline for how teachers' job satisfaction could and should be analyzed. In reality, data from ILSA studies have been used more creatively, for example in the analysis of how teenagers' propensity to claim to have expertise and knowledge about fictitious mathematical concepts varies across countries, gender and socio-economic groups, by developing a 'bullshit scale' in PISA (Jerrim et al., 2019). A related issue pertains to what is included in the TALIS study, and therefore what it is possible to discover through an explorative study. The themes covered in TALIS are determined by the OECD. The large-scale studies in education conducted by the OECD and others have been criticized and questioned in relation to the power that these organization exert by defining what is included and measured in these studies, which influences debates and initiatives concerning education within countries (Ydesen, 2019).

Secondly, and perhaps more crucially, there is a critique that this approach leads to potentially meaningless conclusions as well as questionable research practices such as multiple testing and HARKing (Hypothesizing After the Results are Known). As acknowledged in the article itself, we believe that such approaches necessitate the researcher having subject-specific knowledge, and critically reflecting on whether the results of such exploratory analysis may be helpful in refining existing theories or in generating new theories and hypotheses to test, or if they should be disregarded as meaningless, possibly random correlations. It is important to acknowledge that doing exploratory analysis involves the risk of finding nothing of interest, but also holds potential for generating new insights and ideas, to be tested using other data and methods. Hypothesis testing, which is the common approach used in quantitative educational research, will lead to either a confirmation or a rejection of

an assumption based on theory. During hypothesis testing, various questionable research practices have been applied in attempts to produce statistically significant results, for example by altering data, questions and methods to obtain significant results, which are easier to publish (multiple testing), or simply by formulating hypotheses after the authors know the results (HARKing) (Andrade, 2021; Frias-Navarro et al., 2020). It is easy to see how supervised ML as applied in this thesis poses an opportunity to find strong relationships within a given dataset, which could then be used to form post-hoc hypotheses which are likely to be statistically significant, and thereby provide a better basis for writing a research article that is arguably more likely to get published. While these methods can provide a useful tool for engaging in such questionable practices, these issues precede the use of ML, and simpler strategies such as correlation matrices have been available since the early 20th century, as has the possibility of testing various model specifications and methods. These issues, for the individual researcher, can be reduced to one of maintaining honesty and transparency throughout the research process.

We are still in an early stage of applications of computational methods within social science research in general, and methods and scope of use within various fields are developing quickly, bringing both new methodological possibilities and new challenges; see for example Grimmer et al. (2021) for an overview. I believe we will see more applications of algorithm-assisted approaches in the future, which may turn out to be more fruitful than the approach taken in this thesis, which, with its innate limitations, is among the early applications of such methods to ILSA data, which are among the most accessible and relatively large datasets in the field of education (Immekus et al., 2022; Lezhnina & Kismihók, 2022).

Data and analysis availability

As all data applied in this thesis comes from the TALIS 2018 study, which is publicly available through the OECD website (OECD, 2019a), anyone can access the data and should be able to reproduce the results presented throughout the three articles included. The practice of sharing data and detailed analysis methods aligns with the principles of Open Science, which is gaining popularity to address the replication crisis in research, among other objectives (UNESCO, 2022a). Specifically for the article developing the measure of PLCs in TALIS, presented in chapter 2, it was an ambition that these measures should be made available for other researchers to apply, and therefore, the scripts used to produce these scales, and the scale scores themselves are available in the supplementary materials that are published with the journal article (Christensen, 2022). Additionally, the scripts used to produce the results of all three articles are available from a repository at the Open Science Framework (Christensen, 2024)[4]. The development of the measure of PLCs was primarily done using the MPLUS software package (Muthén & Muthén, 1998-2017), while all other analyses were done using R (R: a language and environment for statistical computing), which is freely available (R Core Team, 2022).

Summary and discussion

The first step of this project was to develop a measure of PLCs in the TALIS 2018 data with the aim of enabling international and cross-national research on PLCs. The measure developed consists of three subdimensions: Collaborative practice with focus on student learning (CPL), Shared vision and responsibilities (SVR), and Supportive conditions (SC),

[4] The articles presented in chapter 3 and chapter 4 have not yet been published, therefore there is a possibility that the analyses of these chapters, and thereby the scripts in this repository will be revised.

and the higher order construct of PLCs. These PLC measures can be applied with the data of 42 of the countries participating in TALIS 2018, and are available for use in further studies on the various themes covered in the TALIS study, either within specific contexts, for comparing contexts, or for analyzing the pooled international data. The second article explores the relationship between the extensive information about the school contextual factors included in the TALIS study, and the teachers' perceived level of the PLCs. The potential influence of contextual factors has arguably received little attention as PLCs have been implemented across the globe. The third article analyzes the relationships between each of the subdimensions of PLCs, as well as the overall measure of PLCs, and teachers' job satisfaction, self-efficacy and their clarity of instruction.

