university of groningen / faculty of behavioural and social sciences / teacher education

# Features of effective professional development interventions in different stages of teacher's careers

*A review of empirical evidence and underlying theory*

**D.W.MAANDAG@RUG.NL**

D.W.MAANDAG, M.HELMS-LORENZ, LUGTHART, E., VERKADE, A.T. & VAN VEEN, K.
2017

# Contents

# Summary in Dutch[1]

## Doelstellingen van de overzichtsstudie

Met de noodzaak tot 'een leven lang leren' tijdens de beroepsloopbaan is er de laatste decennia in algemene zin sprake van toegenomen wetenschappelijke aandacht voor het antwoord op de vraag hoe leren en werken het best samen kunnen gaan en elkaar kunnen versterken (zie bijvoorbeeld Onstenk, 1997). Kenmerkend voor krachtige vormen van professionalisering is onder meer dat de werkgerelateerd zijn. Dit geldt ook voor de professionalisering van (aanstaande) leraren (Van Veen, Zwart, Meirink en Verloop, 2010). De toenemende integratie van leren en werken zien we in het onderwijs internationaal terug op institutioneel niveau, waar de opleidingsfunctie (de lerarenopleiding) en de arbeidsfunctie (de scholen) de laatste jaren verder naar elkaar toegegroeid zijn (Maandag, Deinum, Hofman, & Buitink, 2007). Fenomenen als (academische) opleidingsscholen, inductietrajecten en -arrangementen, professionele leergemeenschappen, zijn hiervan in Nederland een uiting. Leren en werken treffen elkaar daarbij in diverse fasen van de beroepsloopbaan en in diverse stadia van de professionele ontwikkeling. Met uiteenlopende interventies wordt geprobeerd de professionele groei van (aanstaande) leraren te bevorderen. Waarop dergelijke interventies in de diverse stadia van de beroepsontwikkeling het best gericht kunnen zijn en wat daarbij de meest effectieve verschijningsvormen zijn is vaak onduidelijk (Van Veen et. al. 2010). Theorievorming op dit terrein blijft tot op heden uit.

In de nabije toekomst zullen veel ervaren leraren het Nederlandse onderwijs verlaten en plaats maken voor weliswaar star bekwame, maar qua pedagogisch-didactisch niveau, beginnende leraren. Het op deze wijze wegvloeien van pedagogisch-didactische expertise (ervaren leraren hebben gemiddeld hogere vaardigheidsniveaus) zet de kwaliteit van het onderwijs onder druk. Dit maakt het wenselijk professionaliseringsinterventies expliciet te richten op het (versneld) ontwikkelen van pedagogisch-didactische vaardigheden van starters maar ook van ervaren leraren. Momenteel wordt er in samenwerking tussen scholen en lerarenopleidingen, zowel landelijk als regionaal gewerkt aan diverse projecten om dit te stimuleren.

Deze NRO-overzichtsstudie kan worden gezien als een follow-up van de vorige NRO-overzichtsstudie van Van Veen et al. (2010) (zie ook Van Driel, Meirink, van Veen & Zwart, 2012) waarbij getracht wordt de kennisbasis aan te vullen met inzichten van de periode 2010-2016, meer in te gaan op de onderliggende theoretische noties en expliciet het verschil in werk- of ontwikkelingsfases mee te nemen. Drie onderzoeksvragen staan centraal in deze overzichtsstudie:

1. Wat zijn de kenmerken van effectieve professionaliseringsinterventies?
2. Welke theoretische noties liggen ten grondslag aan deze effectieve interventies?
3. Verschillen de effectieve kenmerken met betrekking tot de verschillende stadia van professionele ontwikkeling?

De zoektocht naar overzichts- en empirische studies richtte zich sterk op het vinden van nieuwe inzichten over de kenmerken van effectieve interventies tussen 2010 tot 2016 in aanvulling op de

[1] This study was subsidized by the NRO, project number 405-15-713

bestaande kennis hierover uit 2010 (onderzoeksvraag 1). Volgens Kennedy (2016) is hierbij niet alleen relevant wat precies die kenmerken zijn, maar vooral hoe deze worden ingezet in relatie tot het doel van de interventie, wat zij beschrijft in termen van 'theory of action' (vergelijkbaar met Van Veen et al's theory of improvement). In de selectie van relevante studies was daarom een expliciete 'theorie van verandering of actie' een extra selectiecriterium (onderzoeksvraag 2). Tenslotte, leraren verschillen in de ontwikkeling van hun pedagogisch-didactische vaardigheden (Muijs, Kyriakides, van der Werf, Creemers, Timperley and Earl, 2014) en leerbehoeften (Louws, 2016) in verschillende werkfases. De zoektocht naar professionele ontwikkeling interventies gerelateerd aan de verschillende werkfases leverde niet veel informatie over de verschillende doelgroepen op in termen van ervaring of deskundigheid, met uitzondering van de groep van starters die uitgebreid is en wordt onderzocht. Onderzoeksvraag 3 had daarom betrekking op interventies voor starters en voor ervaren docenten (onderzoeksvraag 3).

## Bevindingen

Naar aanleiding van **de eerste onderzoeksvraag** naar de kenmerken van effectieve professionalisering is een eerste bevinding dat de consensus over deze kenmerken, die in 2010 door veel reviewers werd beschreven, ook in de studies van 2010-2016 wordt teruggevonden. Deze algemene kenmerken zijn, zoals Desimone (2009) samenvatte, een focus op vak inhoud en vakdidactiek, actief leren door leraren, samenhang met de eigen lespraktijk en school, duur en collectieve deelname. Wat opvalt in de meer recente studies is dat deze kenmerken terug komen op verschillende manieren. Niet elk kenmerk is even relevant in elke studie. Dit lijkt bepaald te worden door de specifieke doelen en context.

Wat opvalt bij veel studies is de sterke focus op pedagogical content knowledge (PCK), dat in het Nederlands wordt omschreven als vakdidactiek. Vakdidactiek heeft niet alleen betrekking op (de vele mogelijkheden van) het uitleggen van het vak aan leerlingen (vakdidactisch repertoire variërend van directe instructie tot activeren en differentiëren binnen het vak), maar is in de betekenis van PCK een veel breder begrip dat ook betrekking heeft op inzicht in hoe leerlingen een vak leren en begrijpen of waar leerlingen problemen ondervinden en last hebben van misconcepties. Veel programma's blijken zich hierop te richten. De relevantie van inzicht in vakdidactiek wordt ook bevestigd in onderzoek naar leerbehoeften van leraren in de verschillende werkfasen, waarbij leraren in alle werkfasen aan geven dat zij dit een belangrijk onderwerp blijven vinden (Louws, 2016).

Naast deze kenmerken komen ook nog andere relevante kenmerken naar voren voor interventies voor starters die retentie (het blijven werken in het onderwijs) beogen: het hebben van een coach/mentor van hetzelfde schoolvak; in de gelegenheid zijn om samen lessen voor te bereiden met docenten die hetzelfde schoolvak doceren; regelmatig, geplande afspraken voor overleg en samenwerking; en deel uitmaken van een extern netwerk. Vermindering van werkdruk heeft minder effect op retentie volgens Ingersoll en Smith (2004) terwijl verminderde werkdruk wel invloed heeft op retentie in de Nederlandse context (Helms-Lorenz, van de Grift, & Maulana, 2015). Hoe meer activiteiten worden aangeboden in een inductie arrangement, hoe groter de kans dat de starter in het onderwijs te blijft. Dus hoe meer aandacht voor de starter, des te beter. Wat ook van belang is gebleken is dat de schoolcontext een rol speelt bij de effectiviteit van inductie-arrangementen. Scholen met een hoog percentage leerlingen met een lagere sociaal-economische status ("high

poverty schools") laten minder tot geen effecten van inductie zien (Ingersoll et al., 2004). Ook deze analyse liet zien dat de doelen van de interventie bepalend zouden moeten zijn voor de keuze en opzet van de interventie (backward design). Als een interventie voor starters het doel heeft om het pedagogisch-didactisch handelen te versnellen, dan blijkt werkdrukreductie contraproductief te werken. Een-op-een coaching is hierbij dan effectiever (Helms-Lorenz et al., 2015).

Uit de analyse van studies van na 2010 tot 2016, kwam de volgende opsomming van effectieve activiteiten:

1. Eén op één coaching, gebaseerd op observaties, bevorderen de professionele ontwikkeling van leraren richting meer complexe lesvaardigheden;
2. Het helpt als teacher leaders goed worden opgeleid in het leiding geven aan docent ontwikkelingstrajecten; hoe ze theorie en didactische voorbeelden kunnen uitleggen aan leraren, en zelf als voorbeeld kunnen dienen als expert leraren.
3. Het centraal stellen van vak inhoud en vakdidactiek in workshops en seminars, gecombineerd met coaching om leraren te ondersteunen bij het toepassen van nieuwe lesstrategieën en didactiek.
4. Reflecteren op alleen die lesvaardigheden waar een leraar aan toe is in termen van de eigen ontwikkeling.
5. Leraren zelf hun eigen lessen laten ontwerpen en veranderingen plannen, waarbij dus rekening wordt gehouden met de specifieke context en eigen leerbehoeften.
6. Het leren over de eigen lespraktijk door middel van video-opnames van de eigen lespraktijk.
7. Het koppelen van leraren aan experts in de school, waarbij de experts coachen en als voorbeeld dienen, en langzamerhand steeds meer afstand nemen.
8. Online middelen kunnen erg ondersteunend zijn.
9. Bediscussiëren van voor gestructureerde casussen van klassensituaties.
10. Bestuderen van leerling werk (in relatie tot de eigen lessen).
11. Leergemeenschapmodel (werkgroepen) met een sterke focus op discussie met peers.
12. Een extra uur aan deelname aan professionalisering (vakdidactisch en in gesprek met vakgenoten) leidde tot een behoorlijke voortuitgang in het leerresultaten bij wiskunde.
13. Het samen leren door leraren over wiskunde vakdidactiek en het leren van hun leerlingen) is effectiever dan als leraren dit alleen doen. Het doen van praktijkonderzoek inclusief conferentie-bezoek en –presentatie is eveneens geassocieerd met betere leerling prestaties.
14. Het aanbieden van leerkeuzes aan leraren geeft hen meer de mogelijkheid aan te sluiten bij wat nuttig is voor hun eigen ontwikkeling.

Met betrekking tot **onderzoeksvraag 2**, naar de onderliggende theoretische aannames bij professionalisering over waarom het programma effectief zou zijn voor leraren, blijkt dat bij veel studies dit nauwelijks expliciet wordt aangegeven. Toch wordt er bij nader inzien in de studies tussen 2010-2016 wel degelijk in toenemende mate gebruik gemaakt van dit soort aannames in de zin van dat de meeste studies zich baseren op wat bekend is over effectieve kenmerken van professionalisering. Hierbij lijkt het didactische kernprincipe van *Backward design* een centrale rol te spelen: denkend vanuit het doel dat men wil bereiken met een professionaliseringsinterventie, wordt bepaald hoe de interventie er uit moet zien en dus welke effectieve kenmerken moeten worden meegenomen. Uit deze studies komen dan ook veel concrete voorbeelden hoe bepaalde kenmerken een rol spelen in bepaalde contexten. In steeds meer interventies waar het doel is om de dagelijkse

lespraktijk en het leren van leerlingen te beïnvloeden wordt gebruik gemaakt van vormen waarin de eigen lespraktijk centraal staat, wordt besproken en geobserveerd en waarbij het om de vakdidactiek gaat.

In een beperkt aantal studies wordt wel expliciet verwezen naar leerpsychologische principes over het leren van volwassenen (Meyers, Molefe, Brandt, Zhu, & Dhillon, 2016). Het expliciet hanteren van dit soort theoretische principes lijkt zinvol te zijn bij het ontwerpen en uitvoeren van professionalisering. Het bevestigt de stelling van Kennedy (2016) en Van Veen et al. (2010) dat het expliciteren van de 'theory of action' (de redenen waarom de interventie zou bijdragen aan het leren van leraren) een relevante en cruciale activiteit is voor zowel degenen die de interventie ontwerpen en uitvoeren als 'ondergaan'. Dit sluit ook aan bij het pleidooi van bijvoorbeeld Hattie (2012) die in het kader van het leren van leerlingen stelt dat het nuttig is om naar leerlingen transparant te zijn over de doelen en de daarbij behorende didactiek. Meer algemeen - ook al lijkt dit een vanzelfsprekendheid maar is het niet in veel PD interventies - werd er geconcludeerd dat het leren van leraren sterk lijkt op het leren van leerlingen en dat de zelfde didactische principes van toepassing zijn. Goede professionalisering is als een goede les, waarin alles wat we weten over effectief leren en lesgeven wordt meegenomen. Oftewel, goede professionalisering vraagt ook om goede leraren die het leren van leraren in de PD organiseren.

Uit de resultaten met betrekking tot **onderzoeksvraag 3**, naar het verschil tussen starters en ervaren leraren, blijkt allereerst dat de interventies voor starters andere kenmerken hebben, waarbij namelijk rekening wordt gehouden met de specifieke leerbehoeften van starters, die anders zijn voor ervaren leraren. Starters zijn veel meer gericht op het nog leren van en gesocialiseerd raken in het beroep. Het onderzoek hierna laat duidelijk zien dat ondersteuning hierbij in de vorm van inductiearrangementen zinvol is.

Het verschil in de aanpak van interventies bij starters en ervaren docenten heeft mogelijk te maken met de bereidheid van starters om van ervaren docenten te leren. Als de starter ervan overtuigd is dat de ervaren docent meer kan en meer inzicht heeft, zijn de voorwaarden voor het leren gunstig (uiteraard is dit niet altijd het geval, soms heeft de starter deze overtuiging (terecht) niet). Gezien het verschil in ervaring zal de starter deze overtuiging eerder hebben, vooral als het lesgeven lastig verloopt. Ervaren docenten hebben niet zonder meer deze overtuiging. De ervaren docent, die al veel leerervaringen achter de rug heeft, is minder snel bereid van een collega of een extern "expert" te leren. Bij interventies voor ervaren leraren is het noodzakelijk om veel aandacht te besteden aan *hoe* de interventie de docenten kan laten ontwikkelen. Hiervoor is een "theory of action" van groot belang. Dit dwingt degene die de interventie ontwerpt om na te denken over de doelen van de interventie, de gewenste uitkomsten, om stil te staan bij de beginsituatie van de docenten, en de interventie-inhoud en de activiteiten af te stemmen of de verschillen tussen de deelnemende docenten. Er moet ook ruimte zijn om de gekozen leerstrategie te expliciteren en om andere leerstrategieën van de deelnemers aan bod te laten komen en te betrekken. Bij ervaren leraren is het zinvol om interventies binnen scholen en tussen scholen te organiseren met docenten van soortgelijke vakken, bijvoorbeeld bij Lesson Study. Het is aan te raden om de starters hieraan te koppelen, mits er oog is voor de specifieke behoeften van deze doelgroep.

# Introduction

It is generally acknowledged that 'Teachers matter'; teaching quality is significantly and positively correlated with student achievement (Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009) and is an important factor from an economical point of view (Hanushek, 2011). The last decade, due to demographic trends, the quality of teaching in many western countries has come under pressure. Teachers shortages are growing and teaching experience drains out of schools. At the same time, the demands on the teaching profession change as a result of rapid technological and societal changes. For example The European Union links teaching quality with 'the school's duty to provide young citizens with the competences they need to adapt to globalized, complex environments, where creativity, innovation, initiative, entrepreneurship and commitment to continuous learning are as important as knowledge' (Caena, 2011). Similar arguments in the situation in the Netherlands can be found in policy papers like (Platform Onderwijs2032, 2016). Like elsewhere, teachers in the Netherlands are expected to operate in contexts of growing complexity, like teaching more heterogeneous classes. Initial teacher education can only account for basic preparation for these more complex skills and is therefore commonly seen as (only) the first step into the teaching career. Interventions and programs to promote teaching quality and teachers' professional growth throughout the career are becoming increasingly important.

Individual teaching quality beyond initial teacher education tends to grow with the accumulating number of years of experience of the teacher and the breadth of that experience (Berliner, 2004; Day & Gu, 2007; Ladd & Sorensen, 2015; Muijs et al., 2014; van de Grift, van der Wal, & Torenbeek, 2011). A recent research synthesis by Kini and Podolsky (2016) shows that there is a strong relationship between teachers' years of experience and teacher effectiveness in terms of gains in student outcomes, but that experience is not educative in itself. The mere length of a career does not necessarily lead to the development of expertise and improved performance and not all teachers reach high levels of teaching quality in spite of lengthy careers (Bromme, 2001; Creemers, Kyriakides, & Antoniou, 2013; van de Grift et al., 2011). Some authors contend that teacher effectiveness might even decrease in the long run of a teaching career (Berliner, 2004; Day & Gu, 2007; van de Grift et al., 2011).

It is still puzzling why some teachers become experts while others don't. Researchers have, although sometimes indirectly, addressed this issue by seeking for patterns in teachers' career trajectories, work lives and personal and professional life phases or stages (Berliner, 2004; Fessler & Christensen, 1992; Fuller, 1969; Huberman, 1989; Steffy & Wolfe, 2001). In their longitudinal research, Day and colleagues for example explicitly sought for the relationship between career patterns and teacher effectiveness (Day, Sammons, Stobart, Kington, & Gu, 2007; Day, 2008; Day & Gu, 2007). Developmental studies like these try to explain why teacher professional growth can stagnate in certain career stages from personal or professional perspectives. According to Berliner (2004) they may also help teacher educators and those responsible for planning and delivering professional development programs to think about the abilities and competences of teachers in various stages of their professional careers, thus attempting to match programs to the developmental level of the teachers. This suggests that tailoring PD interventions to levels of teacher experience or competence and to specific needs (Louws, 2016) might be more effective than the traditional 'one size fits all' approach. One of the questions driving this review addresses the issue of

empirical evidence for the effectiveness of PD interventions and programs for specific target groups, grouped on basis of their experience, career stage or achieved level of teaching quality.

Although it is generally acknowledged that teacher professional growth needs time, interventions can accelerate this process in the early career (Helms-Lorenz et al., 2015; Maulana, Helms-Lorenz, & van de Grift, 2015) and more experienced teachers can attain higher levels of teaching quality if they are provided *appropriate* opportunities to learn and develop professionally (Creemers et al., 2013). Interventions and programs to promote and sustain teaching quality and to stimulate teachers' professional growth throughout the career are usually referred to as professional development (PD) interventions or programs or, in case of professional learning throughout the career, programs for continuing professional development (CPD). 'Appropriate' must be conceived as interventions and programs that are 'fit for purpose' and likely to have intended effects. The questions of effectiveness (what are the effects of interventions and programs and what makes them effective?) have been addressed in many empirical causal impact studies and in systematic research reviews and syntheses. Borko (2004) and Desimone (2009) suggest that a consensus exists among researchers in the field on what constitutes effective interventions and programs for professional development. Effective features resulting from reviews and research syntheses are often used as general design principles for PD interventions. However, it is still unclear whether and how these general principles apply to specific situations and target groups in a valid way (van Veen, Zwart, Meirink, & Verloop, 2010). Part of this has to do with the a-theoretical nature of much of the research on the effectiveness of PDI (Creemers et al., 2013; Ingersoll & Strong, 2011; Kennedy, 2016; Opfer & Pedder, 2011). In some studies and reviews attempts are made to explore, understand and explain findings of effectiveness from different theoretical angles (Broad & Evans, 2006; Timperley, Wilson, Barrar, & Fung, 2007). So, one of the aims of the current review is to explore the possibilities of applying theoretical insights to guide the design and implementation of PD interventions for teachers in different stages of their career and teaching quality levels.

This NRO-review can be seen as a follow-up of the previous NRO-review on effective professional development from 2010 (van Driel, Meirink, van Veen, & Zwart, 2012; van Veen et al., 2010) and attempts to update from 2010 to 2016 and replicate the knowledge-basis concerning what works in PD interventions, the actual use of theoretical notions on teacher learning and career development and recent research findings on the development of teacher effectiveness and teaching quality. Regarding the focus on theoretical notions, the recent review of Kennedy (2016) argues that it is more relevant to explore the underlying 'theory of action' (in short, the pedagogy used in the PD to help teachers learn) rather than to explore the effective features. Regarding the relationships between PD in relation to teacher career stages, Day and Gu (2007) showed in detail how those stages can affect teachers' effectiveness, motivation and commitment, and therefore assuming it will also affect teacher learning, though this relationship is still hardly taken into account in PD studies (Louws, 2016). In sum, especially the focus on the theoretical notions or theory of action and the relationship with career stages will contribute to understanding and deliberately enhancing and accelerating teacher growth throughout the career in general, and contribute to tailoring PD interventions to specific contexts and target groups.

To do so this review aims to answer the following questions:

1. What are the features of effective interventions and programs for teacher professional development?
2. What are theoretical assumptions underlying the design of effective interventions and programs as used by designers and researchers?
3. Do features of effectiveness differ across distinctive stages of teacher careers and different levels of teaching quality?

Before turning to the findings, we first describe the framework for this review by elaborating some central concepts and eliciting choices we made to delineate the field of interest. The process of searching, selecting and including the literature is outlined in the section on the method. The findings of this review are presented in the next section. First we revisit influential review studies, syntheses and literature covering research until about 2010 from the perspective of the research questions. Then we report upon the findings of the review of selected individual studies since 2010 and describe whether and how these add up to what was already known. In the concluding section we answer the review questions, discuss the findings and the limitations of this study and give thought to the possible implications for the design and implementation of PD interventions and programs and for further research.

## 1.1 Review framework

Kennedy (2016) states that the research topic of professional development is so popular, that there could be thousands of articles written about it every year. Professional development, professional learning, teacher change and related terms can be defined in many different ways and indeed have been reconceptualized through the years (Borko, Jacobs, & Koellner, 2010; Borko, 2004; Desimone, 2009; Opfer & Pedder, 2011; Webster-Wright, 2009). This is also true for other concepts in the questions guiding this study. To keep focus and to narrow the scope for this review we outline the central concepts below. We also clarify the choices we made to delineate the scope of this study.

*Professional development*

Teachers learn and develop in different ways and only a part of their professional growth is deliberately enhanced by PD interventions, programs and activities. The way 'professional development' is conceptualized in the literature can be confusing. Often it is referred to as the activity or facility organized to promote teacher growth. Guskey and Yoon (2009) named 'What Works in Professional Development' as an activity that must be planned and implemented. In her review of literature on quality in teachers' continuing professional development, Caena (2011) uses the following definition: "professional development is defined <…> as related to activities developing an individual's skills, knowledge, expertise and other characteristics as a teacher, excluding Initial Teacher Education" (p. 3). In quite a similar way, Creemers et al. (2013) state that 'professional development' is usually used in a broad sense, frequently encompassing "all types of learning undertaken by teachers beyond the point of their initial training" (p. 3). The problem arising from these definitions is that it is unclear whether PD refers to a (planned) learning activity (intervention, program), the resulting learning process, or to the outcomes of that learning process (development as effect). Day's definition of PD expresses the nature of the process of continuous teacher learning as related to intended outcomes as follows:

> Professional development consists of all natural learning experiences and those conscious and planned activities which are intended to be of direct or indirect benefit to the individual, group or school, which constitute, through these, to the quality of education in the classroom. It is the process by which, alone and with others, teachers review, renew and extend their commitment as change agents to the moral purposes of teaching; and by which they acquire and develop critically the knowledge, skills and emotional intelligence essential to good professional thinking, planning and practice with children, young people and colleagues throughout each phase of their teaching lives. (Day, 1999, p. 27)

Another example of a conceptualization of PD as being the result of the (dynamic) process of teacher learning can be found in Clarke and Hollingsworth (2002) in their conception of 'domains of consequences'.

Point of departure for this review is that teacher learning aimed at promoting the quality of education as described by Day, can be deliberately enhanced by organizing adequate interventions

and programs of learning activities and opportunities to learn for teachers. Professional development is thus considered as the result of teacher professional learning. The focus of this review is on organized learning only[2].

*The scope of professional development interventions and programs*
Following Caena (2011) and Creemers et al. (2013) we exclude programs and activities related to initial teacher education. From a perspective of career development it can be argued that initial teacher education is the first stage of development of teaching competence. In our view, being a professional formally requires being certified to function as a teacher. Besides, initial teacher education and its effectiveness have been object of research and review in many ways and far more frequently than the stages beyond that initial teacher education (Cochran-Smith, Feiman-Nemser, McIntyre, & Demers, 2008)). Given the limitations of time and space for this review, it was considered sound not to include research on initial teacher education.

*Effectiveness of PD interventions: where do we look for evidence?*
Many studies and reviews on the effectiveness of PD interventions and programs conceptualize the relationship between PD interventions and student outcomes in a logic model as depicted in figure 1 (Desimone, 2009).

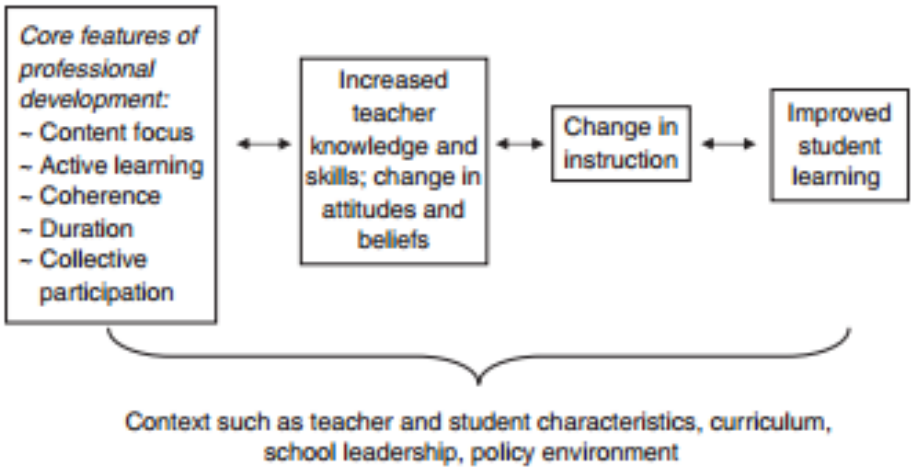



*Figure 1. A model for the input and output of PD interventions (*derived from Desimone, 2009, p185)

Rational models like these (e.g. Yoon, Duncan, Lee, & Shapley, 2008) have been challenged for being too linear, insufficiently meeting the reality of the dynamics of teacher learning and development processes (Clarke & Hollingsworth, 2002; Dall'Alba & Sandberg, 2006; Webster-Wright, 2009). Nevertheless, they can be helpful as an organizing frame when searching for different kinds and levels of effect of PD interventions (e.g. Dunst, Bruder, & Hamby, 2015).

Yoon et al. (2008) describe the logic as follows:

[2] Having said this, does not imply that the authors have the opinion that "unorganized learning" is not worth studying. For an example see Van Waes et al. (2016).

> Professional development affects student achievement through three steps. First, professional development <interventions> enhance(s) teacher knowledge, skills, and motivation. Second, better knowledge, skills, and motivation improve classroom teaching. Third, improved teaching raises student achievement. If one link is weak or missing, better student learning cannot be expected. If a teacher fails to apply new ideas from professional development interventions to classroom instruction, for example, students will not benefit from the teacher's professional development. In other words, the effect of PD interventions on student learning is possible through two mediating outcomes: teachers' learning, and instruction in the classroom (Yoon et al., 2008, p. 3).

Figure 1 makes clear that the student outcomes (improved student learning, or broader, cognitive, affective and behavioral outcomes) are the ultimate aim of PD interventions (Guskey, 2014; Guskey, 2000). This is why some impact studies, reviews and syntheses, in search of what works in PD programs, explicitly focus on student achievement gains only, as effects of PD interventions (e.g. Blank & de las Alas, 2009; Cordingley et al., 2015; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Other studies utilize a broader scope on effectiveness of PD interventions. They also take changes in teachers' knowledge, attitudes and beliefs and changes in classroom practice into account (e.g. van Veen et al., 2010). In terms of effectiveness, teacher professional development and its subsequent consequences can be inferred from changes in teachers' knowledge, attitudes and beliefs (immediate effects), changes in teacher behavior and classroom practices (intermediate effects), changes in student attitudes, motivation, well-being and academic achievement and cultural and organizational changes within teams and schools (all being long term effects). Guskey (2014, 2000) proposes a five level model for evaluating the effects of PD interventions that fits with this broader focus.

In line with the model presented, Guskey (2000) states that PD interventions involve 'those processes, actions and activities designed to enhance professional knowledge, skills and attitudes of teachers so that they might, in turn, improve the learning of students'. Accordingly the author proposes a five level model for evaluating and measuring the effects of PD interventions which is widely referred to when categorizing different possible effects of PD interventions (Guskey, 2014; Guskey, 2000). Level 1 consist of the teachers' reactions to the PD experience or activity, level 2 teachers' learning (referring to new knowledge, new skills, attitudes and dispositions teachers gain), at level 3 organizational support and change are the focus of evaluation, at level 4 the use of teachers new knowledge and skills (changes in professional practice and classroom behavior) are assessed and at level 5 student learning outcomes (cognitive, affective, psychomotor) are measured. Arguably, this model, like the model in figure 1, can be extended school level (learning culture) and even supra school level (retention, staying in the profession, e.g. Ingersoll and Strong (2011).

Establishing a causal link between PD interventions and student achievement, requires meeting very high methodological standards (Yoon et al., 2008). Not surprisingly, systematic reviews like the WWC study of Yoon et al. (2007) only reported about very few studies[3] doing so in a rigorous way, allowing for highly valid and generalizable conclusions. Given the explorative nature of our review regarding career or developmental sensitiveness of PD interventions and possible theoretical underpinnings, we did not limit the scope of effectiveness to student outcomes, thus avoiding the risk of excluding potentially interesting studies. This is why we included studies and reviews mentioning effects at immediate and intermediate levels and long term effects other than student achievement gains.

*Perspectives on professional career stages and the step wise development of teaching quality*
One focus of this review is the possible link between specific features of PD interventions and their effectiveness regarding teachers in different stages of their professional career. The concept of professional career stages can be approached from different perspectives. Some authors (Day & Gu, 2007; Fessler & Christensen, 1992; Fuller, 1969; Huberman, Gronauer, & Marti, 1993) describe the development and stagnation of teacher careers as a sequence of successive personal and professional life phases in terms of effectiveness, motivation and commitment (Louws, 2016). Different phases are generally defined as developmental stages, delineated by years of experience. Others (Berliner, 2004; Berliner, 2001) build on the work of (Dreyfus & Dreyfus, 1980) conceptualing teacher careers as the gradual stage by stage development of professional expertise. In that case, stages of professionalism are described by using labels like novice, advanced beginner, proficient, competent and (adaptive) expert.

Research on teacher effectiveness (Creemers et al., 2013; Muijs et al., 2014; van de Grift et al., 2011), suggests that teacher growth should be conceived as successively reaching higher levels of teaching quality. Research on the development of effective teaching behavior (teaching quality) seems to indicate that clusters of effective teacher behavior can be identified and ranked in terms of increasing complexity (Creemers et al., 2013; Muijs et al., 2014; van de Grift et al., 2011). Teacher learning or teacher professional growth as related to student outcomes may thus be conceived as achieving competence on a certain level and developing competence on the next level, thus gradually becoming more effective as a teacher. What most stage models have in common is the distinction between the phase of initial teacher education, the phase of induction and the phase of the (more) experienced teacher.

While preparing this study, we searched for articles and documents considering the differential impact of effective features of PD interventions on (sub) groups of teachers, classified following either of the perspectives described above. Our searches yielded no studies that both met our quality criteria, and linked effects of PD interventions to very specific target groups in terms of experience or expertise, with the exception of early career teachers. Therefore the review only distinguishes between the induction stage (novices, newly qualified teachers, early career teachers, advanced beginners, 1-4 years) and the experienced stage (4 years +, midcareer teachers, competent, proficient, expert).

[3] Is their study, only nine out of more than 1300 studies identified as potentially addressing the effect of teacher professional development on student achievement in three key content areas.

*The use of theory*

One of the problems regarding many studies on the effectiveness of PD is the absence of theoretical notions used either as guiding principles for the design of PD, or as possible means of explaining how the effectiveness of features of PD interventions fits into a bigger underlying theory about (teacher) learning. It refers to what Timperley et al. (2007) name the black box between particular learning opportunities on their impact on teaching practice.

The first way of using 'theory' is what Wayne, Yoon, Zhu, Cronen, and Garet (2008) refer to as 'the theory of teacher change' and 'the theory of instruction', the two sometimes taken together as 'theory of improvement' underlying PD (van Veen et al., 2010). The theory of teacher change is the intervention's theory about the features of PD that will promote change in teacher knowledge and/or teacher practice, including its theory about the assumed mechanisms through which features of the PD are expected to support teacher learning. It resembles Kennedy's conception of theory of action. A theory of action "…includes two important parts. First, it identifies a central problem of practice that it aims to inform, and second, it devises a pedagogy that will help teachers enact new ideas, translating them into the context of their own practice" (Kennedy, 2016, p. 2). This 'pedagogy' can also be located between the boxes 2 and 3 in figure 1. According to Wayne et al. (2008) this theory of teacher change or action "is not limited to the structural features of the PD, such as its duration and span, but also includes elements of the activities in which teachers are expected to engage during the PD <…> and the intermediate teacher outcomes these activities are expected to foster". The theory of instruction is the "intervention's theory about the links between the specific kinds of teacher knowledge and instruction emphasized in the PD and the expected changes in student achievement" (p. 472). Because of our focus on the relationships between PD and teacher learning, the theory of instruction is out of the scope of this review. When in this review we refer to the question 'What are theoretical assumptions underlying the design of effective interventions and programs as used by designers and researchers?', we refer to the concept of 'theory of teacher change' or 'theory of action' as described here.

*Quality of research: how strong is the evidence?*

Impact studies meeting WWC standards, like the ones included in Yoon et al. (2007), can be said to have technically strong and high quality of evidence, even when qualified as meeting the standards with reservation. They generally make use of experimental or quasi experimental designs with randomized controlled trials to ensure internal validity. Yoon et al. (2008) state that the (rigorous) design must also lead to findings with high degrees of external validity. Studies with less rigorous designs (e.g. case studies, qualitative descriptive studies) must be considered to result in less powerful evidence. One of the limitations in research into the effects of PD intervention is a lack of generalizability of the results. This is another way of looking at the power of the evidence. From this perspective, Borko states that studies can be categorized within three different (developmental) 'phases'.

> Phase 1 research activities focus on an individual professional development program at a single site. Researchers typically study the professional development program, teachers as learners, and the relationships between these two elements of the system. The facilitator and context remain unstudied. In phase 2, researchers study a single professional

> development program enacted by more than one facilitator at more than one site, exploring the relationships among facilitators, the professional development program, and teachers as learners. In phase 3, the research focus broadens to comparing multiple professional development programs, each enacted at multiple sites. Researchers study the relationships among all four elements of a professional development system: facilitator, professional development program, teachers as learners, and context (Borko, 2004, p. 4).

Reconfirming Borko (2004), van Veen et al. (2010) and van Driel et al. (2012) showed that most PD studies can be classified as especially phase 1 studies, some as phase 2 and very rarely as phase 3. However, phase 3 studies result in more powerful evidence in terms of potential impact on research, practice and policy, and therefore will be the main focus of the search of the current review.

# 2. Methodology

The search for empirical research studies strongly focused on finding new evidence for effective PD features after 2010 till 2016 (research question 1). In the selection of relevant studies, an explicit 'theory of change or action' was an additional selection criterion (research question 2). As mentioned earlier, the search did not yield much information on different target groups in terms of experience or expertise, with the exception of early career teachers. For this reason the original search was completed with the results of an available search on beginning teachers and induction (research question 3).

## 2.1 Electronic searching

The first search was directed to the two groups of variables at the end of the process of teaching, the outcomes of interventions, in particular the 1) teacher/classroom practices and on the other hand the stages/phases of teacher expertise/development. For both groups as many as possible keywords were determined. The keyword strings are to be found in the Appendix 2. The keywords within the group are interconnected with "OR" and the two groups are connected with "AND". The search is restricted to the timespan 2006-2016. The search was executed in ERIC, PsycINFO and SocINDEX via EbscoHOST and in Web of Science and SCOPUS. In the different searches various limiters are used to reduce the amount of hits. For more information we refer to the Appendix 1. This search resulted in 826 titles. Examination of the results made clear that some relevant articles weren't covered with these searches. Supplementary searches were done with keywords in the domain of interventions. This additional search resulted in EbscoHOST (ERIC, PsycINFO, and SocINDEX) in 464 titles of which 152 were duplicates of the first searches. In Web of Science the limiters "Web of Science Core Collection" and category "Education educational research" were added. The result was 195 hits of which 165 were duplicates. All searches were imported in Refworks. For further screening the titles were transported to an excel file resulting in 1156 studies. We excluded studies focusing on pre-service teachers. The exclusion resulted in 621 studies for closer scrutiny.

The next step was to select studies for the full text review using the following criteria:

(a) The focus of the study is on professional development of teachers
(b) It is a report of an empirical research; effects of the treatment were measured and reported
(c) An intervention was described

After controlling the interrater reliability on the first 250 studies the criteria were clear. This process was completed and resulted in 282 studies for full text review. Studies on induction (71) were added with the results of a previous search on induction. The next step entailed coding of the quality of the research and the relevance for the research questions. The positively assessed studies included interventions focusing on professional development in a broader setting and explicit outcome measures on teacher beliefs, classroom practices and/or student outcomes in controlled designs. The 32 studies that seemed to satisfy these conditions were finally screened on generalizability. Strict criterions were applied in the final selection. When no experimental design was applied large sample sizes and advanced statistics prevailed: studies with more than 400 participants, applying multilevel statistics accounting for the nested structure and controlling for background variables were selected. Experimental studies with a sample size of at least 100 teachers, teaching in at least 40 different schools met the criteria of the causality claim and representativeness of the sample, and were picked for conclusions. This criterion of 100 teachers is arbitrary; therefore the studies with lower sample sizes are included in the results section. Even though these studies are not included in the conclusion section, they are presented in the results section (if they meet the requirements, but have sample sizes smaller than 100 or in appendices 3 and 4 (if they do not meet the requirements[4]).

The same procedure for the studies on induction resulted in 9 studies. The separate literature search on induction was conducted in the ERIC and PsychINFO bibliographies using the combined keywords "induction" in de abstract, "beginning teacher" in the entire text and since the publication year 2000. The search produced 462 hits. Studies were selected for review on the basis of the following criteria: (a) the source was peer reviewed, (b) an induction arrangement was described (the treatment), and (c) effects of the treatment were measured and reported. This reduced the number of studies to 24. Hits were screened on the research methodology. To determine which publications fit the quality criteria for inclusion, all the abstracts were divided among 6 readers, who classified each publication into one of three groups (yes, no, maybe) based on the probability that it was an empirical study revealing effects and/or effective elements of induction arrangements. Two readers read the full articles that were not clearly outside the scope of the review study (the yes and the maybe groups). When the readers did not agree on inclusion, they argued until agreement was reached. The procedure resulted in a selection of 7 studies meeting the criteria.
Of the 16 studies that were empirical studies with effects and/or effective elements of induction arrangements 4 studies complied with the conditions of the mentioned strict criteria for generalizability. In Appendix 4 the ECT studies that do not comply with these strict criteria are presented.

As the research questions also focused on theory and explanation of effectiveness, the result of the searches was simultaneously checked on relevance for research questions 2 and 3.

[4] This was possible for ECT studies (limited number) but not for the literature covering experienced teachers (too many studies).

Complementary was also the search for relevant reviews on the terrain of professional development. In the initial searches reviews and meta-analyses were set apart. An additional search was made on the search terms meta-analysis and systematic review. Together this resulted in 40 studies. After selection little studies were useful. Most relevant reviews were detected by the snowballing procedure (see 2.2).

## 2.2 Handsearching

All studies of the penultimate selection were used in a snowballing search. References in the 32 research studies and the 19 review studies that resulted after screening of the full text were examined to detect additional studies. Subsequently references of the relevant studies of this first result were examined for more studies. This retrospectively search generated mainly studies from before 2010. Likewise a search was done with the citation-index; insofar these were available in the databases. All literature citing the selected studies was examined on relevance.

# 3 Reviews on the effectiveness of professional development programs

## 3.1 Introduction

Since the turn of the century an increasing number of causal impact studies exclusively linking PD interventions to student outcomes have been carried out. This research aims to reveal structural features of effective PD, like duration, intensity, form and place, but also attempts to shed light on content and pedagogy (methods and activities aimed to facilitate professional development). Some impact studies only focus on the effectiveness of PD interventions from the perspective of student achievement gains. Other studies also focus on effects regarding teacher knowledge, attitudes, beliefs (immediate effects) and teaching practice (intermediate effects). Research on the effectiveness of PD has been reviewed in a systematic way too, since the seminal review of (Kennedy, 1998) on this topic, and more recently authors have attempted to integrate findings and insights from reviews in meta-studies or syntheses.

In this section we explore a selection of reviews and meta-syntheses on the impact of PD interventions and programs for (experienced) teachers in general and for early career teachers in particular. This exploration is done against the background of the three research questions: what do the reviews tell us about the effectiveness, do they contain theoretical underpinnings and/or theoretical interpretation of findings and do findings turn out to be different for distinguished target groups: experienced teachers as opposed to early career teachers. We start by revisiting influential general reviews on the effectiveness of PD program. Desimone (2009) argued that by that time enough empirical evidence existed to suggest that there was a consensus in research on a core set of five critical features of effective professional development programs, namely content focus, active learning, coherence, duration and collective participation.

In paragraph 3.2 we give a brief description of findings from research exclusively focusing on establishing the link between the features of interventions and student outcomes. We do so because, as argued, the ultimate goal of PD is the enhancement of student performance. Furthermore, the findings from these studies can be said to be based on the strongest empirical

evidence. Or, as Wayne et al. (2008) put it: “to have an impact on student achievement of a detectable magnitude”, the immediate impact of the PD intervention must be “quite substantial” (p. 476). Because these impact studies do not take immediate and intermediate effects into account, the processes of teacher learning and enactment in practice are not illuminated. In paragraph 3.3 we pay attention to what is known about the effectiveness of PD interventions for early career teachers (ECTs). These activities are often referred to as induction arrangements, not only aiming at enhancing teaching quality (and subsequently student achievement), but also at successful enculturation in the school organization and retention in school and/or profession. As noted, the results of our searches only allowed to distinguish between early career teachers and experienced teachers. In the final part of this section we reflect on the findings from the perspective of the research questions.

## 3.2 Reviews focusing on effects of professional development of experienced teachers

One of the first systematic reviews on the effectiveness of PD interventions and programs is that of Kennedy (1998). Results of this meta-analysis of the effect sizes of 12 studies in mathematics and science, all taking student achievement into account, reveal that programs with content[5] that mainly focusses on teacher behavior elicit smaller influences on student learning compared to programs with content focusing on teacher knowledge of the subject, on the curriculum and/or on how students learn the subject. The focus of knowledge enhancement of these programs, was not purely about subject matter but also about how students learn that subject matter. Apart from the content, Kennedy found no clear relationship between several other features of the programs and student achievement, concluding that the content (not the form) of a PD program, is critical for its success. This conclusion has been reconfirmed by several other, more recent reviews.

Based on a synthesis of relevant research, Hawley and Valli (1999) propose a series of design principles for effective PD programs. According to the authors effective PD programs (pp137-143):

> 1. have a content focus on what students are to learn and how to address the different problems students may have in learning the material;
>
> 2. are based on analyses of the differences between actual student performance and goals and standards for student learning;
>
> 3. involve teachers in the identification of what they need to learn and in the development of the learning experiences in which they will be involved;
>
> 4. are primarily school-based and built into the day-to-day work of teaching;
>
> 5. are organized around collaborative problem-solving;
>
> 6. are continuous and ongoing, involving follow-up and support for further learning (including support from sources external to the school that can provide necessary resources and new perspectives);

[5] Kennedy’s notion of ‘content’ does not necessarily refer to the school subject-matter content, but to the topics that are dealt with in the program.

7. incorporate evaluation of multiple sources of information on learning outcomes for students and the instruction and other processes that are involved in implementing the lessons learned through professional development;

8. provide opportunities to gain an understanding of the theory underlying the knowledge and skills being learned; and

9. are connected to a comprehensive change process focused on improving student learning.

Knapp (2003) reviewed literature on the effectiveness of PD programs on teaching practice and the learning of students in a narrative way. Knapp identified six features for which there is emerging evidence that they affect teaching and learning positively, which strongly correspond with the ones described by Hawlet and Valli (1999). Knapp argues that, given what is known about adult learning, there is good reason to accept that evidence as theoretically sound. In his review Knapp discusses some theoretical insights on which we will elaborate in section 5.

Our search identified one review (Broad & Evans, 2006) explicitly approaching the theme from the perspective of career stages and ongoing development (research question 3). Their literature review on the content and delivery of PD programs for experienced teachers (beyond the stage of induction) addresses themes including teachers' stages and pathways, delivery methods and practices, effective PD programs and a syntheses of characteristics. The findings on the last themes do not significantly differ from those in the other reviews and syntheses described here. Broad and Evans critically analyzed the usefulness of general stage theories from the perspective of applying them for decision making on the planning and implementation of PD programs. In their review, they incorporated insights resulting from reviews from Dall'Alba and Sandberg (2006) and Darling-Hammond, Bransford, LePage, Hammerness, and Duffy (2005). They conclude that experience and development of expertise and sophistication in knowledge and skills are integrative and interactive and, for several reasons, cannot be captured by the traditional stage theories.

In a systematic review of research on the relationship between PD interventions and programs and improvements in student learning, Yoon et al. (2007) only included studies with experimental and quasi experimental designs. The 9 studies included in the review all regarded primary education in three 'core subjects'. Yoon and colleagues found that successful PD interventions focused on content knowledge or pedagogy or both, provided follow-up support, included at least 30 contact hours, all included some form of workshops or summer courses (contrary to what they expected) and ideas behind theories of instruction were brought in by outside experts. Programs providing more time were not more effective per se, but sustained follow-up proved to be crucial. Finally the reviewers did not find 'best practices' when it comes to learning. Good practices were varied and carefully adapted to specific contents, processes and contexts.

Timperley et al. (2007) conducted a comprehensive "best evidence synthesis" on teacher professional learning and development. The effective features identified by Timperley et al. (2007) regard the content of the programs, the activities organized to promote teacher learning, the learning processes themselves and the influence of the context. The findings are similar to those from reviews described before, but go beyond that from the perspective of comprehensiveness and nuance. This is why we describe the findings of this synthesis in more detail. Regarding the *content* of

effective programs, they conclude that the integration of theory and practice is a key feature. Integration of pedagogical content knowledge, assessment of learning and information about how students learn the subject was an essential part of successful programs. What characterized content in strong programs too, was (1) that there were clear links between teaching and learning and/or the student-teacher relationship established; (2) that assessment was used to focus teaching and the enhancement of self-regulation; and (3) that sustainability was enhanced by promoting teachers' in depth understanding of theory to inform educational decision making and to provide teachers with the skills of inquiry to judge the impact of teaching and learning and to identify next teaching steps. *Activities* that were constructed to promote professional learning were characterized by variety and by alignment of content and form. The content conveyed was more important than the activity itself, professional instruction was sequenced[6], understandings were discussed and in all activities student perspective was maintained. Regarding *learning processes,* they notice that even in powerful programs substantive change is difficult, new understandings may accommodate with existing conceptual frameworks or may not (thus creating cognitive dissonance). Furthermore the authors remark that in few interventions teachers learned to regulate their own and others' learning, which appeared to be crucial for sustainability. Finally on *context*, the authors found that extended time for learning opportunities and the use of external expertise were necessary, but not sufficient, that teachers' engagement in learning at some point was more important than their volunteering, that the opportunities to engage in professional communities of practice were more important than the place, that there was consistency with wider trends in policy and research and that there was active school leadership. Like Knapp (2003) the synthesis of Timperley et al. (2007) refers to theoretical insights[7] on (teacher) learning that offer possibilities to explore the question why identified core features of PD programs are effective from the perspective of teacher change and student learning.