Limitations

Each article has its own limitations, which are addressed within them, and in addition, some of the limitations have been further elaborated and discussed in this introductory chapter. An additional limitation, which has yet to be unfolded, is that all three articles, and therefore all the results of this thesis, are tied to the TALIS 2018 study. This raises a question of the external validity or generalizability of the results and conclusions. As an example, the PLC measure itself is to some extent limited by the information available in TALIS, and the same goes for the information we have available to us for the explorative analysis of the contextual factors. It cannot be determined to what degree the results would have differed if further or other themes and items were included in the TALIS study, and therefore what this means for the results and conclusions of this thesis. As argued in the introduction, the very reason for using the TALIS study was the extensive range of themes it covers, and the large number of countries in which data has been systematically collected and prepared for cross-national comparisons. This thesis benefits from using TALIS in terms of both the availability of items

relevant for developing a measure of PLCs, and the themes it is possible to investigate across many different countries using this measure, which is apparent from the many aspects covered in the official reports and in secondary analyses of the TALIS data (OECD, 2019b; Veletić et al., 2024). Finally, as the ambition of this thesis is to contribute to international research on PLCs, it is worth noting that the majority of countries participating in TALIS are OECD countries, and while countries from many parts of the globe are represented, OECD countries are noticeably overrepresented. We do not know whether or how the results would have been different had the data been collected at a different point in time, from a different group of countries, using different items or other methods entirely.

Overview of the findings

The following section will outline the results achieved across the three research articles in relation to the objectives of this thesis. The first ambition, to develop a measure of PLCs which enables international and cross-national research, is presented in chapter 2 and discussed in various ways throughout this chapter. Ultimately, my conclusion is that the measure created is satisfactory, and can be applied in relation to the other research objectives of this thesis, and can potentially be used in further analysis.

The second objective, to explore the school contextual factors related to PLCs using international data, relates to the explorative analysis presented in chapter 3. The results indicate that several of the important variables in our predictive model of PLCs were in some way related to factors mentioned in the relatively sparse existing literature. That these same factors are among the most effective predictors of PLCs when considering the hundreds of variables included in the international data, suggests that there may to some extent be universally favorable conditions for PLCs to operate in. Furthermore, some aspects that have previously received little or no attention related to PLCs were identified. Some of these

factors have the potential both to inform further studies into the contexts in which PLCs operate, and, if deemed important by further research, to be altered by changes in policies and practices at different levels. Among these are the positive relationship with principals that have experience working as teachers and the prevalence of PLCs; the negative relationship with principals' perceived incentives and possibilities to participate in professional development, and/or their attitude towards professional development; school autonomy regarding staffing; and principals' instructional leadership behavior, which was an important predictor in our analysis, as opposed to the other leadership models mentioned in the PLC literature.

The final objective, to investigate the relationship between PLCs and various teacher outcomes across different national contexts, was addressed in chapter 4. We analyzed the relationship between the overall measure of PLCs as well as the three subdimensions, and their relationship with teachers' self-efficacy in instruction, job satisfaction and clarity of instruction. We analyzed these relationships using the pooled international data, as well as the individual country data. There are many nuances across the analysis covered in the chapter, but the general finding was that the PLC measure, and especially the Shared vision and responsibilities (SVR) dimension, has a robust and positive relationship with teachers' job satisfaction. This relationship is identified across most countries and teacher characteristics. For instance, the relationship between the overall PLC measure and teachers' job satisfaction is significant and positive in 39 of the 40 countries included in the analysis, and among the cases where it is significant, the relationship ranges from 0.150 to 0.351 standard deviations (SD) in magnitude. There are several cases where this relationship is both significantly weaker and stronger than the median across the countries. On the contrary, the relationships between any of the PLC constructs and teacher's self-efficacy in instruction and clarity of instruction are in most cases weak, and in several cases not statistically significant, but there

are exceptions to these general results. While the relationships between the PLC constructs and teachers' self-efficacy and clarity of instruction are often insignificant, there are countries with coefficients that differ significantly from the median coefficient in the analysis of each of these relationships. The analysis of the subgroups of teachers reveals that these relationships are quite stable across teacher experience and gender, and only the group of male teachers with less than three years' experience stand out by having no significant relationships between any of the PLC constructs and neither self-efficacy nor clarity of instruction.