Cordingley, Bell, Isham, Evans, & Firth (2007) were specifically interested in the role specialists play in effective PD interventions and settings. They conducted an in-depth review on 19 empirical studies that met rigorous criteria and showed medium or high weight evidence. All studies reported on effects for both teachers (teacher practice) and students (learning and achievement, and affective development, including attitudes to learning and self-esteem). The findings suggest that teacher practice tended to change as a result of the input of specialists and teacher learning regarding topics like teaching strategies, learning theories, the use of technology, educational policy and subject knowledge. Specialists involvement in learning activities concerned modelling, workshops, observation, feedback, coaching, and planned and informal meetings for discussion. In more than half of the settings specialists observed teachers (and students) and provided feedback. Other activities included discussing student needs and collaboratively examining student work. Peer support proved to be an important feature, while specialists encouraged teachers to take the lead to some degree.

According to Scher and O'Reilly (2009) the results of the study of Timperley et al. (2007) should in part be interpreted with caution, because the study does not systematically examine overall effects. Scher et al. conducted a quantitative meta-analysis, in which they compared programs that focused

[6] Typically this involved: engagement of participants, instruction in the main theory, opportunities to transfer theory into practice and to deepen the understanding of theory

[7] These are discussed in section 5

on subject matter with those that focused on pedagogy (again math and science). Regarding the focus of the interventions they concluded that programs that include both content and pedagogy as part of their intervention have a larger positive impact on student achievement. This is an empirical affirmation of the conclusion regarding the importance of content drawn by Kennedy (1998). They conclude that although earlier findings on other core features are supported, the evidence still is thin.

Blank and de las Alas (2009) conducted a meta-analysis on 16 impact studies within math and science, all having an experimental or quasi experimental design. They were specifically interested in content focused PD interventions and their impact on student learning. Results showed cross study evidence of impact of PD on student learning. The authors found a relationship between some key characteristics of the design of PD programs and student learning gains. These included follow up steps in schools, active learning methods, collective participation and 'substantive attention on how students learn specific content'. The authors suggest the evidence makes very clear that effective programs 'focused on helping teachers improve their knowledge on how students learn in the specific subject area, how to teach the subject with effective strategies, and the important connections between the subject content and appropriate pedagogy so that students will best learn.'

The NWO-review study of van Veen et al. (2010) was a broad exploration and summary of effective professional development in the Netherlands, written in Dutch for a broad audience from researchers to teachers and policy makers, strongly influencing the debate on PD policy in the Netherlands. Next to a review of effective features, confirming Timperley et al. (2007) and Desimone (2009) findings, also the organizational conditions were analyzed that support professional development. A main conclusion in this review was that most schools are not equipped for teachers to learn while most studies point out that learning at the workplace, especially in the own classroom, is very effective. In a follow-up of this review (van Driel et al., 2012), studies in the field of science education were selected and analyzed using the frame of Borko's (2004) three phases, confirming the previous consensus on effective features and showing the prominent issue of Pedagogical content knowledge in PD in the field of science education.

Other general reviews on the effectiveness of PD programs (Avalos, 2011; Broad & Evans, 2006; Caena, 2011; Cordingley et al., 2015; van Driel et al., 2012; van Veen et al., 2010; Wei et al., 2009) also largely confirm the findings by Timperley et al. (2007) and Desimone (2009), some of them being more cautious than others, because many studies lack generalizability.

In a comprehensive literature study by Gersten, Taylor, Keys, Rolfhus, and Newman-Gonchar (2014) the effectiveness of PD programs for K-12 mathematics teachers on student proficiency in this subject was examined. Only 5 studies (4 randomized control trials, 1 quasi experimental design) were included in the final review. Significant positive effects were found in two studies, one approach consisting of an intensive math content course on mathematics, accompanied by follow-up workshop, the other being a lesson study approach. Limited effects were found in one study on cognitively guided instruction (CGI). Two other approaches showed no discernable effects on learning gains. Features of the three (modestly) effective approaches may be described as the employment and collaboration of (external) experts and the combination of a focus on content knowledge and pedagogy and a string focus on student learning and mastery of the subject.

A recent meta-analysis by Dunst et al. (2015) suggests that the evidence base for the features mentioned is growing stronger. The majority of studies they included, were experimental, quasi experimental and/or had pretest-posttest designs. They identified core features and characteristics important for the effectiveness of PD interventions. They hypothesized that PD interventions and programs that included the majority of these features would be associated with positive teacher and student outcomes. The patterns of results of systematic reviews and syntheses taken together in this meta-synthesis provided strong evidence for the relationship between the core characteristics and teacher and student outcomes. According to the authors, the fact that the results were the same or similar in the different types of research syntheses for different types of practices, indicates that there is growing evidence the necessary, (but not the sufficient) conditions for PD programs to be effective.

In a systematic review of empirical research, Evens, Elen and Depaepe (2015) focused on the effectiveness of interventions designed to promote teachers' pedagogical content knowledge (PCK), which is generally accepted as positively impacting teaching quality and student learning (already by Kennedy, 1998, and confirmed by all reviews mentioned above). To answer the effectiveness question Evens et al. limited their analysis to 37 studies that at least had a pre-test post-test design, 16 of them characterized as being quantitative. Most studies were conducted in primary or secondary education and almost all of them focused on science or mathematics. Regarding the design, the authors concentrated on PCK sources (teaching experience, PCK-courses, disciplinary knowledge, observation, cooperation with colleagues, and reflection), on location (on-site/in school or of-site) and main actors. Of the 16 quantitative studies, 13 reported positive effects on growth of PCK. So did all the mixed-method and qualitative studies. With regard to the effective features of the interventions Evens et al state that, due to the fact that the interventions were very similar and often contained a mix of PCK sources, these features could not be isolated well. Nevertheless the authors describe some general tendencies. These include the importance of individual and collective reflection that induces higher order thinking in order for teachers to understand their own learning; cooperation of teachers, when disciplinary knowledge was addressed; this was only effective in combination with PCK. Effective interventions took place of-site or in a combination of of-site and on-site activities, and were always guided by experts. Finally, the authors caution that the evidence is weak because most of the studies included did not meet the conditions of a (quasi)experimental design.

Zwart, Smit and Admiraal's (2015) NWO-review focuses on teacher research as a way of professional development, exploring how teacher research is conducted. In the literature, it is unclear how teacher research is characterized and how this type of research differs from other types of educational research or good teaching. This literature review aimed at providing insight in the characteristics and benefits of teacher research in the context of primary and secondary education, taking into account the various goals of practitioner research. Regularly used educational research databases were systematically searched for articles on teacher research published between 2009 and 2012. Based on an analysis of 160 articles, four most common types of teacher research were discussed: Action research, Lesson-study, Self- study and Design-based research. In addition, examples of these practices were provided. Also, conditions were discussed for rendering teacher research more effectively. The results showed that primary and secondary teachers conduct research mainly to improve their own individual practice or to develop as a professional in a more general

sense. The extent to which this form of PD was effective or how it was effective was not part of this review.

In a recent review of research of Kennedy (2016), the effectiveness of PD programs is approached from a different angle than the reviews mentioned above. Rather than focusing on the array of the different design features PD interventions rely on, in her review Kennedy categorizes 28 studies using rigorous research standards according to their underlying 'theories of action'. In Kennedy's view, a theory of action consists of two parts: "(1) it identifies a central problem of practice that it aims to inform and (2) it devises a pedagogy that will help teachers enact new ideas, translating them into the context of their own practice, facilitate enactment of their ideas" (p. 2). On the first dimension, Kennedy distinguishes four central problems of practice related to teaching and learning: portraying curriculum content, contain student behavior, enlist student participation and finding ways to expose student thinking. On the dimension of pedagogy Kennedy characterizes four different approaches: detailed prescription, focus on strategies and defining goals, enhancing insights ('aha- Erlebnisse) and exposure to a body of knowledge. After correcting for teacher motivation to learn, Kennedy calculated (estimates of) effect sizes from the different programs. She then classified the programs according to the two dimensions of their theory of action. Results suggest that the effectiveness of PD programs is not dependent on the nature of the teaching and learning problem or challenge that is at the heart of the program. Programs on all kinds of teaching problems (not only on portraying curriculum using appropriate subject knowledge and PCK) tended to be equally effective from the point of view of student achievement. On the other hand, the kind of pedagogy used seemed to matter: pedagogies focusing on strategies and defining goals and enhancing teachers' (theoretical) appeared to be more effective than insight than those consisting of detailed prescription or (only) strongly focusing on the exposure to a knowledge base.

## 3.3 Reviews on the effectiveness of programs for early career teachers (ECTs)

Ingersoll & Strong (2011) reviewed the literature since the mid-1980s. They included 15 (un)published empirical studies of induction that compared outcome data from both participants and non-participants in particular induction components, activities, or programs. The authors conclude that most of the studies reviewed provide empirical support for the claim that induction for ECTs and teacher mentoring programs in particular have a positive impact on three sets of outcomes: teacher commitment and retention, teacher classroom instructional practices, and student achievement. Most of the studies reviewed showed that ECTs who participated in some kind of induction performed better at various aspects of teaching, such as keeping students on task, developing workable lesson plans, using effective student questioning practices, adjusting classroom activities to meet students' interests, maintaining a positive classroom atmosphere, and demonstrating successful classroom management. Almost all of the studies on student achievement show that students of ECTs who participated in induction have higher scores, or gains, on academic achievement tests. There were exceptions to this overall pattern, in particular the study of Glazerman et al. (2010). The authors discuss the need for the following kind of research with regard to effective features of induction arrangements: a) the content of the program, b) the duration and

intensity of the program, c) the relative costs and benefits of the program and d) the impact of the context. The *content of the program* should be aligned with the goals and the desired outcome should be well-defined. If the aim of induction is to increase the quality of ECT's teaching, trade-offs and contradicting goals should be avoided; e.g. if a teacher needs to develop the skill of engaging students in higher-order inquiry does not fit with the additional goal to teach standardized test taking. It is still unclear *how long or how intense* induction arrangements need to be. The thresholds below which and above which induction is of little value have not been identified. With regard to *the relative costs and benefits of the program* almost no empirical work has been done. It is not clear if content and duration of effective induction are similar across *different settings*.

Long et al. (2012) reviewed 49 empirical studies on the relationship between induction and mentoring and early career attrition and retention. The authors conclude that the effect of induction (including mentoring) programs on ECT retention is unclear. Multiple factors influence a teacher's decision to stay in or leave the profession. Studies were found showing the quality of teaching may be impacted with induction (including mentoring) but links to retention were often not made or were tenuous. Two studies focusing on systemic and structured observations provoke questions around the common assumption in mentoring and induction programs that observation is always valuable. School culture and the working context seem to be prevailing with regard to retention. School cultures which are highly collaborative, value all teachers' knowledge, which focus on what is most educative for students, and which see students as the responsibility of the whole school, appeared most successful in retaining ECTs. Principals seem to have a very important role in the success of induction programs, setting a tone for collegiality amongst all staff. School cultures supportive of an integrated approach rather than those oriented toward supporting veteran or ECTs were most successful in retaining ECTs.

Schaefer, Long and Clandinin's (2012) review concerns ECT attrition and retention from 1999 to 2010. They summarize the research on attrition as it relates to individual (e.g. burnout, resilience) and contextual factors (e.g. support, PD). Many causes have been identified for retention (e.g. resilience, personal characteristics, personal factors, contextual factors, incentives and access to resources, collaboration: ECTs are more likely to stay if they are part of an integrated professional culture, student issues and teacher education collaboration with schools). They suggest the need to shift the conversation from one focused only on retaining teachers, toward a conversation about sustaining teachers throughout their careers. The authors conclude that working alongside ECTs and working from a narrative conceptualization of identity and school contexts offers a promising way to understand what sustains ECTs. This may offer the possibility of new insights about the kinds of continuing spaces needed on school landscapes to sustain and retain ECTs.

Additional review studies concerning the support of ECTs are presented in Appendix 3. These reviews had less transparent search criteria.

# 4. Recent Empirical studies on the effectiveness of PD

## 4.1 Early career teachers

The aim of this part of the review is to update the body of knowledge regarding the effective features of interventions for early career teachers as from 2010. This cannot be done without mentioning the ground-breaking work of Ingersoll and Smith (2004). Their work has had a great impact on the conceptualization and development and research of the effects of induction arrangements all over the world. Their contribution to the field is especially interesting because they unravelled relative contributions of different features of induction arrangements on teacher retention. Different types and components of induction were available for this study, including mentoring programs, collective group activities, and the provision of extra resources and reduced workload. They ran secondary analyses on data gathered in a large-scale survey study comprising a sample of 3 235 early career elementary and secondary teachers. The surveys (Schools and Staffing) included items designed to elicit information on the range of possible induction and mentoring supports. This data was linked to the Teacher Follow-up Survey data which included all teacher turnover or departures, including those who move to teaching jobs in other schools ("movers") and those who leave the occupation all together ("leavers"). The large-scale study made it possible to statistically control for various teacher and school characteristics. Their central research question was whether receiving support matters for teacher retention. By means of multi-nominal logistic regression analyses, they found support for the association between receiving support and the likelihood of early career teachers moving or leaving at the end of their first year on the job. The strongest induction features contributing to retention were found to be: having a mentor from the same field, having common planning time with other teachers in the same subject, having regular scheduled collaboration with other teachers, and being part of an external network of teachers. The weakest factors contributing to retention were found to be: a reduced teaching schedule, reduced number of preparations, and extra classroom assistance. Furthermore, they found that as the number of components in the packages increased, both the number of teachers receiving the package and the probability of their turnover decreased. The contextual influence of induction effects became apparent in their disaggregated analysis. They revealed that the retention effects were very strong in low-poverty schools and that there were zero retention effects in high poverty schools.

Two longitudinal studies have been found exploring the effects of a number on intervention features acting in concert (Glazerman et al., 2010; Helms-Lorenz et al., 2015). In primary education (Glazerman et al., 2010) revealed positive effects of comprehensive induction arrangements on student achievement in mathematics (effect size 0,20 on average) after a period of three years. No effects however concerning retention, teacher mobility, teacher attitudes towards job satisfaction, and feelings of preparedness were obtained in this study. The effectiveness of the different features of the intervention was not the focus of this study. The intervention entailed carefully selected, trained, full-time mentors. The program provided a curriculum of intensive and structured support for early career teachers, including orientation, professional development opportunities, and weekly meetings with mentors; a focus on instruction, with opportunities for novice teachers to observe experienced teachers; formative assessment tools that permit evaluation of practice on an ongoing basis and require observations and constructive feedback; and outreach to district- and school-based administrators to educate them about program goals and to garner their systemic support for the program. Results show that compared to business-as-usual induction programs, two years of

comprehensive induction can boost student achievement. For teachers who received one year of comprehensive induction, however, there was no impact. Stanulis, Little and Wibbens, (2012) (see Appendix 3) were challenged by these results. They argued that if the quality of the mentoring would be improved, the effects on student achievement would increase. Even though their intervention study (which was based on thorough theoretical underpinnings) does not adhere to our strict selection criteria it does reveal positive results.

(Helms-Lorenz et al. (2015) reported the package effect as a whole, as well as the impact of different support features separately. In this study 71 schools with 338 early career secondary education teachers were randomly allocated to an experimental or a control group. The experimental schools implemented induction arrangements. The teachers in the control group received some induction into the profession, though for a maximum of 1 year. In this longitudinal experimental study the experimental group followed a 3-year induction program under controlled conditions arranged by the schools, involving workload reduction, activities to enhance enculturation in the school and its policy, professional development planning, classroom observations and mentoring, workshops and seminars. Specially trained university supervisors guided and supported the schools with induction, which involved the development of programs following guidelines derived from Smith and Ingersoll (2004). This study revealed that ECTs in the experimental condition showed greater improvements in teaching skills (perceived by their students) compared to the ECTs in the control group (effect size .75 in the experimental group and .46 in the control group). Attrition effects were measured by comparing the rates at which ECTs in the control and experimental groups left the teaching profession. Three years later, 14% of the control group and 12% of the experimental group had left. Leaving the profession could be explained by a lack of certification and low initial teaching skill levels. A multilevel regression analysis revealed which characteristics of ECTs and which elements of the induction program explained attrition. Approximately 55% of teacher attrition within three years can be explained by a lack of certification, low levels of teaching skills at the beginning of the career and a low degree of enculturation in the school. A multilevel regression analysis revealed which personal characteristics and elements in the induction program explained the development in teaching skills. Classroom observations have a positive impact on the development of the ECT's skills. Workload reduction has a negative impact on the teaching skills in year 3.

In another publication the same research group modelled the longitudinal development of early career teaching skills (Maulana et al., 2015). The theoretical underpinning is the self-determination theory of (Ryan & Deci, 2000): the fundamental needs of teachers (autonomy, competence and relatedness) are used to specify the required conditions for psychological and professional growth, integrity, and well-being. Teacher autonomy is supported by taking the individual teacher's concerns into account regarding the conditions needed to thrive. The intervention consisted of (1) considering the need to reduce (or increase) the workload and by attending to ECT concerns in meetings and mentoring sessions. Teacher autonomy is nourished by increasing the competence level, by (2) supporting effective teaching behaviour in the classroom (this is enhanced by working together with colleagues when preparing lessons, observing each other, etc.) and by (3) stimulating professional development (developing transparent induction goals and assessment procedures, clear human resource management and policy). The need for relatedness calls for (4) facilitating activities to get familiar with one's own students, school policy, school rules, and with colleagues. The development tasks proposed by Feiman-Nemser (2001) were used as leading principle to streamline activities into

a coherent arrangement encompassing three years. These principles are: (1) gaining local knowledge of students, curriculum, and school context, (2) designing responsive curriculum and instruction, (3) enacting an early career repertoire in purposeful ways, (4) developing a professional identity and, (5) learning in and from practice. The guidelines provided by Ingersoll et al. (2004) were used as a framework for the development of induction arrangements. ECTs were to be provided with a mentor, ideally from the same subject field and available on a daily basis for timely support, opportunities for common planning time with other teachers in the same subject, opportunities for regular scheduled collaboration with other teachers, opportunities for being part of an external network of teachers, opportunities for structured peer coaching, opportunities for classroom observations with feedback, workshops, opportunities to work with professional development plans, and the opportunity to work with portfolio's. As the number of components in the induction arrangements increases, the probability of retention increases. Workload reduction was added to the arrangements as many studies reveal that ECTs need extra time to reflect, collaborate and to observe others (Algozzine, Gretes, Queen, & Cowan-Hathcock, 2007; Lazovsky & Reichenberg, 2006; and Nielsen, Barry, & Addison, 2007). This study revealed that students perceived more rapid teaching quality increases over time in the experimental condition compared to the control condition. Certification status, gender, and induction programs explained the differences (and changes) in perceived teaching quality.

Kang and Berliner (2012) examined the relationship of teacher induction activities to teacher retention using data from the Schools and Staffing Survey (SASS) and Teacher Follow-up Survey (TFS). The sample only included first-year teachers (n=1556) whose turnover was voluntary and avoidable. Multinomial logistic regression analysis revealed, after controlling for both teacher-level and school-level characteristics, that three of the examined induction activities were beneficial in significantly reducing turnover rates (in terms of moving to another school) for ECTs: participation in seminars or classes (rrr[8] = .67, p = .03), common planning time with teachers in the same subject (rrr = .73, p = .09), and receiving extra classroom assistance, such as the provision of teacher aides (rrr D .57, p D .00). The data did not allow for judgments to be made about the quality or depth of the induction activities, nor were the activities well described.

In Table 1 the ECT studies meeting the strict selection criteria are presented. The last column indicates the sample size which determines the selection of studies for generalizable conclusions.

[8] Relative Risk ratio (rrr) is calculated after controlling for teacher and school characteristics. Rrr reflects the relative risk of moving or leaving as opposed to staying

Table 1: Summary of the ECT studies that meet the strict selection criteria

| Author, year | Secondary or primary | Student population | Specific subject | Effective features | Effect | Generalizability |
|---|---|---|---|---|---|---|
| Ingersoll & Smith, 2004 | Not clear | Full range | all | Having a mentor from the same field, having common planning time with other teachers in the same subject, having regular scheduled collaboration with other teachers, and being part of an external network of teachers. The weakest factors contributing to retention were found to be: a reduced teaching schedule, reduced number of preparations, and extra classroom assistance. Furthermore, as the number of components in the packages increased, both the number of teachers receiving the package and the probability of their turnover decreased. The contextual influence of induction effects became apparent in their disaggregated analysis. They revealed that the retention effects were very strong in low-poverty schools and that there were zero retention effects in high poverty schools. | + retention (very strong effect in low poverty schools and no effect in high-poverty schools)<br><br>+ retention (increasing number of intervention elements increases likelihood of retention) | ++ (N=3235) |
| Glazerman et al., (2010) | Primary | Majority high poverty | Math and reading | Not specified | + (math student achievement after 3 years of induction)<br>0 (retention, teacher mobility, teacher attitudes towards job satisfaction, and feelings of preparedness) | ++ (N= |
| Helms-Lorenz, van de Grift & Maulana, 2015* | Secondary | regular | all | Leaving the profession could be explained by a lack of certification and low initial teaching skill levels<br><br>Classroom observations have a positive impact on the development of the ECT's skills. Workload reduction has a negative impact on the teaching skills in year 3 | + (small retention effect)<br>+ (teacher behavior.32) | ++ (N= 276) |
| Maulana, Helms-Lorenz & Van de Grift, 2015* | Secondary | regular | | Certification status, gender, and induction programs explained the differences (and changes) in perceived teaching quality | + (teaching quality as perceived by students) | ++ (N=276) |
| Kang & Berliner (2012) | | | | Survey study. Seminars or classes, common planning time with teachers in the same subject, and receiving extra classroom assistance (such as the provision of teacher aides) | + (retention) | + (N=1556) |

Note: * Helms-Lorenz et al. (2015) and Maulana et al. (2015) were based on the same sample

## 4.2 Experienced teachers

The aim of this part of the review is to update the body of knowledge regarding the effective features of interventions for experienced teachers as from 2010. All the included studies aimed to improve student achievement and to investigate the impact on the intervention on the expected achievement gains. The studies differ in content and approach: some aim to improve a) instructional behavior or teacher-student interactions and student achievement at the same time and others aim to improve b) teacher knowledge or teacher efficacy and student achievement. A number of studies focus on finding the best approach by c) comparing the impact of different kinds of interventions on student achievement in a single study. In the following sections we discuss these studies by grouping them according to the intended effect as well as by the strength of the generalizability of the results. As discussed in the method section the strength of the generalizability of the findings depends of the study's research design and the sample size.