In sum, the result of this analysis suggests that the relationships between PLCs and the expected outcomes for teachers do to some degree vary between country contexts. We find generally weak and often insignificant relationships with self-efficacy and clarity of instruction. The relationship between PLCs and teachers' job satisfaction is more convincing and stable across our analysis, but it seems that this positive relationship is driven mostly by the SVR construct rather than the overall PLC measure and the two other dimensions.

Discussion and directions for future research

The combined results of this thesis connect and contribute to the body of research on PLCs in various ways, and also provide suggestions for further research in several directions. These will be discussed and related to the general literature on PLCs in the following.

Time will tell if others find the measure of PLCs developed here useful. If so, it may be towards entirely different aspects than those explored and analyzed in this thesis, or towards further expanding on subjects covered within this project. For instance, the analysis of the relationship between PLCs and teacher outcomes included in this thesis explores the relationship across many different countries, but does not go into depth on why these relationships differ in ways that may or may not be expected. It could be interesting to

conduct more detailed analyses of selected countries, or groups of countries, and how they differ in the relationship between PLCs and such outcomes. ILSA studies has previously provided the basis for more in-depth analyses of countries, for example with the inclusion of additional country-specific information (Esping-Andersen, 2008; Hatos & Hatos, 2019; Woessmann, 2009). Additionally, the explorative machine learning analysis provided several directions with a potential to refine and inform further studies into beneficial contextual conditions for PLCs in schools, to be evaluated using different methods and data.

The most concrete and robust finding of the analyses in this thesis is that, across almost all countries included, teachers are more satisfied when working in schools that function as PLCs. This is not the essential goal of PLCs, but it may still be an important finding. In recent years there have been global concerns about teacher shortages, and increased focus on retention and/or attrition. According to UNESCO, teacher shortages and attrition is a major and growing concern in many countries across the globe, and among their recommendations to address these issues are measures to improve teachers' working conditions, and reinforce their educational decision-making (UNESCO, 2022b). The results of chapter 4, which mirror those of several other studies, suggest that teachers' job satisfaction is significantly higher in schools that function as a PLC. Our results suggest that this positive relationship is especially strong for the PLC dimension that covers aspects such as a shared responsibility for school issues and opportunities to participate in school decisions (SVR). The relationship with self-efficacy and clarity of instruction, however, is generally weak in magnitude, but more dependent on the context. On this basis, implementing PLCs in schools may be one way to increase the likelihood of keeping more teachers in the profession, and at the same time develop their professionalism and potentially improve student learning outcomes, although we find less substantial support for the latter.

While the focus of this thesis is on international perspectives on PLCs, the interest came from the Danish context. Since I began working on this thesis, a few studies and evaluations of PLCs in Denmark have been published. These evaluations have generally produced results that are "quite disappointing" (Thorborg & Qvortrup, 2024, p. 163). For instance, a study that involved more than 200 schools from thirteen municipalities did not find any significant relationships (correlations) among seven hypothesized outcomes for teachers and students, which included teachers' job satisfaction and self-efficacy, as well as student achievement (Jensen et al., 2020).

An interesting question arises from the mixed results across many quantitative studies on PLCs, including this thesis: is it expected that PLCs will have a synergistic effect, where successfully implementing all dimensions leads to better results than the sum of the parts? In the work of Hord, DuFour and Hargreaves, it seems that all dimensions or aspects of PLCs are important, but the logic behind combining the various aspects into one seems unclear (DuFour, 2004; Hargreaves, 2007; Hord et al., 2008). Across the many quantitative studies on various effects of PLCs, many find significant relationships of some, but not all PLC dimensions, and most studies include either only subdimensions or only an overall measure. This raises a question of empirical support for the logic behind combining the various aspects or subdimensions, with equal emphasis, into one concept. A similar question was raised by Lomos et al. (2011) in the widely cited meta-analysis of the effects of Professional Communities[5] on student learning:

> What is still unclear, however, is the argument for integrating these variables into one concept, that of professional community. The methodological decision of grouping several essential characteristics into one factor still needs more empirical support. What is required is a proper validation of the professional

[5] While this study is on the related concept of Professional Communities (PC), the studies involved in the meta-analysis include studies on PLCs, and the study is often referred to within the PLC literature.

> community measurement by explicitly relating the concept to its underlying latent constructs (p. 139).

Arguably this question is still relevant today. In the case of teacher outcomes, this can be illustrated through the overview of studies in chapter 4, where most of the studies cited find similarly mixed results, especially in relation to the relationship with teachers' self-efficacy and instructional practices. Such differences between the impact of to the individual dimensions also concerns the research into the effect of PLCs on student learning. For instance, in the study by Bolam et al. (2005), which is frequently referred to as evidence supporting the effect of PLCs on student learning, they find a significant and positive relationship with student achievement for two of their four PLC dimensions. As such, the results of the analyses of the relationship between PLCs and teachers' outcomes in chapter 4 mirrors many other studies, by finding significant relationships between some, but not all PLC dimensions, and in addition, the overall PLC measure has a weaker relationship with the outcomes than the sum of the subdimensions. Is this theoretically expected? Is it a sign of failed implementation or measurement? Whatever the answer may be, the results of this thesis, specifically in terms of the variation in teacher outcomes across countries that we found in chapter 4, suggest that the effects of PLCs, as well as that of the subdimensions, also vary across contexts. Perhaps there are now sufficient studies available to perform meta-analysis of the effects of the various subdimensions of PLCs, for teachers and for students. Again there are challenges related to the comparability of measures applied, but the three common dimensions that Lee et al. (2022) identified across a number of applied measures could may serve as the basis for selecting studies for a meta-analysis into the expected effects of PLCs on students and teachers.

Time will tell if future TALIS studies or other large-scale studies will provide the necessary items to develop more comprehensive measures of PLCs, or even better,

develop and include an official measure of this popular concept. The issues related to defining and measuring PLCs remain an obstacle towards accumulating and comparing what we know about the effects of PLCs. Perhaps the leading researchers within the field could advance the research into PLCs substantially by joining forces with the aim of agreeing on a way to define and measure PLCs, and subsequently test and develop this measure in collaboration with OECD or IEA in the hopes of ending up with an agreed-upon, internationally applicable measure, which covers all important aspects. There are still many questions about PLCs, the contexts in which they operate, and the effects they may have on students and teachers, both within and across different national contexts. This thesis provides some of the initial steps towards internationally comparable research on PLCs and the expected outcomes.

References

Ainley, J., & Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018 Conceptual Framework (OECD Education Working Papers No. 187)*.

Anders, J., Has, S., Jerrim, J., Shure, N., & Zieger, L. (2021, 2021/02/01). Is Canada really an education superpower? The impact of non-participation on results from PISA 2015. *Educational Assessment, Evaluation and Accountability, 33*(1), 229-249. https://doi.org/10.1007/s11092-020-09329-5

Andrade, C. (2021). HARKing, cherry-picking, p-hacking, fishing expeditions, and data dredging and mining as questionable research practices. *The Journal of clinical psychiatry, 82*(1), 25941.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics : an empiricist's companion*. Princeton University Press.

Bellibas, M. S., Bulut, O., & Gedik, S. (2016). Investigating professional learning communities in Turkish Schools: the effects of contextual factors. *Professional development in education, 43*(3), 353-374. https://doi.org/10.1080/19415257.2016.1182937

Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education), 21*(1), 33-46. https://doi.org/10.1007/s11092-008-9064-9

Blankenship, S. S., & Ruona, W. E. A. (2007). *Professional Learning Communities and Communities of Practice: A Comparison of Models, Literature Review* Paper presented at the Academy of Human Resource Development International Research Conference in The Americas Indianapolis. https://eric.ed.gov/?id=ED504776

Bolam, R., McMahon, A., Stoll, L., Thomas, S., Wallace, M., Greenwood, A., Hawkey, K., Ingram, M., Atkinson, A., & Smith, M. (2005). *Creating and Sustaining Effective Professional Learning Communities*.

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational researcher, 33*(8), 3-15. http://www.jstor.org.ez.statsbiblioteket.dk/stable/3699979

Buckley, F. (2000). *Team Teaching: What, Why, and How?* Thousand Oaks. https://doi.org/10.4135/9781452220697

Buczynski, S., & Hansen, C. B. (2010, 2010/04/01/). Impact of professional development on teacher practice: Uncovering connections. *Teaching and Teacher Education, 26*(3), 599-607. https://doi.org/https://doi.org/10.1016/j.tate.2009.09.006

Carlgren, I., Handal, G., Vaage, S., Sveinung, V., Carlgren, I., Handal, G., & Vaage, S. (1994). *Teachers' minds and actions : research on teachers' thinking and practice*. Falmer.