*a. Interventions focusing on the improvement of instructional behavior or teacher-student interaction and student achievement*

Wasik et al. (2011) studied the effect of a teacher coaching program for language and literacy in a randomized control study in primary education[9]. The program had two components: (a) intensive and ongoing staff development and (b) books, materials, and lesson plans that support the development of children's language and literacy. Thirty teachers participated in this study in 3 Head Start centers. Two centers were randomly assigned to the intervention condition and one center to the control condition. There were no significant differences between the groups, most factors were evenly spread. Of the in total 541 children 2/3 were in the intervention program. Nearly all of the children were African American. Four literacy coaches provided the teacher training for the intervention teachers. All intervention teachers participated in a summer literacy institute where teachers were familiarized with the goals of the project and the training and coaching procedures. After that there were 3-4 weeks training cycles. Each cycle addressed one aspect of the five modules of the program, interactive book reading, guiding conversations across the curriculum, phono logical sensitivity, alphabet knowledge, and writing. The cycle started with a 3-hr group training dedicated to explaining research regarding how a specific classroom practice could support later reading acquisition and comprehension, to specific strategies and activities for implementing effective instruction and to reflection and planning own practices. Later on the coaches modelled the target behavior in each teacher's classroom and observed teachers' use of the new strategy. In addition to the training cycles each teacher was videotaped conducting a book reading, a circle-time activity and center activities. The coach and teacher reviewed and discussed the tape together. Teacher quality was measured with observation instruments of literacy environment in the classroom and the quality of the teachers' instruction. The intervention implementation was measured by tapping the degree to which teachers' classroom practices were consistent with the strategies emphasized in the group training and coaching. Furthermore video recordings were used to code effective book readings by observing child talk before, during, and after the book reading. Teachers completed a background questionnaire on education, experience and certification. Achievement measures included student's vocabulary, pre-literacy skills, phonological skills and alphabet knowledge.

[9] Part of the Head Start Project

Because no meaningful or significant variance on teacher practice variables and fidelity outcomes was shared within centers multiple regression analysis was used instead of multilevel models. On the whole, classroom quality was stronger in the intervention classrooms. The intervention proved to be related to improvements in the classroom language and pre-literacy environment, particularly in the area of writing materials. Intervention teachers also provided higher quality literacy instruction at the end of the year relative to comparison teachers ($\beta$ =0.50, p<.01), demonstrating a moderate to large advantage over their peers of one half of a standard deviation or approximately one full point (on a 7-point scale). Overall, the model explained 44% of the variance in spring instructional quality scores... The largest effect was observed for the language modeling subscale of the test equivalent to 1.39 points ($\beta = 0.54$, p < .01). Although the intervention had caused positive changes in teachers' classroom quality there was quite a lot variation across intervention teachers in their practices. One teacher had a low score, ten teachers medium and 3 teachers high after a full year of training. The children outcomes were all separately analyzed with a two-level model, children nested within classrooms. The contribution of the intervention status to the child's spring outcome was examined controlling for child-level and classroom covariates. For the vocabulary measure participation in the intervention explained 17% of the variance between classrooms. Children in the intervention outperformed their peers in the spring controlling for children's fall vocabulary skills, as well as their age, gender, and disability status and their teacher's education and classroom exposure to public preschool. Phonological sensitivity explained 26%; the intervention made a significant contribution to children's skills. Findings from the alphabetical knowledge are less positive: the children whose teachers participated in the intervention did not learn more letters over the course of the year than their peers. After substituting the intervention variable with classroom quality in the multilevel models the results showed that higher scores on the literacy environment as well as well as higher scores on the classroom instructional quality were linked to stronger vocabulary learning among children over the course of the year. Together children's gains were at least partly linked to the stronger quality in the intervention classrooms, particularly in the use of writing materials and activities and the language environment. In addition, nuanced findings suggest that, of specific classroom experiences, children in the intervention talked significantly more than children in the control classrooms ($\beta = 0.45$, p < .02), beyond the effects of teacher and child background factors. Intervention teachers allowed children more opportunities to initiate the use of language during activities such as book reading.

The Literacy Collaborative (LC), a schoolwide reform program, was designed to improve elementary children's reading, writing, and language skills. Primarily through school-based one-on-one coaching of teachers. A multi cohort, longitudinal quasi-experimental research on literacy-coaching professional development was conducted in kindergarten and elementary school (Biancarosa, Bryk, & Dexter, 2010). Teachers selected by their schools, to lead local instructional improvement, received an intensive, graduate-level training program of a year, including learning how to lead a PD-course to introduce theories and instructional practices to teachers and how to use one-on-one coaching as a mechanism to support individual professional growth and development. After their training year, coaches spent approximately half of their time providing PD and coaching to their school colleagues. This started with teachers' participation in a 40-hour course led by the coach. Ongoing courses after the initial year 10–12 hours of PD were offered annually. The most important component of the program was one-on-one working with teachers in their classrooms: observing, modeling, and catalyzing teachers' development toward more expert practice. Participants in this

longitudinal research program were 287 teachers in kindergarten through second-grade classrooms in 17 study schools at some point during the study's 4 years. During each year of the study, approximately 1,150 students were assessed in fall and spring in each grade level from kindergarten through second grade. After the baseline year the effects of the following 3 years of implementation on student literacy were examined. The expected achievement growth for a child exposed to the average instructional conditions in the school was based on the data in the first year where coaches were trained, but no activities took place in the school. Then the observed growth trajectories in the 3 years of implementation were compared to these latent growth trajectories, in the baseline condition. The differences between the expected and observed outcomes are the value-added effects. A mix of reading assessments was used to measure students' literacy learning. The various scales were scaled together using Rasch modeling. After looking at the trends in the mean Rasch literacy development scores for K–2 students in the final analytic sample by grade, semester (i.e., fall or spring), and study year, the most suited hierarchical, crossed- level, value-added-effects model was determined. A two-level model for individual growth in achievement over time, and the two-level model of the value added that each teacher and school contributes to student learning in each particular year were applied to the data. The models estimated both the average value added by LC in each year of implementation and random value-added effects associated with each teacher and with each school year of implementation. The core evidence for LC effects consists of comparing learning gains in each teacher's classroom during each year of program implementation to the gains in that same teacher's classroom during the baseline year. The average LC Value-Added effects were positive. Results demonstrated significant gains in student literacy learning beginning in the first year of implementation and that the effect's magnitude grew larger during each subsequent year of implementation. There was a 16% increase in learning compared to the average baseline growth rate in the first year of implementation, and 28% increase in the second year and a 32% increase in the third year of implementation. Program effects maintained over the subsequent summers. In almost every case the estimated individual school effects were positive. The variance in school-level effects increased across the 3 years of implementation. The vast majority of teachers in most of the participating schools showed substantial value-added effects by the end of the study.

The American Institutes of Research (AIR) published two research reports on the Middle School Mathematics Professional Development Impact Study (Garet et al., 2010; Garet et al., 2011). Randomized controlled study examined the impact of a professional development program for Middle School teachers that was designed to improve the teaching of rational number topics. The program consisted of three 8-hour sessions of instruction on pedagogy and content knowledge during a summer institute and 5 seminars during the following school year. In the week following each of the seminars classroom coaching was provided to assist teachers in applying the new strategy. Seventy seven schools in 12 districts participating in the study were randomly assigned to receive the intensive PD activities or only the PD activities normally provided by the district. In the treatment schools all the $7^{th}$ grade teachers teaching at least one regular seventh-grade mathematics class were offered the intense PD during the first year of implementation. In 6 of the 12 districts this was continued in the second year. The study focused on the impact of the PD program on teacher knowledge, teacher instructional practices and on student achievement in rational number topics. Teacher knowledge was measured for all treatment and control teachers with a specially constructed teacher knowledge test. The test was administered before and after the PD program. During year one classroom observations were conducted for each teacher to measure instructional practice. The

observation delivered three measures the frequency with which the teacher employs key behaviors that elicit student thinking, the teacher using representations and the teacher focusing on mathematical reasoning. Student-level math achievement was measured by a computer-adaptive rational number test restricted to positive rational number content. Background data and participation of teachers were gathered. Observing the summer institute and seminars and reviewing logs maintained by coaches delivered implementation data. Data were analyzed with a three-level model for the impact of professional development on student achievement and two-level models (teachers nested within schools).The PD program did not have a statistically significant impact on overall teacher knowledge. In the first year both teacher groups answered on average over 50% of the items correct with a difference of 4, 6 % in favor of the treatment group. In the second year the teachers on average answered about 75% of items correct and the difference was only 1% in favor of the treatment group. Impact on teachers ' instructional practices was statistically significant in the first year on the frequency with which teachers engaged in activities that elicited student thinking (3,45 times per hour compared with 2,42 times per hour). But the differences on the other teacher practices were smaller and not significant. In the second year of the study classroom practice was not measured anymore. In both the first and the second year of the study the PD program had no impact (effect size = –0.01, p = 0.94 in the second year) on student achievement in rational numbers.

In later years the American Institutes of Research (AIR) published another research report on content-intensive PD on teachers' math content knowledge in elementary-middle schools (Garet et al., 2016). This study examines the impact of content-intensive PD on teachers' math content knowledge, instructional practice, and their students' achievement. The PD consisted of three elements, an intensive workshop focused on deepening teachers' knowledge of grades K-8 mathematics, 5 collaborative meetings during the schoolyear thereafter and a series of three one-on-one coaching sessions with video feedback where teachers 'lessons were observed and critiqued. The study was based on 221 grade 4 teachers of 94 schools in six districts and five states. The teachers were randomly assigned within their school to a treatment group and a control group. Due to leavings and incomplete data the final dataset includes 165 teachers. Implementation of the PD was followed with activity logs, video-recorded group sessions that were coded on the engaged activities and attendance records of the teachers. On three occasions in the PD program, teacher knowledge was measured with an assessment on five mathematical domains. Instructional practice was video-recorded during two time points: after the workshop and at the end of the program. The videos were scored with the Mathematical Quality of Instruction (MQI) instrument. Focusing on three dimensions of instructional practice: richness of mathematics (the conceptual aspects of math, such as the use and quality of mathematical explanations), student participation in mathematics (student mathematical contributions, explanations, and reasoning) and errors and imprecision (incorrect, unclear, and imprecise use of math). Student achievement was measured with a content adapted assessment and the scores of the state mathematics assessment. In this study the PD program had a positive impact on teacher knowledge and on some aspects of instructional practice. On average the scores of the treatment teachers were higher on *Richness of mathematics.* The scores on the other two aspects were also higher for the treatment group, but not statistically significant. The impact on teacher knowledge was larger for teachers with higher baseline knowledge. Both knowledge and instructional practice were not statistically significantly associated with student achievement. Only the *Errors and imprecision* aspect was significantly and negatively related to student achievement.

Van Kuijk, Deunk, Bosker, and Ritzema (2016) investigated whether student reading comprehension could be improved with help of a teacher PD program consisting of three components: 1) setting standards and performance goals for every student, 2) applying formative assessment and data use, and 3) knowledge and instruction for reading comprehension. All components were equally important within the PD program. Second- and third-grade teachers, principals and internal support coordinators participated in the program. Throughout the school year of 2011–2012, the time investment of the teachers was scheduled for 40 hours, including attending 9 after-school meetings and completing homework assignments. Participation was voluntary and free of charge. The effectiveness of the PD program on 2nd- and 3rd-grade student achievement was examined using a quasi-experimental pretest-posttest control group design. This study formed part of a larger conglomerate of teacher PD intervention studies. The authors used the conglomerate to construct a suitable control condition; second and third-grade groups from schools that had no intervention in these grades. In total, 35 groups in the experimental condition (420 students) were matched to 35 control groups (399 students) by applying propensity score matching. Reading comprehension assessments and mathematics assessments were used for the estimation of the propensity score and the results on these instruments were also used to investigate the effect of the program on student achievement. Matching was conducted at the group level, whereas the analyses on the effect of the program were conducted on the student level. A multilevel regression analyses controlling for several covariates (students' sex, students' grade, and students' performance on the pretest and on the mathematics assessment) was used to analyze the data. Students in the experimental condition scored significantly higher on a standardized assessment than the control condition (d = .37, 90% CI). The authors checked for the robustness of these results using different model specifications and found similar though smaller effect sizes (d = .29 and d = .30, respectively). No differential effects of the PD program were found in relation to initial reading performance or grade. At the end of the program, students in the experimental condition were more than half a year ahead of students in the control condition.

Ostermeier, Prenzel, and Duit (2010) evaluated the SINUS-program (a program focussed on improving efficiency of the mathematics-science education) in a large-scale comparison study between SINUS schools and a representative sample of German schools tested in PISA. The aim was to improve the mathematics and science instruction, in particular the classroom instruction. After an inventory of the problem areas, 11 program modules were composed. Schools had to determine their problem area and choose two modules of the program. The modules served as a starting point to improve teaching. In addition written materials, in-service training or consultation was offered to the teachers developing their own classroom instruction. The program stimulates cooperation and collaboration on different levels, especially between the teachers participating in the program. The modules are a frame of reference for support and are based on the current state of research in science and mathematics education and on learning and instruction in general. Since the modules address key problem areas the teacher can locate class-related problems within the frame of modules. Examples are provided to solve the problems. The evaluation of the program includes the engagement of the teachers in the program, the kind of support they want, the products and understandings they develop and the changes in student competencies and motivation. The degree to which the teachers appreciated the program and its goals was investigated with questionnaires in two surveys. This data could only be used on school level because of privacy regulations. The teachers seemed to be engaged in the program and the appreciation of the professional

development program was high. The teachers rated nearly all aspects rather positively and the ratings increase in the second survey. The request for support was quite different; almost half of the teachers wanted more support while the other half wanted less. The groups in the schools developed a large number of materials, teaching units, classroom projects, curricula and tasks that could be shared with the other schools. The researchers concluded that the modules showed concrete ways to improve instruction step by step, and that changes could easily be integrated in teachers' routines. The PISA instruments were used to determine student achievement. In 144 schools of the SINUS program the instruments were assessed twice. At the school and program level the progress in mathematics and science performance, the interest and the perception of instruction experience of the students was evaluated and compared with a national sample of schools not participating in SINUS. The results provide evidence for the positive change in SINUS schools. The teachers in the SINUS schools reported more cooperation activities at the school level and the students perceived more cognitively activating classroom teaching than in the PISA schools. Both student interest and competencies were higher in SINUS schools compared with other schools.

In the My Teaching Partner–Secondary program (MTP-S), a coaching program focused on improving teacher-student interactions in secondary classrooms. The instructional and motivational qualities of teachers' ongoing daily interactions with students were the focus of the study. The aim of this randomized controlled trial experiment was to examine the generalizability and the sustainability of the impact of the intervention on changes in teaching and in student achievement (Allen, Pianta, Gregory, Mikami, & Lun, 2011). Twelve secondary schools with 78 teachers participated for a period of 13 months in the MTP-S and 2 years in the evaluation of the program. Half of the teachers participated in the program and the other half in the regular in-service training. The teachers were randomly assigned to groups. In the intervention year 1267 students in 76 classrooms participated; in the post- intervention year 970 additional students in 61 classrooms participated. The program started with a workshop-based training. During the school year teachers got personalized coaching based on video recordings of their class sessions. The trained teacher consultants review these recordings, selected illustrative segments with positive teacher interactions and areas for improvement. The teacher was requested to observe his or her behavior and student reactions and to respond to the consultant prompts. Then a phone conference followed in which the consultant and the teacher discussed ways to enhance interactions using the CLASS-S system. During the school year about every two weeks this cycle was repeated. The evaluation was concentrated on the second year, when the training was ended, to see the generalizable and sustainable changes in teaching. In the analysis differences in subject matter and in populations of adolescents were accounted for. Hierarchical linear models were used for the analysis. Differences in end-of-year student achievement test scores were examined, after accounting for predictions from achievement test scores from the previous year and teacher and student demographic characteristics. Intervention effects on test scores were non-significant in the first year, but significantly positive in the second year with an average increase in student achievement from the $50^{th}$ to the $59^{th}$ percentile. Interaction qualities were observed at the end of the intervention year, with analyses examining whether they potentially reflected an enduring change in classroom qualities that would mediate effects of the intervention on achievement for a new class of students in the post-intervention year. The Interaction qualities were observed at the end of the intervention year and displayed a significant indirect effect on student achievement in the post-intervention year through changes in teacher-

student interaction qualities. The analysis revealed no interaction effects with subject matter or across classrooms or teachers with different sociodemographic and structural characteristics.

The qualities of the teacher-student interactions were predicted by participation in the intervention, an indirect effect of the intervention on student achievement through these observed qualities was observed. It seems that the intervention causes an enduring change to the teacher and to the classroom as a behavior setting, since the effects on teachers carried into the next year and new students, when there was no coaching and 30% of the teachers were teaching at least slightly different content material than in the first year. The lack of effects on student achievement in the intervention year suggests the difficulty of rapidly changing classrooms in ways that lead to student achievement gains.

A research project on low science achievement of English Language Learning (ELL) students implemented a professional development intervention aimed at improving science and literacy achievement in urban elementary schools (Santau, Maerten-Rivera, & Huggins, 2011). The goal for ELL students was to use English in social settings, to achieve academically in all content areas and to use English in appropriate ways. Therefore teachers of ELL students need to create classroom environments that promote development of general and content-specific academic language. The idea is that Inquiry-based science promotes students' communication of their understanding in a variety of formats. The intervention focused on professional development through the provision of curriculum units and teacher workshops. This study examined the results of science achievement of fourth-grade students in a large urban school district during 3 years of the intervention. Of the 33 disadvantaged schools in the district with many ELL students 14 participated in the intervention, 6 in the treatment group and 8 in the comparison group. A total of 55 teachers, 83 classrooms of grade 3, 4 and 5 participated in the intervention over the three years. During the first year third grade teachers began the intervention, in the second year 4 grade teachers and in the third year 5$^{th}$ grade teachers. In the 4$^{th}$ year the 3th grade teachers phased out of the intervention and so-on. The entire science curriculum for grades 3-5 was developed in curriculum units that promote science inquiry with students who may be less familiar with scientific practices. The units gradually progress along the continuum of teacher-explicit to student-initiated inquiry and to higher levels of complexity in terms of both science concepts and the level of inquiry required. Teacher guides provides the backgrounds, information and explanation of the key-concepts, the learning of students, how to incorporate English language and literacy and supplementary materials. The student booklets focused on standards-based, inquiry-driven science learning and highlight activities or strategies to foster reading and writing as part of science instruction and provide explicit guidance to promote English proficiency. The teachers followed full-day workshops during the school year. In the workshops for science experimental designs, procedures for gathering data, multiple ways of displaying the data, and conclusions based on data and evidence were discussed. In addition teachers discussed how to promote student initiative in conducting inquiry as they gradually reduce their level of guidance. For English language and literacy various literacy strategies were discussed and how they can reinforce these strategies in their instruction. A project-developed science test was administered to the fourth-grade students at the beginning and end of science instruction over the school year. Each student received a gain score in science achievement, which was computed by subtracting the pretest Rasch score from the posttest Rasch score. Analyses were conducted with the 1,758 fourth-grade students who took both the pre- and posttest. Analysis were done with

hierarchical linear modeling (HLM) with student background independent variables at the level of the classroom by converting them to proportions that were reflective of the classroom-level values of the variables. All students made significant achievement gains from pre- to posttest. The gains were not significantly different for the ELL students and did not vary based on the year of the intervention. Comparison of test-scores of students in the treatment group on items of the NAEP and TIMSS-tests with the tests-cores of the norm group revealed the students in the treatment group scored substantially lower on these items than the norm group at the pretest but better than the norm group at posttest. The conclusion was that the intervention didn't narrow achievement gaps for ELL students, although the gaps didn't widen.

As part of a larger project, "E-Learning for Educators" (efe), four randomized experiments were conducted with teachers from multiple states to evaluate the effects of online professional development (OPD) on teacher knowledge, teacher practices and student (Dash, Magidin, O'Dwyer, Masters, & Russell, 2012; de Kramer, Masters, O'Dwyer, Dash, & Russell, 2012; Masters, Kramer, O'Dwyer, Dash, & Russell, 2012; O'Dwyer et al., 2010) achievement four independent trials had the same research design, but focused on a single grade level and subject area: fourth grade English language arts (ELA), fifth grade mathematics, seventh grade ELA, and eighth grade mathematics. The four online workshops for upper elementary and middle school teachers were developed through a collaborative process among various stakeholders and include both theoretical information as well as pedagogical techniques. The workshops were based on a learning community model with a strong emphasis on peer-to-peer discussions. A workshop started with an orientation session and continued with six content sessions during the schoolyear. The sessions involved readings (articles, books etc.), activities (online video's, work with existing classroom materials, etc.) and discussions (responses on questions related to the readings and activities). Finally the teachers had to develop an action plan or lesson plan based on the workshop content. Each experiment was conducted across multiple states and included three rounds of data collection, each spanning three school semesters. Teachers who volunteered to participate were randomly assigned to the treatment or control group; students were grouped based on the assignment of their teacher. Teachers in the treatment group participated in each of the three years in an OPD workshop of 4-6 hours per week for seven weeks. Teachers in the control group could participate in their normal professional development activities. Numbers of teachers varied from 71 to 110 per trial, comprised of 34 to 46 % of the recruited group who completed all requirements. Less than half of the participants were in the treatment condition. Most teachers taught in high-needs schools. There were almost no significant differences in background variables between the treatment and control groups in the four experiments. Only in the fourth grade there was a significant difference between the teacher groups in gender (all 5 man were randomly assigned to the control group) and in the eighth grade students differed significantly in the racial and ethnic characteristics. Tests for the ELA trials measured teachers' knowledge and frequency of instructional practices on three broad areas: vocabulary, reading comprehension, and writing. Knowledge items measured content knowledge and were both closed and open response. Practice items measured self-reported frequency of desirable instructional practices as defined by each OPD workshop. They were closed response items on a Likert scale. The tests for the students measured students 'knowledge and practices in vocabulary, reading comprehension and writing'. Tests for the Mathematics workshops had items that fell in the following broad content areas: fractions, algebraic thinking, and measurement. Reliability of most tests of English Language Arts was good, except the vocabulary test of the students in the fourth grade and the writing knowledge scale

for the students in the seventh grade. For mathematics the overall mathematics knowledge scale and the practice subscales for teachers had adequate reliability, but the pre-test scores on the fractions and measurement knowledge scales were quite low. Same was the case with the teacher practice scales. The treatment effect for the teacher outcomes are estimated with analysis of covariance (ANCOVA). With hierarchical linear modelling is examined whether teachers' group membership was a significant predictor of students' post- test scores after controlling for pre-test scores. Outcomes on each post-test knowledge and practice scale were modelled as a function of the students' pre-test scores (level-1) and as a function of teachers' group membership (level-2). In the fourth grade ELA trial (Masters et al., 2012) the multilevel analysis gives statistically significant positive coefficients associated with the teachers' membership group for the vocabulary and overall ELA knowledge scores (b=0.036, t=2.19, df=106, p=0.014 and b=0.030, t=2.40, df=106, p=0.013, respectively) and for students' writing practice scores (b=0.049, t=2.73, df=106, p=0.042). In the seventh grade ELA study the mean differences showed that the average change in the treatment group was larger than the average change in the control group (de Kramer et al., 2012). For the teachers the standardized gains for the treatment group were larger than the standardized gains for the control group, although the gains in reading comprehension practices were of comparable magnitude. Between 19% and 21% of the total variability on the post-test scores could be attributed to between-teacher differences. The results for the student knowledge scores show that the standardized gains for the treatment group were larger than the standardized gains for the control group based on the pre- to post/test effect size. The results on the practice scores follow the same line, but the gains were small for each subject. The effect sizes calculated as the difference between the adjusted means for the post-test divided by the pooled standard deviation the OPD had no measurable impact on vocabulary knowledge (0,05), reading comprehension knowledge (0.07), and the overall ELA knowledge combined score (0.08) and on practice scores for reading comprehension (0.11) or writing (0.17). The multilevel regression models revealed that there was a significant amount of variation in students' post-test knowledge and practice scores between teachers, 19% and 21% of the total variability on the post-test scores could be attributed to between-teacher differences. The outcomes of hierarchical linear regression models indicate that after controlling for students' pre-test practice scores, students in the treatment group were predicted to have statistically significant higher post-test reading comprehension scores than students in the control group. The coefficients for students' writing post-test scores were not statistically significant different from zero after controlling for the pre-test measures. The fifth grade mathematics group was analysed with a repeated measures analysis of variance (RM-ANOVA) (Dash et al., 2012). The results of this analysis confirmed a significant interaction between group membership (experimental or control) and total pedagogical content knowledge, $F(1, 77) = 22.36$, $p < .001$. So there was variability by group membership regarding pre to post gains in teachers' pedagogical content knowledge. Group membership (experimental or control) emerged as a significant main effect, $F(1, 77) = 8.73$, $p = .004$, confirming that teachers in the experimental group received significantly higher overall pedagogical content knowledge scores ($M = .52$, $SE = .02$) than did teachers in the control group ($M = .44$, $SE = .02$). The same analysis was done for teachers pedagogical practices. The results of this analyses revealed a significant interaction between group membership (control or experimental) and pedagogical practices, $F(1, 77) = 46.10$, $p < .001$. The significant interaction is indicative of variability by group membership. The effect size associated with this interaction as measured by $\eta p^2$ was quite large (.38). Group membership (control or experimental) did not emerge as a significant main effect, $F(1, 77) = .230$, $p = .633$, but the results indicated a significant main effect for overall pedagogical

practices (pre/post) F(1, 77) = 84.53, p <.001. The effect size as measured by $np^2$ was large (.52). Controlling for students' pre-test scores the HLM-analysis reported that teachers' participation in the online professional development treatment did not predict students' achievement in mathematics (b = .027, t = 1.89, df = 77, p =,.062). There were no significant differences in students' mathematics achievement, as a function of teachers' group membership (control or experimental). The other experiment on mathematics took place in the eighth grade. The OPD in that trial had a medium effect on the overall mathematics knowledge scale, which was a significant effect (b=0.042, df=1,68, p=0.023). The effect on the separate knowledge scores was small and not significant. Independent samples t-tests showed no statistically significant difference in average pre-test scores on any of the practice content scales for the treatment and control groups (t=0.40, df=69, p=0.968 for the proportional reasoning scale; t=1.446, df=69, p=0.153 for the geometric measurement scale; and t=0.149, df=69, p=0.882 for the functions scale). For all three practice content sub-scales, the pre-to-post gain in mean score for the treatment group was higher than that of the control group (by approximately half a standard deviation). The ANCOVA results showed significant treatment effects for all three scales, after controlling for differences in pre-test scores. Effect sizes were large for the proportional reasoning practices scale ($\eta p^2$= 0.156) and the functions practices scale ( $np^2$= 0.207) and moderate for the geometric measurement practices scale ($np^2$ = 0.082). For student content knowledge the mean differences show that the average change in the treatment group was larger than the average change in the control groups. The effect sizes for the post-test means adjusted for the pre-test means showed no measurable effect of the OPD on the student scores of any of the scales. Multilevel modelling showed statistically significant positive effects associated with the teachers' membership in the treatment or control group for students' overall mathematics knowledge scores, geometric measurement knowledge scores and functions knowledge scores (b=0.030, p=0.017; b=0.052, p=0.024, and b=0.041, p=0.034, respectively). The coefficients indicate that after controlling for students' pre-test scores, students in the treatment group were predicted to have statistically significant higher overall mathematics knowledge scores, and geometric measurement and functions knowledge scores than students in the control group. In conclusion: the four trials provide strong evidence that participation in a coordinated series of three OPD workshops showed positive effects on teachers' self-reported instructional practices and content knowledge. And although much smaller and less consistent it also significant effects on the students were revealed (at least one student measure in each trial).