Choi, Á., & Jerrim, J. (2016, 2016/04/02). The use (and misuse) of PISA in guiding policy reform: the case of Spain. *Comparative Education, 52*(2), 230-245. https://doi.org/10.1080/03050068.2016.1142739

Christensen, A. A. (2022). A Global Measure of Professional Learning Communities. *Professional development in education*, 1-17. https://doi.org/10.1080/19415257.2022.2065516

Christensen, A. A. (2024). *Ph.D.-Thesis*. OSF.io Online data repository. https://osf.io/yqejs/

Curry, M. (2008). Critical Friends Groups: The Possibilities and Limitations Embedded in Teacher Professional Communities Aimed at Instructional Improvement and School Reform. *Teachers College Record, 110*(4), 733-774. https://doi.org/10.1177/016146810811000401

Danmarks Evalueringsinstitut. (2018). *Professionnelle Læringsfællesskaber - tættere på undervisningen [Professional Learning Communities: Closer to the Instruction]*. https://www.eva.dk/grundskole/professionelle-laeringsfaellesskaber-taettere-paa-undervisningen

Dean, S. E., & Witherspoon, C. F. (1962). *Team teaching in the elementary school*. Office of Education, Division of State and Local School Systems, Instruction ….

Doğan, S., & Adams, A. (2018). Effect of professional learning communities on teachers and students: reporting updated results and raising questions about research design. *School Effectiveness and School Improvement, 29*(4), 634-659. https://doi.org/10.1080/09243453.2018.1500921

DuFour, R. (2004). What is a "professional learning community"? *Educational leadership, 61*(8), 6-11.

DuFour, R., DuFour, R. B., Eaker, R. E., Many, T. W., & Mattos, M. (2016). *Learning by doing : a handbook for professional learning communities at work* (Third ed.). Solution Tree Press.

DuFour, R., & Marzano, R. J. (2011). *Leaders of learning : how district, school, and classroom leaders improve student achievement*. Solution Tree Press.

Esping-Andersen, G. (2008). Childhood investments and skill formation. *International tax and public finance, 15*, 19-44.

Fox, N. J. (2008). Post-positivism. *The SAGE encyclopedia of qualitative research methods, 2*(1), 659-664.

Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production? *European Journal of Education, 55*(4), 618-631. https://doi.org/https://doi.org/10.1111/ejed.12417

Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. S. (2001, 12/01). What Makes Professional Development Effective? Results From a National Sample of Teachers. *American Educational Research Journal - AMER EDUC RES J, 38*, 915-945. https://doi.org/10.3102/00028312038004915

Gorard, S. (2003). Finding secondary data: the 'idle' researcher. In *Quantitative methods in social science*. Bloomsbury Publishing Plc.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science, 24*(1), 395-419. https://doi.org/10.1146/annurev-polisci-053119-015921

Guo, L., & Wang, J. (2021). Relationships between teacher autonomy, collaboration, and critical thinking focused instruction: A cross-national study. *International journal of educational research, 106*, 101730. https://doi.org/10.1016/j.ijer.2020.101730

Hadar, L. L., & Brody, D. L. (2013). The Interaction Between Group Processes and Personal Professional Trajectories in a Professional Development Community for Teacher Educators. *Journal of teacher education, 64*(2), 145-161. https://doi.org/10.1177/0022487112466898

Hairon, S., Goh, J. W. P., Chua, C. S. K., & Wang, L.-y. (2017). A research agenda for professional learning communities: moving forward. *Professional development in education., 43*(1), 72-86. https://doi.org/10.1080/19415257.2015.1055861

Hanushek, E. A., & Woessmann, L. (2014). Institutional Structures of the Education System and Student Achievement: A Review of Cross-country Economic Research. In R. Strietholt, W. Bos, J.-E. Gustafsson, & M. Rosén (Eds.), *Educational policy evaluation through international comparative assessments* (pp. 145-175). Waxmann.

Hargreaves, A. (2007). Sustainable professional learning Communities. In L. Stoll & K. S. Louis (Eds.), *Professional learning communities: divergence, depth and dilemmas*. Open University Press.

Hargreaves, A. (2019, 2019/07/04). Teacher collaboration: 30 years of research on its nature, forms, limitations and effects. *Teachers and Teaching, 25*(5), 603-621. https://doi.org/10.1080/13540602.2019.1639499

Hargreaves, A., & Fullan, M. (2012). *Professional capital : transforming teaching in every school*. Routledge.

Harris, A., Huffman, J. B., & Jones, M. S. (2017). *Teachers leading educational reform : the power of professional learning communities*. Routledge. https://doi.org/10.4324/9781315630724

Hatos, A., & Hatos, R. (2019). Varieties of capitalism and academic achievement. Does the welfare production regime have an impact on 15 years' old Mathematics results? *Journal of Social Research & Policy, 10*(1).

Hernán, M. A. (2018, May). The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am J Public Health, 108*(5), 616-619. https://doi.org/10.2105/ajph.2018.304337

Hernández-Torrano, D., & Courtney, M. G. R. (2021, 2021/07/30). Modern international large-scale assessment in education: an integrative review and mapping of the literature. *Large-scale Assessments in Education, 9*(1), 17. https://doi.org/10.1186/s40536-021-00109-1

Hord, S. M. (1997). *Professional Learning Communities: Communities of Continuous Inquiry and Improvement*. Southwest Educational Development Laboratory.

Hord, S. M., Sommers, W. A., & Hargreaves, A. (2008). *Leading professional learning communities : voices from research and practice*. Corwin Press, a SAGE Company.

Hudson, C. (2023). A Conceptual Framework for Understanding Effective Professional Learning Community (PLC) Operation in Schools. *Journal of education (Boston, Mass.)*. https://doi.org/10.1177/00220574231197364

IEA. (2017). *List of Studies.* IEA ILSA-Gateway. Retrieved April 1, 2024 from https://ilsa-gateway.org/studies

Immekus, J. C., Jeong, T.-s., & Yoo, J. E. (2022, 2022/12/27). Machine learning procedures for predictor variable selection for schoolwork-related anxiety: evidence from PISA 2015 mathematics, reading, and science assessments. *Large-scale Assessments in Education, 10*(1), 30. https://doi.org/10.1186/s40536-022-00150-8

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher Effects and Teacher-Related Policies. *Annual Review of Economics, 6*(Volume 6, 2014), 801-825. https://doi.org/https://doi.org/10.1146/annurev-economics-080213-040845

Jensen, L. B., Lorentsen, A., Andersen, M. B., Christiansen, N. L. S., & Hansen, I. S. (2020). *Udfordringer og succeser i skoleudvikling: når tallene taler [Challenges and Successes in School Development: When the Numbers Speak]*. Aalborg Universitetsforlag.

Jerrim, J., Parker, P., & Shure, D. (2019, 01/01). Bullshitters. Who are They and What Do We Know About Their Lives? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3390272

Joseph, C., & Russell, K. S. (2011). *Research Methods in Education*. SAGE Publications, Inc. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1236063&site=ehost-live

Kools, M., & Stoll, L. (2016). What Makes a School a Learning Organisation? *OECD Education Working Papers*. https://doi.org/https://doi.org/10.1787/5jlwm62b3bvh-en

Kruse, S. D., & Johnson, B. L. (2016). Tempering the normative demands of professional learning communities with the organizational realities of life in schools: Exploring the cognitive dilemmas faced by educational leaders. *Educational management, administration & leadership, 45*(4), 588-604. https://doi.org/10.1177/1741143216636111

Lave, J. (1991). Situating learning in communities of practice. In L. Resnick, B. Levine, & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition*. American Psychological Association.

Lee, M., Kim, J. W., Mo, Y., & Walker, A. D. (2022). A review of professional learning community (PLC) instruments. *Journal of educational administration, 60*(3), 262-287. https://doi.org/10.1108/JEA-03-2021-0060

Lezhnina, O., & Kismihók, G. (2022, 2022/03/15). Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA. *International Journal of Research & Method in Education, 45*(2), 180-199. https://doi.org/10.1080/1743727X.2021.1963226

Little, J. W. (1993). Teachers' Professional Development in a Climate of Educational Reform. *Educational Evaluation and Policy Analysis, 15*(2), 129-151. https://doi.org/10.3102/01623737015002129

Lomos, C. (2017). To what extent do teachers in European countries differ in their professional community practices? *School Effectiveness and School Improvement, 28*(2), 276-291. https://doi.org/https://doi.org/10.1080/09243453.2017.1279186

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). Professional communities and student achievement – a meta-analysis. *School Effectiveness and School Improvement*, 121-148. https://www.tandfonline.com/doi/full/10.1080/09243453.2010.550467

Louis, K. S., Marks, H. M., & Kruse, S. (1996, 1996/01/01). Teachers' Professional Community in Restructuring Schools. *American Educational Research Journal, 33*(4), 757-798. https://doi.org/10.3102/00028312033004757

Moosa, V., Salleh, S., & Hamid, L. (2022). Defining and operationalizing professional learning communities: what does the literature say? *Asia Pacific journal of education*, 1-13. https://doi.org/10.1080/02188791.2022.2070907

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art - teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231-256. https://doi.org/10.1080/09243453.2014.885451

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eighth Edition.* Los Angeles. Muthén & Muthén.