In Texas the SimCalc approach was evaluated in several studies. Roschelle et al. (2010) present two randomized controlled experiments and one embedded quasi-experiment of the SimCalc, an approach which integrates an interactive representational technology, paper curriculum, and teacher professional development originally developed in design research. The program concerns the mathematics of change and variation and is meant to highlight the strand of mathematics relating to algebra and leading to calculus. The intention is a more coherent and fruitful mathematical experience for both disadvantaged and advantaged learners in middle school by creating new opportunities for students to learn complex and conceptually difficult mathematics. The software presents animations of motion. Students can control the motions of animated characters by building and editing mathematical functions in either graphical or algebraic form. Often they are asked to tell stories that correspond to the functions (and animations). Two randomized experiments were implemented; one in the seventh-grade that lasted two years and a one-year lasting eighth-grade experiment. Schools were randomly assigned to a treatment or control group at the beginning of

each study. The control group of the seventh grade study began to use the SimCalc replacement unit in the second year (delayed treatment). It provided the possibility to make a within-teacher quasi-experimental comparison between classrooms of the delayed-treatment teachers in Year 1 with the classrooms of the same teachers in Year 2, when those teachers received the SimCalc replacement. The research evaluated the impact of the intervention as a whole. The intervention integrated pedagogy, curriculum, professional development, assessment and school leadership. The target mathematics in seventh and eighth grade curriculum is proportionality and linear function, which is also important for student's science learning. Two replacement units were designed recovering this content. The seventh year contextual theme is that students must serve as soccer team managers - training players, ordering uniforms, planning trips to games, and negotiating their salaries. The contextual theme in the eight year is that students are designers of electronic games who must use mathematics to make the games. For each of the studies, treatment teachers got professional development to strengthen their mathematical content knowledge, learn to use the curriculum materials, and/or plan specifically how to use the materials. After a two day workshop (TEXTEAMS) on the mathematical knowledge for teaching rate and proportionality, the PD got on with a 3-day summer workshop introducing the SimCalc units. In a 1-day workshop specific plans were made for how and when to use the SimCalc materials in their classrooms (writing lesson plans and thinking through their own logic for the unit). In the control groups, the business-as-usual curriculum addressed, within the same time frame as the SimCalc unit, similar basic concepts but provided less coverage of more complex concepts. Two different teacher professional development delivery models were used: a consistent model where two highly experienced mathematics teachers led all the workshops and a train-the-trainers model. Instead of standardized tests that did not capture the conceptual depth students could reach using the SimCalc technology and curricula the research team developed two tests. The tests were thoroughly validated. Within each study, the identical assessment was administered at pre-test and post-test. Next to these tests data on student demographics, classroom implementation, teacher background and school-level data were collected. Each day of the unit, teachers filled out a page in the log. Data were analysed with hierarchical multilevel modelling with 2 –levels (students nested within schools), while 70 % of the schools had only one teacher. The first level predicted student gain scores as a function of a school-specific intercept and P student-level covariates. At Level 2, the school-specific intercept was modelled as the sum of a grand mean, a fixed effect for treatment assignment $T_j$ Q school-level co- variates, and a random deviation. In the 7th Grade 95 classrooms completed the Year 1 experiment and 67 teachers completed Year 2. For the Seventh grade Quasi-Experiment there were 30 delayed treatment teachers who finished both year 1 and 2. In the 8th grade Experiment 56 teachers completed the study. Teachers administered the student assessments immediately before teaching the unit and immediately after teaching it. Each day of the unit, they filled out a page in the log. The treatment and control groups began with similar pre-test scores in all three studies. However, gains from pre-test to post-test were significantly higher for treatment students. The effect sizes were large, the gains differences between the two groups occurred mostly on the items for the more complex concepts (.63, .50, .56 total; .10, .13, .19 for basic concepts and .89, .69 and .81 for complex concepts). Students in the treatment condition spent much more time in the computer lab (41,5% of the days and 3,5% in seventh grade and 72,8% and 6,6% in the Eight Grade) and in the Eight Grade the amount of time spent in the lab was a predictor of student learning gains. Across all three studies there was not a significant correlation between the number of days spent teaching the unit and gains on the complex concepts, so probably time on task does not explain the main effects.

The Chicago Public Schools started in 2007 a schoolwide reform program, their version of the Teacher Advancement Program (TAP). The findings from the four year implementation period are described in various reports and summarized in a last report (Glazerman, Seifullah, & Mathematica, 2012). The research investigated the implementation of the program and what impact it had on student achievement and teacher retention. In the program teachers get performance incentives, along with tools to track their performance and improve instruction. They can earn extra pay and responsibilities through promotion to mentor or master teacher and can earn annual performance bonuses based on a combination of their value added to student achievement and observations of their classroom teaching. The program includes weekly meeting of teachers and mentor, regular classroom observations by a school leadership team and pay for principals who meet implementation benchmarks. The research had a hybrid study design with both the random assignment of schools to year of implementation ("experimental design") and the careful matching of Chicago TAP schools to non-TAP schools in the district ("quasi-experimental design"). The matching was based on demographic and academic characteristics, school size, teacher retention, school accountability status, student achievement, student race/ethnicity, student poverty, student special education status, student language proficiency and charter school status. In the first year 16 elementary schools were randomly assigned to the TAP-group (8 schools) or the comparison group (8 schools) that started a year later with TAP and were control group first year. The second year 18 additional schools were randomly assigned in the same way to TAP-group and comparison group. Students in grade 4 and 7 participated in the study; well over 7000 for reading and mathematics and more than 1700 for science. Data used for the study include teacher surveys and principal interviews as well as standardized student test scores, teacher administrative records, teacher observation scores, payouts by teacher, and scores on a program review. The teacher survey focused on the following areas: teachers' educational background and professional experience, their certification status, and their current teaching assignment; the types of professional development and support that teachers receive at their schools; the leadership roles and responsibilities teachers have assumed in addition to their regular classroom teaching duties; the compensation associated with teachers' performance and that of their students; teachers' attitudes about and satisfaction with various aspects of their school and the opportunities provided to them; and teachers' basic demographic characteristics. The study obtained administrative data from the Chicago Public Schools (CPS) on students, all teachers and Chicago TAP participants. Test data were from the State Assessment in mathematics and reading for grade 3 through eight and in science for grade 4 and 7. Student and teacher background information data was available for most experimental years. Also from the CPS were teacher scores on the Skills, Knowledge, and Responsibilities (SKR) classroom observation rubric performance; payouts by teacher; and scores on a program review that tells each Chicago TAP school how well they have been implementing the program over the current school year. SKR scores are assigned based on observed classroom performance in four domains: designing and planning instruction, learning environment, instruction, and responsibilities. The administrative data sample showed few statistically significant differences between Chicago TAP and matched comparison schools in teacher characteristics, higher percentage of teachers holding National Board certification and lower percentages of teachers in academic but non tested grades or subjects and less late-career teachers in Chicago TAP schools than in comparison schools. The TAP-program was not fully implemented in the Chicago TAP schools, elements of TAP had been introduced, but TAP implementation had not been "rigorous.", especially in the field of performance-based compensation. The average program review score was around 3 out of 5 for each cohort of schools in each of the first three years of

rollout. In terms of mentoring, weekly meetings, career opportunities, and compensation, teacher reports were consistent with the goals of the program, and Chicago TAP teachers' experiences were sufficiently different from those of teachers in non- TAP schools to conclude that real change occurred. Derived from a regression model controlled for students' prior achievement and other factors there were no impacts on student achievement in the first year of implementation. Test scores for students in Chicago TAP schools were statistically indistinguishable from those of students in control schools (<3 % of a standard deviation for math and reading, 14% of a standard deviation for science, but due to the small sample for science this is in the margin of error.18). Separate analyses by grade level revealed similar results. Experimental comparisons of schools with different amounts of experience implementing Chicago TAP suggested there was no advantage associated with having spent an extra year of implementing. Conclusion after lot of various analysis and matching methods was that the program did not consistently raise student achievement as measured by growth in Illinois Standards Achievement Test (ISAT) scores. There were both positive and negative test score impacts in some selected subjects, years, and cohorts of schools, but overall there was no detectable impact on math, reading, or science achievement that was robust to different methods of estimation. Teacher retention was another outcome measure. In the analysis increase of school retention of teachers was found, although the impacts were not uniform or universal across years, cohorts, and subgroups of teachers. For some of the schools teacher retention outcomes improved, inducing teachers to stay longer in their school, but not much evidence was found that Chicago TAP helped prevent teachers from leaving the district. In conclusion it can be stated the Chicago TAP program was only partially successfully in achieving its' goals. Implementation of Chicago TAP increased the amount of mentoring, promotion opportunity, and compensation relative to non-TAP schools, but these changes did not result in higher student achievement.

In a randomized controlled trial (Meyers et al., 2016) evaluated the eMINTS (enhancing Missouri's Instructional Networked Teaching Strategies) PD program. The overall goal of eMINTS is to help seventh- and eighth-grade mathematics teachers, especially those in high-need regions, develop student-centered, purposeful instruction fostered by technology utilization in order to increase their students' engagement and achievement in communication arts, mathematics, and 21st-century skills. The program provides 240 hours differentiated PD sustained over 2 years grounded in constructivist pedagogy, with opportunities for "hands-on" work that connects to teachers daily instructional practice. Collective participation in this program promotes teacher communication and collaboration to support instructional changes. eMINTS embeds these features by focusing on the promotion of four specific strategies that address issues the authors have identified as barriers to the consistent use of standards-based instruction; inquiry-based learning, high-quality lesson design, community of learners and technology integration. In addition to this intensive 240 hour PD program eMINTS also includes: 1) a specific set of school and classroom technology equipment, 2) intensive on-site training for school principals, district and school technology coordinators, and classroom teachers, 3) job-embedded coaching to enhance teachers' classroom practices and written curricula, and 4) just-in-time learning opportunities via online courses to help teachers improve their practice over time. One third of the study schools were randomly assigned to receive an additional third year of PD that used Intel® Teach courses and online tools to further enhance teachers' technology integration skills by providing teachers with additional PD and Web-based tools to build on what they learned during the first 2 years of the program. The study includes approximately 200 teachers and 3,000 students in 60

high-poverty rural middle schools across Missouri who were randomly assigned to one of three groups: treatment 1 (eMINTS), treatment 2 (eMINTS + Intel), or control. Analyses of school-, teacher- and student characteristics showed that treatment and control groups were relatively similar (no significant differences between groups). Data were collected using a teacher survey, classroom observations, a student engagement survey, a 21st-Century skills assessment and Missouri Assessment Program data of student outcomes. Analyses were conducted to examine the difference in student and teacher outcomes between year 3 results and pretest measures collected at baseline. Rasch analyses of teacher selfreport data showed no statistically significant differences between treatment and control teachers for a community of learners after year 3. Both treatment groups reported significantly higher levels of inquiry-based learning (effect sizes 0.73 and 0.96, respectively) and technology integration (effect sizes 1.43 and 1.56, respectively) relative to the control group. The eMINTS + Intel group reported significantly higher levels of highquality lesson design (effect sizes 0.50). No significant differences were found between the eMINTS program and the eMINTS + Intel program. Rasch analyses of all observed domains revealed significant positive results for both treatment groups versus control group with medium effect sizes: community of learners (0.50 and 0.52, respectively), inquiry-based learning (0.68 and 0.56, respectively), and technology integration (0.58 and 0.78, respectively). Again, no significant differences were found between the two treatment groups. The observation instrument did not include a high-quality lesson design domain. Student analyses using two-level hierarchical linear models with students nested within schools showed positive and significant impact estimates of mathematics achievement for both treatment groups (effect sizes of 0.128 and 0.178, respectively). No significant impacts on 21st-century skills or communication arts were found when comparing treatment groups with the control group. Impacts on student engagement were not statistically significant and there were no significant differences in mean student outcomes between both treatment groups. The authors note that the findings of this study do not generalize beyond rural schools enrolling seventh- and eighth-grade students, and perhaps then only rural Missouri schools.

*b) Interventions focusing on the improvement of teacher knowledge or teacher-efficacy and student achievement*

Watson and Beswick (2011) investigated the effectiveness of a teacher knowledge enhancement program designed to enhance student's numeracy skills in the middle years of schooling. Teachers were supported in developing their own skill efficiency, their conceptual understanding and they were supported in their considerations of how they could provide similar opportunities for their pupils. The 3 two-day sessions involved presentations on student difficulties, trials of the activities in their class, and time for sharing and reflecting on the implemented changes. The "teacher profiling instrument" measured seven types of teacher knowledge: teacher confidence, beliefs about numeracy in the classroom, knowledge of students as learners (KSL) and pedagogical content knowledge (PCK). The program ran in three consecutive years with 35 teachers of 12 schools. Pupils were surveyed at the beginning and the end of the third year of the program. The program was focused on mental computation, proportional reasoning and assessment. At the start of the program teachers were asked about their needs for professional development and at the end they were asked to what extent their needs had been met. Pupil's attitudes to mathematics and their answers on the frequency of mathematics activities taking place in their classrooms were surveyed and they were asked to explain their understanding in relation to basic numeracy concepts. A total of 674 pupils participated in the study. Results reveal that the average confidence of teachers in teaching various

topics increased, but teachers' beliefs related to the mathematics classroom hardly changed. Teacher behavioral change was apparent in the raise of the number of various aids provided for problems. Of the 25 questions pupils answered about their own beliefs, significant change was seen on six items. These items reflected some of the items of the CPD; largest change was on being asked to explain their math's thinking. The performance on student's mathematical thinking improved significantly and for some classrooms a highly significant change was found. The authors conclude that it is warranted that dedicated time is provided for teachers to be challenged and assisted to extend their own mathematical thinking, to appreciate the typical thinking of pupils, to allow for planning to revise their teaching practices, and to reflect on change. Teachers and their pupils change together.

In a large-scale study with elementary school teachers, an intervention was implemented to enhance science teaching efficacy measured aiming to enhance student learning (Lumpe, Czerniak, Haney, & Beltyukova, 2012). Each year of the project teachers participated in six, 2-week long summer programs that were focused on inquiry-based instruction, science content knowledge, and science process culled from the districts' adopted curriculum. Elementary school teachers were given full-time release, to provide assistance to classroom teachers during bi-weekly visits. All the school principals were involved in this science reform effort. The changes in beliefs of 450 teachers concerning teaching science, were measured with reliable belief scales. These scales measured self-efficacy, outcome expectancy and a teachers' belief about the supportiveness of their professional context. Demographics, in particular gender, grade level, years of experience, educational degree, number of days a week teaching science, and length of a science lesson were included in the analyses. The State Science Achievement Test scores for over 8,000 students across 43 elementary school buildings were collected at the end of the schoolyear in which their teachers were participating in the professional development program. The analysis included paired t-tests and multiple regression. The self-efficacy beliefs of the teachers were more positive after participating in the program, but there was a slight drop in context beliefs. Outcome expectancy didn't change significantly. Time period teaching science predicted teacher beliefs. Males displayed more positive beliefs than female teachers. For student outcomes teacher beliefs and the number of hours participating in the research-based professional development program were significantly predictors, although a small portion of variance was explained.

*c) Studies of the effects on student achievement by comparing different kinds of interventions in one study*

(Antoniou, Kyriakides, & Creemers, 2011; Antoniou & Kyriakides, 2013) conducted a group-randomization study on teacher development programs in 130 primary schools in Cyprus, following their dynamic integrated approach (DIA). The teachers had to adopt their teaching skills to the content of mathematics that was discussed in the training sessions. The aspect that was compared between the two conditions was the different focus in the critical reflection and the action plans. In the DIA-group the specific focus was on skills that belonged to the development stage of the teacher and in the other group on the whole spectrum of knowledge, skills, attitudes and beliefs about teaching. So although "both experimental treatments encouraged and utilized critical reflection of teachers on their teaching practices, teachers employing the DIA were asked to reflect on those aspects of their teaching practice that were found to be related with their priorities for improvement based on the stage at which they were situated". The study was done in four phases. At first in the initial evaluation the teaching skills of all 130 teachers were evaluated by external observers, data of

student achievement were collected from their 2356 students and teacher and student questionnaires measured potentially explanatory variables. Teachers were classified into developmental stages according to their teaching skills and randomly evenly allocated into the two experimental groups. Then the teachers of each group attended nine training sessions. The first session introduced the main phases and the importance of the professional development program and trained the development of an action plan. Next sessions were separately for the two conditions organized every month and based on “active teaching”. The DIA-group was divided in four groups according to their developmental stage. Sessions concentrated on the teaching skills which correspond to their developmental stage. The teachers developed their own action plans. In the sessions teachers reported teaching practices and commented on them, shared ideas and teaching materials and exchanged and discussed their experiences. These were focused on effective and non-effective teaching practices, identifying the significance of the factors and comprehend how these factors could be linked with effective teaching and learning. The teachers in the HA-group did the same things as the DIA group without the focusing on practices on their own level. The researchers conducted interviews and collected the teachers’ reflective diaries for analysis. These revealed that teachers of both experimental groups faced similar issues in their efforts to implement their action plans, like lack of time, leadership support, etc. At the end of the sessions the teaching skills and student achievement were measured and a final evaluation meeting was conducted. Background variables of students were gender and SES and of teachers gender, position/function and teaching experience. Pupils’ reported opportunity to learn was by estimating the time spent doing homework and on private tuition. From the student-level data contextual classroom variables like average SES-score were aggregated. Independent observers made four observations with three instruments that measured the teaching factors of the dynamic model and were used in a series of other studies. These observations provided the score for Quality of teaching. At last the extent to which each teacher had put effort into implementing their improvement strategies and their action plans was measured based on diaries of and interviews with the teachers. The impact of the teacher professional development program on the teaching skills was established with the observation measures of the quality of teaching. Comparing the classification of teachers into different stages at the beginning and at the end of the intervention, showed that from the DIA-group one-third of the teachers moved from one stage to another, while the other group had no teachers that moved to a higher stage. The Quality of teacher scores were significantly higher at the end on the DIA-group, the teachers in the other group showed no significant progress. . Nevertheless, one year after the end of the intervention stimulus the two groups neither managed to improve nor decline in teaching skills. The impact on student outcomes was measured with multilevel analysis. An empty model consisting of student, teacher, and school level represented the best solution. Then 73% of the total variance was at the student level, 18,5 % at the classroom level and 10,2 % at the school level. When entering explanatory variables at different steps the multilevel analysis provided evidence that the DIA yields better results in student achievement. The findings of the experimental study seem to provide support to the argument that improvement of teaching skills takes place gradually.

Heller, Daehler, Wong, Shinohara, and Miratrix (2012) studied the effects of various intervention treatments by systematically comparing closely related versions of professional development interventions. Using a randomized experimental design implemented on a large scale in six states, this project compared the differential effects of three teacher interventions as well as a ‘‘business as

usual'' control condition. The interventions were: 1) a Teaching Cases course (with discussions of pre-structured written cases of classroom practice), 2) a Looking at Student Work course (involving analysis of teachers' own student work in conjunction with concurrent teaching) and 3) a Metacognitive Analysis course (in which teachers engage in metacognitive reflection on their own learning experience). The courses contained the same subject matter, but differed in the ways they supported development of teacher pedagogical content knowledge. All three interventions used key features of effective professional development, took 3 x 8 hours, had an identical science content component, but differed in pedagogical content components. The science content contained science investigations to extend the teachers' conceptual understanding of key scientific ideas. Participants were 446 teachers at the $4^{th}$ grade, participation was voluntary. Teachers were randomly assigned to one of the experimental conditions (324) or the control group (122). The analytic teacher sample with complete data was reduced to 271 by drop-out of teachers. Students were the students in participating teacher classes. Of 253 teachers student data were present. The randomized teacher level trial design had a pretest-posttest follow-up experimental design with four conditions in two rounds of interventions and data collection during the study year and only data collection in the following year. The subject specific knowledge on how students learn and how to teach the specific science content was topic in each intervention, but the way they approached it differed. Science knowledge was measured with traditional response tests and written justifications for answers. The written justifications were scored using item-specific rubrics. Teachers 'professional experience and background in science and teaching, school setting and curriculum information was measured with an online teacher survey for the course and after the finish of the teaching the specific content. For each student the teacher gave background information (sex, race/ethnicity and English language competence. Student tests were administered 2 weeks before and after completion of the unit. Implementation and attendance rates were registered. Al the courses produced significant increases in teacher and student outcomes. Student test scores were well beyond those of controls, and the effects were even stronger a year later. The impact on written justifications of answers was a bit different between the three courses. Written justifications of selected-response answers revealed conceptual understandings and ability to communicate about science that did differentiate among the effects of the courses. Science content situated in activities and scenarios involving student curricula and instruction, in combination with analysis of student work and classroom pedagogical practice caused the greatest achievements. In all three experimental conditions students demonstrated significantly higher estimated gains than control group students on the science content test, but there were no differences between the experimental groups, including the follow-up schoolyear. Students' written justifications only showed effects caused by "Looking at Student Work" in the first year. "Looking at Student Work" and "Teaching Cases" both demonstrated gain scores compared to the control group. For "Metacognitive Analysis" there were no differences compared to the control group.

Penuel, Gallagher, and Moorthy (2011) investigated whether and how professional development can help teachers design sequences of instruction that lead to improved science learning. The basic point is that teachers' selection and enactment of materials can and do vary in ways that could limit their efficiency. In three different programs and a control condition teachers learned in a way to organize and work with expert-designed Earth systems science curricula. The study took place in an urban district that recently adopted an *Understanding by Design* (UbD) approach for the teacher's professional development. All teachers in the district had to organize their Earth science instruction

into a 9-week unit. The UbD approach is a framework for designing curricular units of instruction that centers on the big ideas, essential questions, and authentic performances. The models provide detailed guidance on pedagogical strategies linked to specific subject matter domains. The programs had essential features for high-quality professional development and focused on supporting for teachers learning to design sequences of instructions. A total of 53 teachers from 19 middle schools participated in the study. Teachers were randomly assigned to the programs. One group got explicit instruction in the principles of (UbD) to teach Earth system science. They didn't get a specific curriculum, but designed instructional experiences and reorganized existing curricular materials (Earth Science by Design, ESBD). Another group worked with a 10-module middle school curriculum (Investigating Earth Science, IES) with the specific purpose of preparing teachers to implement instructional sequences using the materials as laid out. They did not receive explicit instruction in the models of teaching that underlay IE. The third experimental group was a blend of the other two conditions (Hybrid). Teachers received explicit instruction in the UbD models of teaching and were also immersed in the practice of design as in the ESBD condition. The control group did not receive any guidance but were free to pursue available professional development opportunities. There were no significant differences among groups in characteristics of teachers and students. The extent to which teachers engaged in the conduct referred was observed once in the classroom during the implementation of the unit. Teachers completed a survey administered online on the support, attitudes toward study participation, and the motivation behind particular pedagogical practices. The specialists developed unit tests that were administered twice (pretest and posttest) to the students. Each test contained a brief survey of student background characteristics. Data were analyzed with hierarchical linear modeling with students nested within teachers and teachers nested within schools. The analysis on student achievement revealed differences between the groups. Students from teachers who participated in the ESBD or the Hybrid program outperformed students of teachers in the Control condition. The results of the IES group didn't significantly differ from that of the control group. This means that in explicit instruction in models of teaching can improve student learning. Guidance in selecting curricula that promotes student investigations is not clearly effective for student outcomes. Teachers in the ESDB condition had the most explicit instructions as laid out in the *Understanding by Design* (UbD) approach.

Powell, Diamond, Burchinal, and Koehler (2010) investigated the effects of a one-semester literacy-focused PD intervention called *Classroom Links to Early Literacy*, involving expert coaching with Head Start teachers (n=88) on classrooms and prekindergarten children (n=759). The study was designed to provide two sets of randomized controlled trial comparisons: 1) examining intervention effects in relation to classrooms randomly assigned to a wait-listed control group, and 2) examining effects of remote versus on-site delivery of the PD intervention. Data were collected using classroom observations (Early Language and Literacy Classroom Observation instrument) and child assessments (Peabody Picture Vocabulary Test) at the beginning and at the end of the intervention. The goal of the intervention was to improve teachers' use of evidence-based literacy instruction that in turn would lead to significant improvements in children's literacy achievement. Primary emphasis was given to classroom strategies to improve children's oral language skills and code-focused skills. Attention to oral language included instructional practices aimed at improving children's vocabulary knowledge, listening comprehension skills, and syntactic knowledge. Teachers were given printed copies and demonstrations of phonological awareness activities and were encouraged to emphasize letter sounds when teaching letter names and to use writing to promote letter knowledge. The

intervention gave secondary emphasis to instructional practices to improve children's knowledge of print concepts. Teachers participated voluntarily and received a modest level of monetary compensation for their time. The intervention comprised a 2-day workshop providing an overview of the intervention content, with emphasis on demonstration and guided discussion of evidence-based practices, followed by expert coaching. Coaching sessions (seven in total) were to focus on each of the two primary child outcome areas (oral language and code-focused skills), with the specific sequence and content individualized to classroom circumstances. In the on-site coaching condition, the coach observed classroom activities of the teacher followed by feedback. In the remote coaching condition, teachers submitted a videotape of a targeted instructional practice for the coach to review. During the semester of participation in the intervention, teachers in the remote condition were given use of a case-based hypermedia resource developed for the intervention. In the on-site condition coaches brought a laptop computer with the hypermedia resource for the purpose of showing a video exemplar. Teachers in the on-site condition also could receive copies of published articles included in the hypermedia resource. Hierarchical linear model analyses, with the year of participation as a covariate in the classroom–teacher outcome analyses and child race/ethnicity, child gender, and cohort as covariates in child outcome analyses, revealed positive PD intervention effects on general classroom environment (*d* =0.99), classroom supports for early literacy and language development (*d*=0.92), and on children's letter knowledge (*d* =0.29), blending skills (*d*=0.18), writing (*d*=0.17), and concepts about print (*d*=0.22). Teachers in the intervention condition demonstrated a trend toward larger gains in code-focused instruction (*d*=0.62). No significant intervention effects on teaching practices and children's outcomes related to oral language were found. Children in intervention classrooms had no significant gains for receptive language, letter-word identification, or initial sound matching. There were no differential effects of remote versus on-site delivery of literacy coaching; no consistent pattern emerged. Teachers who received on-site coaching demonstrated significantly larger gains in code-focused instruction (*d* =0.71) and had higher scores on code-focused instruction at post-intervention (*d* =0.67) than teachers in the remote coaching condition. However, children whose teachers received remote coaching showed larger gains on the Peabody Picture Vocabulary Test (*d* =0.13) and were more likely to exhibit initial sound-matching skills (odds ratio=1.85).

The study of Fishman et al. (2013) examined differences in outcomes of professional development by comparing online and face-to-face conditions. The experiment compares teachers' learning how to implement new secondary science curriculum in both conditions. They conducted a cluster randomized experiment where 49 secondary teachers from across the United States (in schools that adopted the curriculum) were randomly assigned to either a face-to-face or online condition. 80% of the teachers were certified in environmental science and had about five years' experience in teaching this subject. The professional development content is held constant in both conditions while the research is focused on the differences in teachers' learning in terms of changes in beliefs and knowledge, teachers' classroom practices and student outcomes. The PD in this study is designed to prepare high school teachers to implement a year-long environmental science curriculum and employs a pedagogical approach called Learning-for-Use, designed to develop understanding through cycles of motivation, knowledge construction, and knowledge organization. The PD is intended to increase the probability that teachers' curriculum enactment is consistent with designers' intentions, and leads to desired learning outcomes. The curriculum consists of three units meant to be completed over an academic year. One condition was a week-long (48 hr) face-to-face

workshop spread over 6 days. The other was an online workshop designed to be completed by teachers asynchronously at their own pace. The researchers examined differences with respect to changes in teacher CK, teacher beliefs about self-efficacy to teach environmental science, and teacher beliefs about teaching science in general. This was measured with online surveys including questions about teachers' background and preparation for teaching science. Classroom practice was measured with videotaped observations of teaching practice. Looking for differences in how the teachers taught the curriculum, the videos were scored on quality of curriculum enactment/general teaching and quality of PD enactment, thereby focusing on three central strands of the curriculum (making connections, decision making and technology use). Differences between the two conditions for student knowledge outcomes were determined with test scores related to the curriculum. To analyse the data a two-level model (students nested within teachers) was used. Using a matched pre–posttest design, student learning was operationalized as the gain from pre to post. With respect to CK, teachers in both conditions improved on average. The slight difference between the teachers in both conditions was not significant. Teachers in both conditions improved with respect to personal beliefs and did not improve with respect to impersonal beliefs. With respect to teacher beliefs there were no treatment effects. Teachers in the online condition were more likely to read through and discuss essential questions and lesson overviews when introducing a lesson than the face-to-face teachers. Online teachers were less likely to substitute components of a lesson than the face-to-face teachers. In addition, online teachers were more likely to refer to the text when teaching than the face-to-face teachers. Students in both conditions improved their science scores. The difference in means between the two conditions, although positive, favouring the online condition, was not significant and in the two-level analysis the overall treatment effect was not different from zero and thus the benefit was similar.

Taylor, Roth, Wilson, Stuhlsatz, and Tipton (2016) studied the impact of the analysis-of-practice, videocase-based STeLLA PD program on upper elementary students' science learning. The STeLLA program supports teachers in classroom-based learning over time. The program seeks to promote changes in science teaching and learning through an approach where teachers' science content learning and their learning about science pedagogy emerge from collaborative, videocase-based inquiries into practice. A conceptual framework that guides teachers' analysis of practice in the PD program and their classroom science teaching practice is developed from research on effective science teaching. The study compares the effect on students' science achievement of a teacher professional development program that integrates science content deepening with analysis-of-practice (i.e., the treatment) relative to a comparison PD program of equal duration and intensity that includes only content deepening (teacher content knowledge and confidence in teaching science effectively). Central in the program is the mutual interaction of teachers in small grade-level study groups, discussing each other's lessons and video-cases. The program starts with a two-week summer institute. During the fall teachers teach the study-lessons and analyze the videos in small groups and in the winter this shifts from implementing provided lesson plans to developing their own lesson plans. The teachers in the Content Deepening Program did not participate in analysis-of-practice activities, nor were they introduced to the STeLLA conceptual framework and teaching strategies. The form consisted of a mixture of hands-on science investigations, creation and analysis of content representations, science notebook writing, large and small group discussion, short lectures and readings, and field trips. In a period of two years schools were recruited across two cohorts. The analytic sample consisted of 2823 students of 77 schools. In the design schools (clusters)

were the units randomly assigned to one of two treatment conditions (a cluster-randomized design). In each school, only students of consenting teachers participated in the study. Participants were blinded as much as possible. For instance university faculty providing content deepening to both treatment groups were not aware of the treatment assignment of the teacher participants. Student achievements were measured with tests designed by the researchers using items from current assessments as well as NAEP and TIMSS assessments and writing new items on learning goals that did not exist. Data were analyzed with a two-level modeling approach (student outcomes, treatment assigned at the school level) to estimate pretest adjusted mean differences in outcomes. The analysis reveals that students whose teachers experienced content deepening integrated with analysis-of-practice in their professional development program (i.e., STeLLA) reached higher levels of science achievement than did students whose teachers received content deepening alone. The researchers conclude that the research supports the claim that engaging teachers in scientific inquiry activities to deepen their content knowledge is not enough; analysis-of-practice should be included in the PD.

In Missouri a longitudinal statewide survey of 467 middle school mathematics teachers in 91 schools was conducted in 2009, 2010, and 2011 to examine various types and amounts of teachers' professional learning activities and what impacts these activities have on student achievement (Akiba & Liang, 2016). The six types of professional learning activities focused on mathematics teaching and learning were: (a) professional development programs, (b) teacher collaboration, (c) university courses, (d) professional conferences, (e) informal communications, and (f) individual learning activities. The research questions focused on the levels of teacher participation in formal and informal professional learning, how they are associated with one another and associated with student achievement growth over four years. Data on teacher participation in professional learning activities were collected with the Teachers Opportunity to Learn (TOTL) survey that measured teachers' participation in various professional learning activities and work contexts that influence that participation. It was conducted over three years. The amounts of professional learning activities were aggregated at the school level. Of the middle schools with a Grades 6-8 configuration in the state of Missouri, 91 schools had at least 50% of the eligible mathematics teachers that participated in the TOTL project across the three years. These schools had 11.192 students with complete MAP scores on mathematics over the 4 years. Almost all teachers engaged in individual learning activities. About 75% of the teachers participated in professional development programs, in teacher collaboration and communicate informally with their colleagues to discuss mathematics teaching and learning. Attending professional conference(s) or an audience and taking university courses were less common. The amounts of teachers' participation in formal and informal professional learning activities were only weakly associated with one another. To analyze the impact of professional learning activities on student achievement growth a three level hierarchical linear model was used. School data and Teacher data were aggregated at the school level. To examine the value-added effects of six types of professional learning activities on student achievement growth, independent of other student and school-level predictors of student achievement, the analysis was controlled for both student characteristics (poverty level, ethnic minority status, and gender) and school-level background characteristics (poverty level, percentage of ethnic minority students, school size, school location, percentages of teachers with a mathematics certification, master's degree or above, mathematics major, mathematics education major, five or fewer years of teaching experience, and more than 15 years of experience). The HLM analysis for each teacher professional learning activity (professional development program, teacher collaboration, university or college courses,

professional conferences, informal communication, and individual learning activities) was conducted separately. Modest yet significant positive associations were identified between student achievement growth rates in MAP scores and school average amounts of participation in teacher collaboration, professional conferences, and informal communications with colleagues. Controlling for individual student poverty level, ethnic minority status, gender, teacher qualification and school background characteristics, one hour increase in school average amount of teacher participation in teacher collaboration results in a .01-point increase in the annual growth rate in student mathematics achievement over three years. One hour increase in school average amount of teacher participation in professional conference and informal communication results in a .15-point increase and a .23-point increase in the annual growth rate in students' mathematics scores. School average amounts of teacher participation in professional development, university courses, and individual learning activities were not significantly associated with student achievement growth. Apparently teacher-centered collaborative learning activities on mathematics teaching and learning seem to be more effective in improving student achievement than learning activities that do not necessarily involve such communications (professional development programs, university courses, individual learning activities). Teacher-driven research activities including professional conference presentation and participation were also found to be associated with student achievement growth in mathematics.

In Table 2 the studies concerning experienced teachers meeting the strict selection criteria are presented . The last column indicates the sample size which determines the selection of studies for generalizable conclusions.

Table 2: Summary of the studies concerning experienced teachers that meet the strict selection criteria

| Author, year | Secondary or primary | Selected special students | Specific subject | Intervention | Effect | Generalizability [10] |
|---|---|---|---|---|---|---|
| *Interventions focusing on the improvement of instructional behavior or teacher-student interactions and student achievement (teacher behavioral focus)* | | | | | | |
| Wasik et al., 2011 | Primary | Majority African American | Language and literacy | a) intensive, ongoing staff development and (b) books, materials, and lesson plans that support the development of children's language and literacy. 3-4 weakly training cycles. | + (Teachers' classroom quality)<br>+ (student achievement) | -<br>(N=30) |
| Biancarosa, Bryk & Dexter, 2010) | Primary school and kindergarten | - | reading, writing, and language skills | School-based one-on-one coaching of teachers observing, modeling, and catalyzing teachers' development toward more expert practice.. Six core components form the LC comprehensive literacy framework for kindergarten through second grade: interactive read-aloud, shared reading, guided reading, interactive writing, writing workshop, and word study. The components vary in their use of student grouping and the level of scaffolding provided, as well as in their focus on reading, writing, or word-level skills and knowledge. | + (children's skills, specifically literacy learning)<br>+ (value added effects of teachers)<br>effect sizes of .22, .37, and .43 in years 1, 2, and 3, respectively | ++<br>(N=287)<br>17 schools |
| Garet et al., 2010, 2011 | Primary school | - | Mathematics (rational number topics) | Three 8-hour sessions of instruction on pedagogy and content knowledge during a summer institute and 5 seminars during the following school year. In the week following each of the seminars classroom coaching was provided to assist teachers in applying the new strategy. The planned PD activities included opportunities for teachers to solve mathematics problems individually and in groups, make short oral presentations to explain how they solved problems, receive feedback on how they solved and presented their solutions, engage in discussions about the most common student misconceptions associated with topics in rational numbers, and plan lessons that they would teach during the follow-up coaching visits. The primary purpose of the coaching component of the PD program was to help teachers apply material covered in the institutes and | 0 (teacher knowledge)<br>+ the frequency with which teachers engaged in activities that elicited student thinking (effect size = 0.48).<br>0 (student achievement) | +<br>(Year 1 = 165)<br>(Year 2 N=39<br>77 schools |

[10] If no experimental design was applied and the sample size was > 400 using multi-level advanced statistics taking the nested structure into account and controlling for background variables the generalizability was scored +.
If an experimental design was applied, and no evidence of representativeness was given a sample size > 100 the generalizability was scored +.
If an experimental design was applied, and sample size > 200 the generalizability was scored ++.
If an experimental design was applied, and sample size > 200 and different schools, sectors and teachers with different teaching subjects were included the generalizability was scored +++.