Nigel, N., & John, E. (2012). *Curriculum, Pedagogy and Educational Research: The Work of Lawrence Stenhouse*. Taylor and Francis. https://doi.org/10.4324/9780203818114

OECD. (2019a). *TALIS 2018 and TALIS Starting Strong 2018 User Guide*. https://www.oecd.org/education/talis/TALIS_2018-TALIS_Starting_Strong_2018_User_Guide.pdf

OECD. (2019b). *TALIS 2018 Results (Volume I)*. https://doi.org/https://doi.org/10.1787/1d0bc92a-en

OECD. (2019c). *TALIS 2018 Techincal Report* https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf

OECD. (2020). *TALIS 2018: Insights and Interpretations*. www.oecd.org/education/talis/TALIS2018_insights_and_interpretations.pdf

Oxford English Dictionary. (2024). school, n.[1], sense II.5.a. In *Oxford English Dictionary*.

Pedersen, A. K. (2021). *Samarbejde i skolen: En empirisk undersøgelse af, hvordan samarbejde tematiseres i kommunikationen i et kommunalt dansk skolesystem med fokus på, om og hvordan samarbejde bidrager til kobling mellem professionelles praksis og elevers læring [Collaboration in the school ...]* AU Library Scholarly Publishing Services. https://ebooks.au.dk/aul/catalog/book/414

Pennell, J. R., & Firestone, W. A. (1996). Changing Classroom Practices through Teacher Networks: Matching Program Features with Teacher Characteristics and Circumstances. *Teachers College Record, 98*(1), 46-76. https://doi.org/10.1177/016146819609800105

Qiao, X., Yu, S., & Zhang, L. (2018). A review of research on professional learning communities in mainland China (2006–2015): Key findings and emerging themes. *Educational management, administration & leadership, 46*(5), 713-728. https://doi.org/10.1177/1741143217707523

R Core Team. (2022). *R: A Language and Environment for Statistical Computing.* In R Foundation for Statistical Computing. https://www.R-project.org/

Rutkowski, D., Thompson, G., & Rutkowski, L. (2020). Understanding the Policy Influence of International Large-Scale Assessments in Education. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment : Understanding IEA's Comparative Studies of Student Achievement* (pp. 261-277). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_15

Scheerens, J. (2016). Defining the Key Factors in Instructional Effectiveness. In J. Scheerens (Ed.), *Educational Effectiveness and Ineffectiveness: A Critical Review of the Knowledge Base* (pp. 51-76). Springer Netherlands. https://doi.org/10.1007/978-94-017-7459-8_3

Schleicher, A. (2016). *Teaching Excellence through Professional Learning and Policy Reform*. OECD Publishing. https://doi.org/https://doi.org/10.1787/9789264252059-en

Schön, D. A. (1992). *The Reflective Practitioner: How Professionals Think in Action*. Routledge. https://doi.org/10.4324/9781315237473

Senge, P. M. (2006). *The fifth discipline : the art and practice of the learning organization* (Rev. and updated. ed.). Doubleday/Currency.

Sleegers, P., Brok, P., Verbiest, E., Moolenaar, N., & Daly, A. J. (2013). Toward Conceptual Clarity: A Multidimensional, Multilevel Model of Professional Learning Communities in Dutch Elementary Schools. *The Elementary school journal, 114*(1), 118-137. https://doi.org/10.1086/671063

Stancel-Piątak, A., & Schwippert, K. (2022). Comprehensive Frameworks of School Learning in ILSAs. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 145-173). Springer International Publishing. https://doi.org/10.1007/978-3-030-88178-8_8

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct Meaning in Multilevel Settings. *Journal of educational and behavioral statistics, 41*(5), 481-520. https://doi.org/10.3102/1076998616646200

Stenhouse, L. (1975). *An introduction to curriculum research and development*. Heinemann.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional Learning Communities: A Review of the Literature. *Journal of Educational Change, 7*(4), 221-258. https://doi.org/10.1007/s10833-006-0001-8

Stoll, L., & Louis, K. S. (2007). Professional learning communities: elaborating new approaches. In *Professional learning communities: divergence, depth and dilemmas*. Open University Press.