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | seminars to their classroom instruction. The coaching component was designed to consist of 10 days of coaching provided through five 2-day visits to each school. During the coaching visits at each school, the facilitators focused their activities on the school's seventh-grade mathematics teachers. Each 2-day coaching visit was designed to occur immediately after one of the 5 seminar days and to link to the preceding seminar, using both individual and group activities. | | |
| Garet et al., 2016 | Elementary - Middle school | - | Mathematics | an intensive workshop focused on deepening teachers' knowledge of grades K-8 mathematics, 5 collaborative meetings during the schoolyear thereafter and a series of three one-on-one coaching sessions with video feedback where teachers 'lessons were observed and critiqued | + (teachers' mathematics content knowledge)<br>Partly + (instructional practice)<br>0 (students' achievement, on one aspect negative effect) | + (N=165, questionable validity of observation instrument) |
| Santau, Maerten-Rivera, & Huggins, 2011 | Primary school (3rd 4th and 5t grades) | English Language Learning students, one disadvantaged school | Science | The entire science curriculum for grades 3-5 was developed in curriculum units that promote science inquiry with students who may be less familiar with scientific practices. The units gradually progress along the continuum of teacher-explicit to student-initiated inquiry and to higher levels of complexity in terms of both science concepts and the level of inquiry required. Teacher guides provides the backgrounds, information and explanation of the key-concepts, the learning of students, how to incorporate English language and literacy and supplementary materials. The student booklets focused on standards-based, inquiry-driven science learning and highlight activities or strategies to foster reading and writing as part of science instruction and provide explicit guidance to promote English proficiency. The teachers followed full-day workshops during the school year. In the workshops for science experimental designs, procedures for gathering data, multiple ways of displaying the data, and conclusions based on data and evidence were discussed. In addition teachers discussed how to promote student initiative in conducting inquiry as they gradually reduce their level of guidance. For English language and literacy various literacy strategies were discussed and how they can reinforce these strategies in their instruction. | 0 (science achievement) | -<br>One large urban school<br>Specific target group, small sample |
| Glazerman & Seifullah, 2012 | Elementary | Grade 4 and 7 | Reading and mathematics | In the program teachers get performance incentives, along with tools to track their performance and improve instruction. They can earn extra pay and responsibilities through promotion to mentor or master teacher and | + & -<br>the program did not consistently raise student | +<br>N=195<br>Implementation |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | can earn annual performance bonuses based on a combination of their value added to student achievement and observations of their classroom teaching. The program includes weekly meeting of teachers and mentor, regular classroom observations by a school leadership team and pay for principals who meet implementation benchmarks | achievement<br>mixed effects regarding teacher retention<br>+ compensation effects (performance-based bonuses were paid) | was not completed |
| Van Kuijk, Deunk, Bosker and Ritzema (2016) | Primary | 2nd and 3rd grade | reading comprehension | PD program consisting of three components: 1) setting standards and performance goals for every student, 2) applying formative assessment and data use, and 3) knowledge and instruction for reading comprehension. All components being equally important within the PD program. Second- and third-grade teachers, principals and internal support coordinators participated in the program. Throughout the school year of 2011–2012, the time investment of the teachers was scheduled for 40 hr, including attending 9 after-school meetings and completing homework assignments. Participation was voluntary and free of charge. | + reading comprehension<br>0 teacher attitudes and school climates | -<br>(N=70) |
| Ostermeier, Prenzel & Duit (2010) | Secondary? | | Mathematics and science | 11 program modules were composed. Schools had to determine their problem area and choose two modules of the program. The modules served as a starting point to improve teaching. In addition written materials, in-service training or consultation was offered to the teachers developing their own classroom instruction. The program stimulates cooperation and collaboration on different levels, especially between the teachers participating in the program. The modules are a frame of reference for support and are based on the current state of research in science and mathematics education and on learning and instruction in general. Since the modules address key problem areas the teacher can locate class-related problems within the frame of modules. Examples are provided to solve the problems. | + (self-reported perception of improved classroom instruction, more cooperation activities)<br>? (student achievement, interest and competencies) | ++<br>(144 representative SINUS schools N=557 year 2000; N=527 year 2002) |
| Allen, Pianta, Gregory, Mikami, & Lun, 2011 | Secondary | - | - | a coaching program focused on improving teacher-student interactions in secondary classrooms. The instructional and motivational qualities of teachers' ongoing daily interactions with students were the focus of the study. workshop-based training. During the school year teachers got personalized coaching based on video recordings of their class sessions. The trained teacher consultants review these recordings, selected illustrative segments with positive teacher interactions and areas for improvement. The teacher was requested to observe his or her behavior and student | + (teacher-student interactions)(there was a change in teacher-student interaction qualities, but whether this is a positive of negative change is not mentioned) | -<br>(small sample size, N=76 in year 1 and N=61 in year 2) |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | reactions and to respond to the consultant prompts. Then a phone conference followed in which the consultant and the teacher discussed ways to enhance interactions using the CLASS-S system. During the school year about every two weeks this cycle was repeated. | + (student achievement, 0 in 2nd year)<br>Intervention effects on student achievement were non-significant in the first year, but significantly positive in the second year) | |
| (Dwyer et al., 2010b)(Dwyer et al., 2010a) (Magadin de Kramer et al., 2012) (Masters, Magidin De Kramer, Dwyer, Dash, & Russell, 2012) (Dash, Magidin De Kramer, O'Dwyer, Masters, & Russell, 2012) | Upper elementary and middle school | High-need schools | fourth grade English language arts (ELA), fifth grade mathematics, seventh grade ELA, and eighth grade mathematics | Online Professional Development courses. The workshops were based on a learning community model with a strong emphasis on peer-to-peer discussions. A workshop started with an orientation session and continued with six content sessions during the schoolyear. The sessions involved readings (articles, books etc.), activities (online video's, work with existing classroom materials, etc.) and discussions (responses on questions related to the readings and activities). Finally the teachers had to develop an action plan or lesson plan based on the workshop content. Each experiment was conducted across multiple states and included three rounds of data collection, each spanning three school semesters. | + self reported class-room practices<br>+ student achievement (at least one measure per trial)<br>Effect sizes ca 0.55 | ++<br>N=71 to 110 per trail, relative small sample sizes<br>More than one sector, more than one subject, experimental, more than one grade |
| Roschelle et al. (2010 | Middle schools | Advantaged and disadvantaged students | 7th and 8th grade mathematics | The intention is a more coherent and fruitful mathematical experience for both disadvantaged and advantaged learners in middle school by creating new opportunities for students to learn complex and conceptually difficult mathematics. The software SimCalc presents animations of motion. Students can control the motions of animated characters by building and editing mathematical functions in either graphical or algebraic form. Often they are asked to tell stories that correspond to the functions (and animations). The intervention integrated pedagogy, curriculum, professional development, assessment and school leadership. The target mathematics in seventh and eighth grade curriculum is proportionality and linear function, which is also important for student's science learning. The seventh year contextual theme was that students must serve as soccer team managers - training players, ordering uniforms, planning trips to games, and negotiating their salaries. The contextual theme in the eight year is that students are | + (student gains in complex concepts)<br>(time on task did not explain the student gains) | N= 67 plus 25 (relative small sample) |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | designers of electronic games using mathematics to make the games. After a two day workshop (TEXTEAMS) on the mathematical knowledge for teaching rate and proportionality, the PD got on with a 3-day summer workshop introducing the SimCalc units. In a 1-day workshop specific plans were made for how and when to use the SimCalc materials in their classrooms (writing lesson plans and thinking through their own logic for the unit). In the control groups, the business-as-usual curriculum addressed, within the same time frame as the SimCalc unit, similar basic concepts but provided less coverage of more complex concepts. Two different teacher professional development delivery models were used: a consistent model where two highly experienced mathematics teachers led all the workshops and a train-the-trainers model. Instead of standardized tests that did not capture the conceptual depth students could reach using the SimCalc technology and curricula the research team developed two tests. | | |
| Meyers, Molefe, Brandt, Zhu and Dhillon (2016) | $7^{th}$ and $8^{th}$ grades | High needs regions | Mathematics | The program provides 240 hours differentiated PD sustained over 2 years grounded in constructivist pedagogy, with opportunities for "hands-on" work that connects to teachers daily instructional practice. Collective participation in this program promotes teacher communication and collaboration to support instructional changes. eMINTS embeds these features by focusing on the promotion of four specific strategies that address issues the authors have identified as barriers to the consistent use of standards-based instruction; inquiry-based learning, high-quality lesson design, community of learners and technology integration. In addition to this intensive 240 hour PD program eMINTS also includes: 1) a specific set of school and classroom technology equipment, 2) intensive on-site training for school principals, district and school technology coordinators, and classroom teachers, 3) job-embedded coaching to enhance teachers' classroom practices and written curricula, and 4) just-in-time learning opportunities via online courses to help teachers improve their practice over time. | + (inquiry-based learning effect size ca 0.80)<br>+ technology integration (effect size ca 1.48)<br>+ mathematics achievement (effect size ca 0.16) | +<br>N=200 |
| *Interventions focusing on the improvement of teacher knowledge or teacher-efficacy and student achievement* | | | | | | |
| Watson & Beswick, 2011 | Seconday (middle years of schooling) | - | Numeracy skills | Teachers were supported in developing their own skill efficiency, their conceptual understanding and they were supported in their considerations of how they could provide similar opportunities for their pupils. The 3 two-day sessions involved presentations on student difficulties, trials of the activities in their class, and time for sharing and reflecting on the implemented changes. | + (teacher confidence)<br>0 (teachers' beliefs related to mathematics classroom)<br>(teacher knowledge)<br>+ (students' mathematical thinking) | -<br>no control group, (N=35; small sample) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Lumpe et al., 2012 | Primary | - | Science | Each year of the project teachers participated in six, 2-week long summer programs that were focused on inquiry-based instruction, science content knowledge, and science process culled from the districts' adopted curriculum. Elementary school teachers were given full-time release, to provide assistance to classroom teachers during bi-weekly visits. All the school principals were involved in this science reform effort | + (teachers' self-efficacy beliefs of the teachers)<br>- (teachers' context beliefs)<br>0 (outcome expectancy) | -<br>N=450<br>No control group |
| *Studies of the effects on student achievement by comparing different kinds of interventions in one study* | | | | | | |
| Antoniou & Kyriakides, 2011, 2013 | Primary | - | Mathematics | The teachers were expected to adopt their teaching skills to the content of mathematics that was discussed in the training sessions. The aspect that was compared between the two conditions was the different focus in the critical reflection and the action plans. In the DIA-group the specific focus was on skills that belonged to the development stage of the teacher and in the other group on the whole spectrum of knowledge, skills, attitudes and beliefs about teaching. So although both experimental treatments encouraged and utilized critical reflection of teachers on their teaching practices, teachers employing the DIA were asked to reflect on those aspects of their teaching practice that were found to be related with their priorities for improvement based on the stage at which they were situated". | + (quality of teacher scores)(DIA group)<br>0 (quality of teacher scores 1 year after intervention)(both groups)<br>+ (student achievement)(DIA) | ++<br>(N=130)<br>Effect size = 0.38 |
| Heller, Daehler, Wong, Shinohara, & Miratrix, 2012 | Primary (4th grade) | - | Science | The interventions were: 1) a Teaching Cases course (with discussions of pre-structured written cases of classroom practice), 2) a Looking at Student Work course (involving analysis of teachers' own student work in conjunction with concurrent teaching) and 3) a Metacognitive Analysis course (in which teachers engage in metacognitive reflection on their own learning experience). | + (teacher outcomes)( "Looking at Student Work" and "Teaching Cases")<br>+ (student test scores)<br>13% more growth compared to control group | ++<br>(N=271) |
| Penuel, Gallagher, & Moorthy, 2011 | Secondary | - | Science | recently adopted an *Understanding by Design* (UbD) approach is a framework for designing curricular units of instruction that centers on the big ideas, essential questions, and authentic performances. The models provide detailed guidance on pedagogical strategies linked to specific subject matter domains. The programs had essential features for high-quality professional development and focused on supporting for teachers learning to design sequences of instructions. A total of 53 teachers from 19 middle schools participated in the study. Teachers were randomly assigned to the programs. One group got explicit instruction in the principles of (UbD) | + (student performance)(ESBD and Hybrid programs)<br>0 (student performance)(IES group) | -<br>N=53, small sample |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | to teach Earth system science. They didn't get a specific curriculum, but designed instructional experiences and reorganized existing curricular materials (Earth Science by Design, ESBD). Another group worked with a 10-module middle school curriculum (Investigating Earth Science, IES) with the specific purpose of preparing teachers to implement instructional sequences using the materials as laid out. They did not receive explicit instruction in the models of teaching that underlay IE. The third experimental group was a blend of the other two conditions (Hybrid). Teachers received explicit instruction in the UbD models of teaching and were also immersed in the practice of design as in the ESBD condition. The control group did not receive any guidance but were free to pursue available professional development opportunities. | | |
| Powell, Diamond, Burchinal, & Koehler, 2010 | Kindergarten classrooms | - | Early literacy | Primary emphasis was given to classroom strategies to improve children's oral language skills and code-focused skills. Attention to oral language included instructional practices aimed at improving children's vocabulary knowledge, listening comprehension skills, and syntactic knowledge. Teachers were given printed copies and demonstrations of phonological awareness activities developed by Adams, Foorman, Lundberg, and Beeler (1998) and were encouraged to emphasize letter sounds when teaching letter names and to use writing to promote letter knowledge. The intervention gave secondary emphasis to instructional practices to improve children's knowledge of print concepts. Teachers participated voluntarily and received a modest level of monetary compensation for their time. The intervention comprised a 2-day workshop providing an overview of the intervention content, with emphasis on demonstration and guided discussion of evidence-based practices, followed by expert coaching. Coaching sessions (seven in total) were to focus on each of the two primary child outcome areas (oral language and code-focused skills), with the specific sequence and content individualized to classroom circumstances. In the on-site coaching condition, the coach observed classroom activities of the teacher followed by feedback. In the remote coaching condition, teachers submitted a videotape of a targeted instructional practice for the coach to review. During the semester of participation in the intervention, teachers in the remote condition were given use of a case-based hypermedia resource developed for the intervention. In the on-site | + gains in code-focused instruction<br>+ children whose teachers received remote coaching showed larger gains on the Peabody Picture Vocabulary Test and were more likely to exhibit initial sound-matching skills | -<br>N= 88, small sample |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | condition coaches brought a laptop computer with the hypermedia resource for the purpose of showing a video exemplar. Teachers in the on-site condition also could receive copies of published articles included in the hypermedia resource. | | |
| Fishman et al., 2013 | Secondary | - | Science | The professional development content is held constant in both conditions while the research is focused on the differences in teachers' learning in terms of changes in beliefs and knowledge, teachers' classroom practices and student outcomes. The PD in this study is designed to prepare high school teachers to implement a year-long environmental science curriculum and employs a pedagogical approach called Learning-for-Use, designed to develop understanding through cycles of motivation, knowledge construction, and knowledge organization. The PD is intended to increase the probability that teachers' curriculum enactment is consistent with designers' intentions, and leads to desired learning outcomes. The curriculum consists of three units meant to be completed over an academic year. One condition was a week-long (48 hr) face-to-face workshop spread over 6 days. The other was an online workshop designed to be completed by teachers asynchronously at their own pace. | + (teacher CK)(both online and face-to-face conditions)<br>+ (teachers' personal beliefs)( both online and face-to-face conditions)<br>0 (teachers' impersonal beliefs)( both online and face-to-face conditions)<br>+ (students' science scores improved in both conditions) | -<br>(N=49, small sample size) |
| Taylor et al. 2016 | Upper elementary | | Science | A conceptual framework that guides teachers' analysis of practice in the PD program and their classroom science teaching practice is developed from research on effective science teaching. The study compares the effect on students' science achievement of a teacher professional development program that integrates science content deepening with analysis-of-practice (i.e., the treatment) relative to a comparison PD program of equal duration and intensity that includes only content deepening (teacher content knowledge and confidence in teaching science effectively). Central in the program is the mutual interaction of teachers in small grade-level study groups, discussing each other's lessons and video-cases. The program starts with a two-week summer institute. During the fall teachers teach the study-lessons and analyze the videos in small groups and in the winter this shifts from implementing provided lesson plans to developing their own lesson plans. | + student achievement effect size = 0.52 (engaging teachers in scientific inquiry activities to deepen their content knowledge is not enough; analysis-of-practice should be included in the PD) | +<br>N= 144 |
| (Akiba & Liang, 2016). | Middle schools | | Grades 6-8 configuration, mathematic | Not an effect study. Survey research to examine various types and amounts of teachers' professional learning activities and what impacts these activities have on student achievement (a) professional development | + modest yet positive association between PD features and student achievement | +<br>(not an effect study)<br>N= 467 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | s | programs, (b) teacher collaboration, (c) university courses, (d) professional conferences, (e) informal communications, and (f) individual learning activities. The research questions focused on the levels of teacher participation in formal and informal professional learning, how they are associated with one another and associated with student achievement growth over four years | One hour increase in school average amount of teacher participation in professional conference and informal communication results in a .15-point increase and a .23-point increase in the annual growth rate in students' mathematics scores. | |

# 5 Answering the research questions

In chapter 3 the review studies were summarized including features of effective interventions. The following activities, which are in line with features of effective PD identified in review studies, were identified in our review of empirical studies in chapter 4:

1. One-on-one coaching, entailing observing, modelling, and catalysing teachers' development toward more expert practice.
2. Training teacher leaders how to lead PD courses, how to introduce theories and instructional practices to teachers and on growth supporting coaching techniques, to serve as expert teachers and school based literacy coaches, works too. Well-specified and well supported coaching initiatives cause positive changes in student learning.
3. Introducing a strong focus on subject knowledge and pedagogy in workshops and seminars and follow up coaching to assist teachers to apply new teaching strategies.
4. Reflecting only on the teaching skills that belong to the teacher's development stage.
5. Allowing teachers to deliberately design and plan their own changes in classroom practice, thus taking into account specific contexts and personal needs.
6. Classroom-based learning in which a video-case based analysis of practice plays a central role.
7. Experts model, coach and fade (as in cognitive apprenticeship) and teachers gradually and collaboratively focus on more teacher-directed change and enactment.
8. Online tools can effectively support teacher change.
9. Discussions of pre-structured written cases of classroom practice.
10. Looking at student work (involving analysis of teachers' own student work in conjunction with concurrent teaching.
11. Learning community model with a strong emphasis on peer-to-peer discussion.
12. An hour increase in school average amount of teacher participation in professional conference and informal communication was associated with a 0.15-point increase and a 0.23-point increase in the annual growth rate in students' mathematics scores
13. Teacher centered collaborative learning activities on mathematic teaching and learning are more effective than activities that do not involve such collaboration.
14. Presenting teachers with different modules giving them the opportunity to choose relevant modules for their own development.

Furthermore, studies focusing on ECT reveal that features of effective induction depend on the goals of the PD. If retention is the goal then initial teacher preparation (certification), to ensure an acceptable level of beginning teaching skills, is to be targeted. Once this phase has been passed, school culture, incentives and resources, workload reduction, classroom assistance and working with more experienced teachers, help retaining ECTs. If increasing teaching quality is the aim then the desired outcome should be clearly defined and conflicting aims should be avoided. Workload reduction constrains the development of teaching skills. This conflicts with the aim of raising retention rates, as well as with the well-being of beginning teacher, for which workload reduction in beneficial.

What the reviews on the effectiveness of PD interventions and programs, including those on induction, do not explicitly reveal are theories of action or theoretical assumptions underlying the design principles and their explanatory power. Although some reviews mention the importance of these theoretical insights (Kennedy, 2016; van Veen et al., 2010), our search did not yield reviews explicitly and systematically analyzing these theories as a feature of effectiveness. This can be due to the fact that this theme was out of scope in the reviews, or, more likely, because the design of PD interventions and programs is only incidentally guided by theory on (adult) learning.

Although most studies on the effectiveness of PD interventions are characterized as being a-theoretical, most of them are based on those recent notions of effective features of PD, and in this way do have a theory of action.

Next, some reviews and (meta) syntheses attempt to theoretical underpinnings and explain empirical findings. Explorations include theories on adult learning, motivational theories, and cognitive and (social) constructivist perspectives on learning. Contemporary theoretical insights on how students learn and on how teachers learn seem to overlap, inviting to align teacher learning activities with those of those of students. Examples of attempts in this direction can be found in Knapp (2003), Broad & Evans (2006) and Timperley et al. (2007).

Knapp states that what is going on in classroom practice (the enactment of student learning opportunities) and the results of that practice (student learning) may shape what participants gain in professional development, either as the content for their professional learning, the motivation for it, or both. In turn, what professionals learn through engagement with professional development activities can clearly shape what they do subsequently in the classroom. Knapp suggests that contemporary theoretical approaches to learning and teaching (e.g. Bransford, Brown and Cocking, 1999) can be applied to the learning of students as well as to the learning of teachers/professionals and may well serve to explain why the features mentioned are effective.

Theoretical arguments similar to the ones used by Knapp (2003) can be found in the very comprehensive 'best evidence synthesis' on teacher professional learning and development by Timperley et al. (2007). Theorizing on the findings, the authors contend that the evidence of their syntheses supports the idea that PD interventions and programs with sustainable and substantive effects on student outcomes combine some important elements that build on and interact with each other. Elements like grounding learning in immediate problems of practice, deepening pedagogical content knowledge and knowledge on assessment, and building on and challenging theories of practice in an ongoing process of inquiry. The evidence seems to indicate that teacher knowledge on these issues has to be developed in the context of the analyses of student outcomes. With regard to teacher learning processes, like Knapp (2003) in this synthesis it is suggested that ‘what goes on in the black box of teacher learning, is fundamentally similar to student learning’ (pXV). From a cognitive point of view, three iterative processes play an important role: the process of cuing prior knowledge, the process of developing awareness for new information and the process of creating (cognitive) dissonance. According to the authors, professional learning of experienced teachers is different from that of pre-service teachers, in that the former ones bring a wealth of ideas, (often tacit) knowledge and well-formed positions to the learning process. Reconstruction of professional knowledge often is more difficult than its original construction. Theoretical insights like these are

again in line with the general theories on learning as described by Bransford, Brown and Cocking (1999) and more specific applications of those theories to teacher learning (Bransford, Derry, Berliner, Hammerness, & Beckett, 2005). These include "the engagement of prior understandings and preconceptions about how the world works; a deep foundation of factual and conceptual knowledge, organized in ways that allow their retrieval; and a 'metacognitive' awareness that allows them to take control of their own learning by defining learning goals and monitoring progress towards them" (Timperley et al., 2007). Drawing on this synthesis Timperley (2008) stresses the importance of teachers' developing the self-regulatory skills they need in order to, independently of context and experts, judge the impact of their teaching on valued student outcomes. The extent to which teachers develop self-regulatory skills is one of the most powerful determinants of sustainable and ongoing improvement. These theoretical notions seem to match the characteristics of effective PD programs described by the authors.

Broad et al. (2006) argue that differentiation of professional development practices is critical to meeting the unique learning needs of experienced teachers due to their individual developmental and experiential career paths and contexts. Furthermore, for experienced teachers (adult learners), motivation for professional learning is influenced by perceived relevance, meaningfulness and choice. Thus, it might not be the specific career stage in which a teacher resides that is decisive for learning needs and drivers. From a theoretical perspective, they argue that the key to teachers undertaking professional learning is developing competence and a sense of self-efficacy by directing one's efforts to meet student needs. They theoretically underpin their argument citing Wlodkowski, by stating that Professional development occurs when teachers engage in "substantive actions, either individually or together, that require complex thinking to construct new skills or deeper meaning" (Wlodkowski, 2003, p. 40 cited by Broad & Evans, 2006, p. 15). Adults are motivated to learn when they are in an atmosphere that is safe and respectful, when they are related to others, when content is meaningful, when they can foster some autonomy, when learning tasks include their understanding and beliefs and when they experience that the learning goals are worth striving for.

In short, it is worthwhile approaching the effectiveness of PD interventions and programs from (different) theoretical angles and there are theories on (teacher) learning that are persuasive to be applied to the empirical findings on effective features of PD. Based on her review covering the most important research on PD effectiveness from 1979-2011, Kennedy, on this topic states:

> "… education research is at a stage in which we have strong theories of student learning, but we do not have well-developed ideas about teacher learning, nor about how to help teachers incorporate new ideas into their ongoing systems of practice <….> If we can tie our research designs and our PD models more closely to underlying theories of teacher motivation and teacher learning, we will learn more from our studies" (Kennedy, 2016).

Authors have repeatedly highlighted the importance of explicating a 'theory of action or change' underlying the design of PD interventions and programs (Kennedy, 2016; van Veen et al., 2010;

Wayne et al., 2008). Timperley et al. (2007) and Muijs et al. (2014) contend that contemporary theories on learning are quite universal and may well apply to students as well as adult learners/teachers.

In our review we did not find ample evidence for differential effectiveness of PD programs across distinctive stages of the teaching careers. Effective features seem to be quite universal even though teacher needs change over time. The universal features include meeting the needs of the target population. This means that the learning activities should vary across the career. What is important, regardless of career stage, is that we bear in mind that all learning is based on previous learning. ECT's and experienced teachers (as their students) learn with preconceptions about how the world works. If their initial understanding is not engaged, they may fail to grasp the new concepts and information that are taught or may learn them superficially and revert to their preconceptions in real situations (Bransford et al, 1999). This principle asks to be addressed in all situations. ECTs have specific needs that experienced teachers do not have or they have them to a lesser degree. Induction programs are specific PD programs catering for these needs. Reducing workload and enculturation activities seem less necessary or appropriate for experienced teachers.
From a perspective of professional growth, conceptualized as growing teaching quality, studies reveal that effectiveness of PD can be enhanced by adapting content, form and activity to the specific professional level (level of teaching quality) of the teacher (Antoniou et al., 2011; Antoniou & Kyriakides, 2013; Kyriakides, Creemers, & Antoniou, 2009; van de Grift et al., 2011). Ostermeier et al. (2010) also revealed that experienced teachers report different kinds of support needs. Future research on the enhancement of the development of teaching quality through PD interventions may also contribute to pedagogies to enhance teacher learning in all career stages.

# 6. Conclusions

With respect to the first research question, concerning features of effective PD, recent studies have confirmed and replicated previous insights. These general features summarized by Desimone in 2009 are: content focus, active learning, coherence, duration and collective participation. What stands out in the more recent studies is that these features are implemented in different ways. Not every feature is equally relevant in each program. This seems to be determined by the specific goals of the program and context in which the program is implemented. The role of pedagogical content knowledge (PCK) has been confirmed in recent studies too. PCK revolves around the many different ways to explain content as well as considering the way students learn, which problems the encounter, misconceptions. The relevance and importance of PCK is resonated by teachers in all career stages (Louws, 2016).

Besides these effective PD features, other relevant characteristics become apparent based on ECT studies. Effective features of interventions aiming to reduce early career attrition (leaving education) are: having a coach / mentor of the same school subject; having the opportunity to prepare classes together with teachers who teach the same school subject; regularly scheduled appointments for consultation and cooperation; and being part of an external network. Workload reduction has less effect on retention according to the work of Ingersoll et al. (2004), while workload reduction reduces attrition in the Dutch context (Helms-Lorenz et al., 2015). As the number of components in the induction programs increase, the probability of ECT turnover decreases. So the more attention is better. What is also proved to be of importance is that the school context plays a role in the effectiveness of induction-arrangements. Induction programs are less/not effective in schools with a high percentage of students with lower socio-economic status ("high poverty schools") (Ingersoll et al., 2004). This analysis showed that the intervention goals should determine the choice and design of the intervention (backward design). If an intervention for ECTs aims to accelerate the development of teaching skills, then workload reduction appears to be counter-productive. One-on-one coaching is more effective for achieving this goal (Helms-Lorenz et al., 2015).

With regard to our second research question, concerning the underlying theoretical assumptions of PD interventions, we found that theoretical assumptions are rarely explicitly stated. Yet, upon reflecting on the studies between 2010-2016 we observe that indeed that most studies rely on what is known about effective features of professional development. These insights are used as design principles. Here the didactic core principle of Backward Design seems to play a central role: thinking from the goal to be achieved with a professional intervention, determines the intervention activities and a relevant selection is made of effective features. These studies demonstrate how features play a role in certain contexts. In more and more interventions where the goal is to influence the daily teaching practice and student learning, the own teaching is central, discussed and observed, and involve attention for PCK.

A limited number of studies explicitly refer to psychological principles of adult learning (Meyers et al., 2016). The explicit use of this kind of theoretical principles appears to be useful in the design and implementation of PD interventions. Kennedy (2016) and van Veen et al. (2010) explicate the need of

a ‘theory of action' (the reasons why the intervention would contribute to the learning of teachers) as a relevant and crucial activity for both those who design and undergo the intervention. This is also consistent with the call, for example, of John Hattie (2010) who, in the context of student learning, urges that it is useful to be transparent to students about the objectives and the associated didactics. More in general, it is concluded, perhaps quite obvious though often absent in PD programs, that teacher learning is similar to student learning and that the same principle apply. A good PD program is like a good lesson for students, taking into account all that we know from effective learning and teaching. In other words, good PD requires good teacher educators as well to organize teacher learning in the PD.

With regard to our third research question, concerning the differences between interventions for ECTs and experienced teachers, we conclude that interventions for ECTs have different characteristics, taking the specific learning needs of ECTs into account. These needs differ from the needs of experienced teachers. Starters are much more focused on learning more and becoming socialized into the profession. The research presented clearly shows that supporting these needs of ECT is useful in the form of induction arrangements.

The difference in approaching ECTs and experienced teachers may have to do with the willingness of novices to learn from experienced teachers. When the ECT is convinced that the experienced teacher is an (more of an) expert, the conditions for learning become advantageous (of course this is not always the case, sometimes the ECT is (rightly) not convinced). Given the difference in experience the ECT more readily be convinced of the potential of learning from the more experienced teacher, especially when his/her teaching is a “bumpy road”. Experienced teachers are not easily convinced. The experienced teacher, who has many learning experiences including failures, is less willing to learn from a colleague or an external "expert". For interventions for experienced teachers, it is necessary to pay close attention to how the intervention can influence the development of teachers. A "theory of action" can be of great value in this case. Designing a theory of action, forces the person who designs the intervention to think about the objectives of the intervention, the desired outcomes, to reflect on the initial situation of the teachers, and to adapt the intervention content and activities to the different needs of the participating teachers. There should also be space to explicate the own pedagogy and to allow the venting of other learning strategies of the participants and to involve the chosen learning strategy. With experienced teachers, it makes sense to organize interventions within schools and between schools with teachers of similar subjects, as in Lesson Study. It is advisable to link ECTs in these interventions, provided that the specific needs of ECTs and experienced teachers are not undermined.

**Limitations and future research**

The number of strictly controlled studies is scare, but at the same time an increase in this kind of costly and difficult to implement research is evident. Most[11] of the studied interventions contain multiple features, yet the reported analyses do not include the impact of isolated features in order to investigate their unique contribution to the effectiveness of the program. It is possible that the systemic interaction of different features within certain contexts determines effectiveness, and at the same time that unique distinct features contribute uniquely to the effects. Future research should focus on replicating previous small scale studies to strengthen the body of knowledge. Large

[11] Exceptions are Akiba et al. (2016), Helms-Lorenz et al. (2015), Ingersoll et al. (2004) and Kang et al. (2012).

scale intervention implementation has a drawback because it introduces additional uncontrollable implementation complexity. Controlled trails with representative samples should be preferred to determine effectiveness. After proven effectivity large scale implementation should not be expected to replicate effects, but should rather focus on implementation processes and improvement. Additionally designs should ideally be refined to allow for analyses to determine the impact of different features of interventions.