Strietholt, R., Gustafsson, J.-E., Rosén, M., & Bos, W. (2014). Outcomes and Causal Inference in International Comparative Assessments. In R. Strietholt, W. Bos, J.-E. Gustafsson, & M. Rosén (Eds.), *Educational policy evaluation through international comparative assessments* (pp. 9-18). Waxmann.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of research in science teaching, 37*(9), 963-980. https://doi.org/10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0

Thorborg, M., & Qvortrup, L. (2024). Professional learning and development in Denmark: From formal competence development and professional learning communities to an emergence approach. In (1 ed., Vol. 1, pp. 160-171). Routledge. https://doi.org/10.4324/9781003322610-13

UNESCO. (2022a). *An introduction to the UNESCO Recommendation on Open Science*. https://doi.org/10.54677/XOIR1696

UNESCO. (2022b). *Transforming education from within. Current trends in the status and development of teachers*. https://unesdoc.unesco.org/ark:/48223/pf0000383002/PDF/383002eng.pdf.multi

Valckx, J., Vanderlinde, R., & Devos, G. (2020). Departmental PLCs in secondary schools: the importance of transformational leadership, teacher autonomy, and teachers' self-efficacy. *Educational studies, 46*(3), 282-301. https://doi.org/10.1080/03055698.2019.1584851

Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017, 2017/01/01/). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education, 61*, 47-59. https://doi.org/https://doi.org/10.1016/j.tate.2016.10.001

Veletić, J., Mejía-Rodríguez, A. M., & Olsen, R. V. (2024). A systematic literature review of TALIS secondary research: Trends and future directions. *Review of Education, 12*(1), e3469. https://doi.org/https://doi.org/10.1002/rev3.3469

Veletić, J., & Olsen, R. V. (2021). Developing a shared cluster construct of instructional leadership in TALIS. *Studies in educational evaluation, 68*, 100942. https://doi.org/10.1016/j.stueduc.2020.100942

Vieluf, S., Kaplan, D., Klieme, E., & Bayer, S. (2012). *Teaching practices and pedagogical innovation: Evidence from TALIS*. OECD. https://doi.org/10.1787/9789264123540-en

Wagemaker, H. (2020a). Introduction to Reliability and Validity of International Large-Scale Assessment. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment : Understanding IEA's Comparative Studies of Student Achievement* (pp. 1-5). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_1

Wagemaker, H. (2020b). Study Design and Evolution, and the Imperatives of Reliability and Validity. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment : Understanding IEA's Comparative Studies of Student Achievement* (pp. 7-21). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_2

Watson, C. (2014). Effective professional learning communities? The possibilities for teachers as agents of change in schools. *British Educational Research Journal, 40*(1), 18-29. https://doi.org/10.1002/berj.3025

Webb, R., Vulliamy, G., Sarja, A., Hämäläinen, S., & Poikonen, P.-L. (2009). Professional learning communities and teacher well-being? A comparative analysis of primary schools in England and Finland. *Oxford review of education, 35*(3), 405-422. https://doi.org/10.1080/03054980902935008

Wenger, E. (1998). *Communities of practice : learning, meaning, and identity*. Cambridge University Press.

Woessmann, L. (2009). *International evidence on school tracking: A review* (1613-6373). (CESifo DICE Report, Issue. https://www.ifo.de/DocDL/dicereport109-rr1.pdf

Ydesen, C. (2019). *The OECD's Historical Rise in Education: The Formation of a Global Governing Complex* (1st ed. 2019. ed.). Springer International Publishing. https://doi.org/10.1007/978-3-030-33799-5

Zhang, J., Yin, H., & Wang, T. (2020). Exploring the effects of professional learning communities on teacher's self-efficacy and job satisfaction in Shanghai, China. *Educational studies*, 1-18. https://doi.org/10.1080/03055698.2020.1834357

Zieger, L. R., Jerrim, J., Anders, J., & Shure, N. (2022, 2022/11/02). Conditioning: how background variables can influence PISA scores. *Assessment in Education: Principles, Policy & Practice, 29*(6), 632-652. https://doi.org/10.1080/0969594X.2022.2118665

Chapter 2

A Global Measure of Professional Learning Communities

Written by Anders Astrup Christensen

The article presented in this chapter is published in Professional Development in Education: https://doi.org/10.1080/19415257.2022.2065516.

Chapter 3

Exploring School Factors Related to Professional Learning Communities: A Machine Learning Approach Using Cross-National Data

Written by Anders Astrup Christensen, Kristoffer Laigaard Nielbo & Sedat Gümüş

The article presented in this chapter is published in Educational Studies:

https://doi.org/10.1080/03055698.2024.2369855

Chapter 4:

Professional learning communities and teacher outcomes.

A cross-national analysis

Written by Anders Astrup Christensen & John Jerrim.

The article presented in this chapter is published in Teaching and Teacher Education:

https://doi.org/10.1016/j.tate.2024.104920

Danish School of Education
Aarhus University
Tuborgvej 164
2400 København NV

Tel.: 8716 1300
E-mail: dpu@au.dk
edu.au.dk