Furthermore, the field would be advanced with more fundamental research into how teachers differ in the way they learn (based on gender, concerns, motivation, subject pedagogy). This aspect would advance the quality of interventions tailored to the needs of teachers.

Finally, in line with Kennedy (2016) it is recommended to focus not so much on the specific effective features, rather how they are combined in a theory of action.

# References


Achinstein, B. (2006). New teacher and mentor political literacy: Reading, navigating and transforming induction contexts. *Teachers and Teaching, 12*(2), 123-138. doi:10.1080/13450600500467290

Akiba, M., & Liang, G. (2016). Effects of teacher professional learning activities on student achievement growth. *The Journal of Educational Research, 109*(1), 99-110. doi:10.1080/00220671.2014.924470

Algozzine, B., Gretes, J., Queen, A. J., & Cowan-Hathcock, M. (2007). Beginning teachers' perceptions of their induction program experiences. *The Clearing House, 80*(3), 137-143.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science (New York, N.Y.), 333*(6045), 1034-1037. doi:10.1126/science.1207998

Antoniou, P., Kyriakides, L., & Creemers, B. (2011). Investigating the effectiveness of a dynamic integrated approach to teacher professional development. *Center for Educational Policy Studies Journal, 1*(1), 13-41.

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education, 29*, 1-12. doi:10.1016/j.tate.2012.08.001

Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education, 27*(1), 10-20. doi:http://dx.doi.org/10.1016/j.tate.2010.08.007

Berliner, D. C. (2004). Expert teachers: Their characteristics, development and accomplishments. *R.Batllori i Obiols, A.E Gomez Martinez, M.Oller i Freixa, & J.Pages i.Blanch (Eds.), De La Teoria….A L'aula: Formacio Del Professorat Ensenyament De Las Ciències Socials,* , 13-28.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463-482. doi:http://dx.doi.org/10.1016/S0883-0355(02)00004-6

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal, 111*(1), 7-34.

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

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. P. B. McGaw (Ed.), *International encyclopedia of education (third edition)* (pp. 548-556). Oxford: Elsevier. doi:http://dx.doi.org/10.1016/B978-0-08-044894-7.00654-0

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15. doi:10.3102/0013189X033008003

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bransford, J., Derry, S., Berliner, D., Hammerness, K., & Beckett, K. L. (2005). Theories of learning and their roles in teaching. In L. Darling-Hammond, & Bransford J. (Eds.), *Preparing teachers for a*

*changing world: What teachers should learn and be able to do* (pp. 40-87). San Francisco, CA: Jossey-Bass.

Broad, K., & Evans, M. (2006). *A review of literature on professional development content and delivery modes for experienced teachers*. Toronto: University of Toronto, Ontario Institute for Studies in Education.

Bromme, R. (2001). Teacher expertise. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 15459-15465). Oxford: Pergamon. doi:http://dx.doi.org/10.1016/B0-08-043076-7/02447-5

Bullough, R. V., Jr. (2012). Mentoring and new teacher induction in the united states: A review and analysis of current practices. *Mentoring & Tutoring: Partnership in Learning, 20*(1), 57-74.

Caena, F. (2011). *Literature review; quality in teachers' continuing professional development.* European Commission; Education and training 2020-thematic working group "professional development of teachers".

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947-967. doi:http://dx.doi.org/10.1016/S0742-051X(02)00053-7

Cochran-Smith, M., Feiman-Nemser, S., McIntyre, D. J., & Demers, K. E. (2008). *Handbook of research on teacher education: Enduring questions in changing contexts* Taylor \& Francis.

Cordingley, P. H., S. Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., & Coe, R. (2015). *Developing great teaching - lessons from international reviews of effective professional development*. Teacher Development Trust.

Cordingley, P., Bell, M., Isham, C., Evans, D., & Firth, A. (2007). *What do specialists do in CPD prgrammes for which there is evidence of positive outcomes for pupils and teachers? technical*

*report.* ( No. 1504T). London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Creemers, B., Kyriakides, L., & Antoniou, P. (2013). *Teacher professional development for improving quality of teaching*. Dordrecht: Springer Science + Business Media. doi:10.1007/978-94-007-5207-8_2

Dall'Alba, G., & Sandberg, J. (2006). Unveiling professional development: A critical review of stage models. *Review of Educational Research, 76*(3), 383-412. doi:10.3102/00346543076003383

Darling-Hammond, L., Bransford, J., LePage, P., Hammerness, K., & Duffy, H. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do.* (1st ed.). San francisco, CA: Jossey-Bass.

Dash, S., Magidin, d. K., O'Dwyer, L. M., Masters, J., & Russell, M. (2012). Impact of online professional development or teacher quality and student achievement in fifth grade mathematics. *Journal of Research on Technology in Education, 45*(1), 1-26. doi:10.1080/15391523.2012.10782595

Day, C. (1999). *Developing teachers: The challenges of lifelong learning*. London: Falmer Press.

Day, C. (2008). Committed for life? variations in teachers' work, lives and effectiveness. *Journal of Educational Change, 9*, 243-260. doi:10.1007/s10833-007-9054-6

Day, C., Sammons, P., Stobart, G., Kington, A., & Gu, Q. (2007). *Teachers matter: Connecting work, lives and effectiveness*. Maidenhead: Open University Press.

Day, C., & Gu, Q. (2007). Variations in the conditions for teachers' professional learning and development: Sustaining commitment and effectiveness over a career. *Oxford Review of Education, 33*(4), 423-443.

de Kramer, R. M., Masters, J., O'Dwyer, L. M., Dash, S., & Russell, M. (2012). Relationship of online teacher professional development to seventh-grade teachers' and students' knowledge and practices in english language arts. *Teacher Educator, 47*(3), 236-259.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199. doi:10.3102/0013189X08331140

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition.* (research report No. ORC-80-2). Berkeley: University of California (Berkeley). Operations Research Center.

Dunst, C. J., Bruder, M. B., & Hamby, D. W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews, 10*(12), 1731-1744. doi:10.5897/ERR2015.2306

Evens, M., Elen, J., & Depaepe, F. (2015). Developing pedagogical content knowledge: Lessons learned from intervention studies. *Education Research International.* doi:10.1155/2015/790417

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103*(6), 1013-1055.

Fessler, R., & Christensen, J. C. (1992). *The teacher career cycle: Understanding and guiding the professional development of teachers.* Needham-Heights: M.A. Allyn and Bacon.

Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education, 64*(5), 426-438. doi:10.1177/0022487113494413

Fresko, B., & Nasser-Abu Alhija, F. (2015). Induction seminars as professional learning communities for beginning teachers. *Asia-Pacific Journal of Teacher Education, 43*(1), 36-48. doi:10.1080/1359866X.2014.928267

Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6*(2), 207-226. doi:10.3102/00028312006002207

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., . . . Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (evaluation report No. NCEE 2016-4010). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., . . . and Doolittle, F. (2011). *Middle school mathematics professional development impact study findings after the second year of implementation.* ( No. NCEE 2011-4024). Washinton, DC.: IES; National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., . . . Doolittle, F. (2010). *Middle school mathematics professional development impact study findings after the First year of implementation.* ( No. NCEE 2010-4009). Washinton, DC.: National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences, U.S. Department of Education.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches.* ( No. REL 2014-

010).Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, M. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study.* ( No. NCEE 2010-4027). washington DC: NCEE.

Glazerman, S., Seifullah, A., & Mathematica, P. R. (2012). *An evaluation of the chicago teacher advancement program (chicago TAP) after four years. final report*. Mathematica Policy Research, Inc.

Guskey, T. R. (2014). Evaluating professional learning. In S. S. Billett, C. Harteis & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (pp. 1215-1235). Dordrecht: Springer Netherlands. doi:10.1007/978-94-017-8902-8

Guskey, T. R. (2000). *Evaluating professional development.* SAGE Publications.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495-500. doi:10.1177/003172170909000709

Haigh, M. A., & Anthony, G. J. (2012). Induction and efficacy: A case study of new zealand newly qualified secondary science teachers. *Journal of Science Teacher Education, 23*(6), 651-671. doi:10.1007/s10972-012-9285-0

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review, 30*(3), 466-479. doi:10.3386/w16606

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY, US: Routledge/Taylor & Francis Group.

Hawlet, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In l. Darling-Hammond, & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127-150). San Francisco, CA: Jossey-Bass.

Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching, 49*(3), 333-362. doi:10.1002/tea.21004

Helman, L. (2006). What's in a conversation? Mentoring stances in coaching conferences and how they matter. Developing new leaders for new teachers. In B. Achinstein, & S. Z. Athanases (Eds.), *Mentoring in the making: Developing new leaders for new teachers* (pp. 69-82). New York, NY: Teachers College Press.

Helms-Lorenz, M., van, d. G., & Maulana, R. (2015). Longitudinal effects of induction on teaching skills and attrition rates of beginning teachers. *School Effectiveness and School Improvement,* , 1-27. doi:10.1080/09243453.2015.1035731

Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education, 25*(1), 207-216. doi:10.1016/j.tate.2008.09.001

Huberman, M., Gronauer, M. M., & Marti, J. (Eds.). (1993). *The lives of teachers*. New York: Teacher College Press.

Huberman, M. (1989). On teachers' careers: Once over lightly, with a broad brush. *International Journal of Educational Research, 13*(4), 347-362. doi:http://dx.doi.org/10.1016/0883-0355(89)90033-5

Ingersoll, R., & Smith, T. M. (2004). Do teacher induction and mentoring matter? *NAASP Bulletin, 88*(638), 28-40.

Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research, 81*(2), 201-233. doi:10.3102/0034654311403323

Kang, S., & Berliner, D. C. (2012). Characteristics of teacher induction programs and turnover rates of beginning teachers. *Teacher Educator, 47*(4), 268-282.

Kennedy, M. (1998). *Form and substance in inservice teacher education. research monograph no. 13*. Arlington, VA: National Science Foundation, University of Wisconsin-Madison.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research,* doi:10.3102/0034654315626800

Kini, T., & Podolsky, A. (2016). *Does teaching experience increase teacher effectiveness? A review of the research.* Palo Alto: Learning Policy Institute.

Knapp, M. S. (2003). Professional development as a policy pathway. *Review of Research in Education, 27*, 109-157. doi:DOI: 10.3102/0091732X027001109

Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12-23. doi:http://dx.doi.org/10.1016/j.tate.2008.06.001

Ladd, H. F., & Sorensen, L. (2015). *Returns to teacher experience: Student achievement and motivation in middle school.* (working paper No. 112). Washington: Calder (National Center for Analysis of Longitudinal Data in Education Research).

Lazovsky, R., & Reichenberg, R. (2006). The new mandatory induction programme for all beginning teachers in israel: Perceptions of inductees in five study tracks. *Journal of Education for Teaching, 32*(1), 53-70. doi:10.1080/02607470500510977

Long, J. S., McKenzie-Robblee, S., Schaefer, L., Steeves, P., Wnuk, S., Pinnegar, E., & Clandinin, D. J. (2012). Literature review on induction and mentoring related to early career teacher attrition and retention. *Mentoring & Tutoring: Partnership in Learning, 20*(1), 7-26.

Lopez, A., Lash, A., Schaffner, M., Shields, P., & Wagner, M. (2004**).** *Review of research on the impact of beginning teacher induction on teacher quality and retention.* ( No. D-01-CO-0059/0004). Washington, D.C.: SRI International.

Louws, M. L. (2016). *Professional learning : What teachers want to learn*. Leiden: ICLON Phd Dissetations Series..

Luft, J. A., Firestone, J. B., Wong, S. S., Ortega, I., Adams, K., & Bang, E. (2011). Beginning secondary science teacher induction: A two-year mixed methods study. *Journal of Research in Science Teaching, 48*(10), 1199-1224. doi:10.1002/tea.20444

Lumpe, A., Czerniak, C., Haney, J., & Beltyukova, S. (2012). Beliefs about teaching science: The relationship between elementary teachers' participation in professional development and student achievement. *International Journal of Science Education, 34*(2), 153-166. doi:10.1080/09500693.2010.551222

Maandag, D. W., Deinum, J. F., Hofman, A. W. H., & Buitink, J. (2007). Teacher education in schools: An international comparison. *European Journal of Teacher Education, 30*(2), 151-173. doi:10.1080/02619760701275552

Masters, J., Kramer, R. M. d., O'Dwyer, L., Dash, S., & Russell, M. (2012). The effects of online teacher professional development on fourth grade students' knowledge and practices in english language arts. *Journal of Technology and Teacher Education, 20*(1), 21-46.

Maulana, R., Helms-Lorenz, M., & van, d. G. (2015). A longitudinal study of induction on the acceleration of growth in teaching quality of beginning teachers through the eyes of their students. *Teaching and Teacher Education, 51*, 225-245. doi:10.1016/j.tate.2015.07.003

Meyers, C. V., Molefe, A., Brandt, W. C., Zhu, B., & Dhillon, S. (2016). Impact results of the eMINTS professional development validation study. *Educational Evaluation and Policy Analysis, 38*(3), 455-476. doi:10.3102/0162373716638446

Moir, E., & Gless, J. (2001). Quality induction: An investment in teachers *Teacher Education Quarterly, 28*(1), 109-114.

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. doi:10.1080/09243453.2014.885451

Nielsen, D. C., Barry, A. L., & Addison, A. B. (2007). A model of a new-teacher induction program and teacher perceptions of beneficial components. *Action in Teacher Education, 28*(4), 14-24.

O'Dwyer, L. M., Masters, J., Dash, S., Magidin De Kramer, R., Humez, A., & Russell, M. (2010). *E-learning for educators; effects of on-line professional development on teachers and their students: Findings from four randomized trials*. Chestnut Hill: Technology and Assessment Study Collaborative. Lynch School of Education, Boston College.

Onstenk, J. H. A. M. (1997). *Lerend leren werken. brede vakbekwaamheid en de integratie van leren, werken en innoveren.*(Phd).

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376-407. doi:10.3102/0034654311413609

Ostermeier, C., Prenzel, M., & Duit, R. (2010). Improving science and mathematics instruction: The SINUS project as an example for reform as teacher professional development. *International Journal of Science Education, 32*(3), 303-327. doi:10.1080/09500690802535942

Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal, 48*(4), 996-1025. doi:10.3102/0002831211410864

Platform Onderwijs2032. (2016). *Ons Onderwijs2032. Eindadvies*. Den Haag: Platform Onderwijs2032.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology, 102*(2), 299-312. doi:10.1037/a0017763

Reiman, A., & Thies-Sprinthall, L. (1998). *Mentoring and supervision for teacher development*. New York: Longman.

Reiman, A. J. (1999). The evolution of the social roletaking and guided reflection framework in teacher education: Recent theory and quantitative synthesis of research. *Teaching and Teacher Education, 15*(6), 597-612. doi:http://dx.doi.org/10.1016/S0742-051X(99)00016-5

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., . . . Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing

middle school mathematics: Three large-scale studies. *American Educational Research Journal,* doi:10.3102/0002831210367426

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78. doi:10.1037/0003-066X.55.1.68

Santau, A. O., Maerten-Rivera, J. L., & Huggins, A. C. (2011). Science achievement of english language learners in urban elementary schools: Fourth-grade student achievement results from a professional development intervention. *Science Education, 95*(5), 771-793. doi:10.1002/sce.20443

Schaefer, L., Long, J., & Clandinin, D. (2012). Questioning the research on early career teacher attrition and retention. *Alberta Journal of Educational Research, 58*(1), 106-121.

Scher, L., & O'Reilly, F. (2009). Professional development for K-12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness, 2*(3), 209-249. doi:10.1080/19345740802641527

Simon, F. A. (2011). *Efficacy development in new teacher study groups* (Order No. 3449473). Available from ProQuest Dissertations & Theses A&I. (862553673).

Stanulis, R. N., Brondyk, S. K., Little, S., & Wibbens, E. (2014). Mentoring beginning teachers to enact discussion-based teaching. *Mentoring & Tutoring: Partnership in Learning, 22*(2), 127-145. doi:10.1080/13611267.2014.902556

Stanulis, R. N., Little, S., & Wibbens, E. (2012). Intensive mentoring that contributes to change in beginning elementary teachers' learning to lead classroom discussions. *Teaching and Teacher Education: An International Journal of Research and Studies, 28*(1), 32-43.

Steffy, B. E., & Wolfe, M. P. (2001). A life-cycle model for career teachers. *Kappa Delta Pi Record, 38*(1), 16-19. doi:10.1080/00228958.2001.10518508

Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A. M., & Tipton, E. (2016). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness,* , 1-31. doi:10.1080/19345747.2016.1147628

Timperley, H. (2008). *Teacher professional learning and development.* UNESCO. International Academy of Education. International Bureau of Education.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development : Best evidence synthesis iteration (BES).* (). Wellington, N.Z.: Ministry of Education.

van de Grift, W., van der Wal, M., & Torenbeek, M. (2011). Development in teaching skills. *Pedagogische Studien, 88*(6), 416-432.

van Driel, J. H., Meirink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education, 48*(2), 129-160. doi:10.1080/03057267.2012.738020

van Kuijk, M. F., Deunk, M. I., Bosker, R. J., & Ritzema, E. S. (2016). Goals, data use, and instruction: The effect of a teacher professional development program on reading achievement. *School Effectiveness and School Improvement, 27*(2), 135-156. doi:10.1080/09243453.2015.1026268

van Veen, K., Zwart, R. C., Meirink, J. A., & Verloop, N. (2010). *Professionele ontwikkeling van leraren; een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren.* (). Leiden: ICLON/Expertisecentrum Leren van Docenten.

Van Waes, S., Moolenaar, N. M., Daly, A. J., Heldens, H. H. P. F., Donche, V., Van Petegem, P., & Van den Bossche, P. (2016). The networked instructor: The quality of networks in different stages of professional development. *Teaching and Teacher Education, 59*, 295-308. doi:http://dx.doi.org/10.1016/j.tate.2016.05.022

Wang, J., Odell, S. J., & Schwille, S. A. (2008). Effects of teacher induction on beginning teachers' teaching : A critical review of the literature. *Journal of Teacher Education, 59*(2), 132-152.

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology, 103*(2), 455-469. doi:10.1037/a0023067

Watson, J., & Beswick, K. (2011). School pupil change associated with a continuing professional development programme for teachers. *Journal of Education for Teaching, 37*(1), 63-75. doi:10.1080/02607476.2011.538273

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37*(8), 469-479. doi:10.3102/0013189X08327154

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research, 79*(2), 702-739.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the united states and abroad; technical report.* (). Dallas: National Staff Development Council.

Wlodkowski, R. J. (2003). Fostering motivation in professional development programs. *New Directions for Adult and Continuing Education, 2003*(98), 39-48. doi:10.1002/ace.98

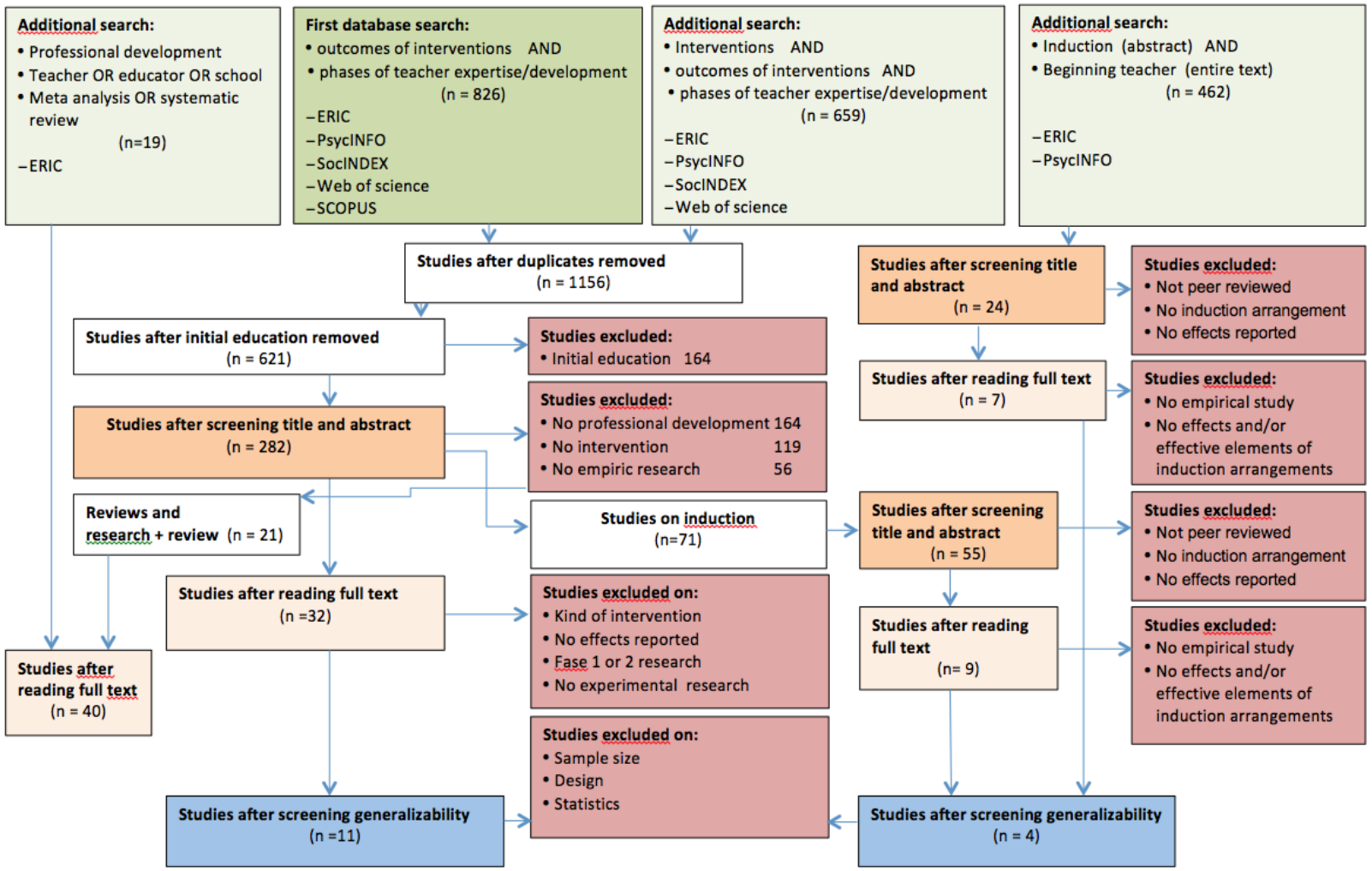
Wood, A. L., & Stanulis, R. N. (2009). Quality teacher induction: "Fourth-wave" (1997-2006) induction programs. *New Educator, 5*(1), 1-23.

Yoon, K. S., Duncan, T., Lee, S. W. -., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement.* ( No. Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Yoon, K. S., Duncan, T., Lee, S., & Shapley, K. (2008). The effects of teachers' professional development on student achievement: Findings from a systematic review of evidence. *American Educational Research Association Annual Meeting.*

Zwart, R. C., Smit, B., & Admiraal, W. F. (2015). Docentonderzoek nader bekeken: Een reviewstudie naar de aard en betekenis van onderzoek door docenten. = A closer look at teacher research: A review study into the nature and value of research conducted by teachers. *Pedagogische Studiën, 92*(2), 131-148.

# Appendix 1: Literature search diagram



# Appendix 2: Search strings and databases

In the first round the following databases were used:

- ERIC (with EbscoHOST)
- PsycINFO (with EbscoHOST)
- SocINDEX (with EbscoHOST)
- Web of Science
- SCOPUS

In the second round SCOPUS was no longer available.

The searches:

| | **Electronic database** | **Search terms** | **limiters** | **result** |
|---|---|---|---|---|
| 1 | ERIC<br><br>(EbscoHOST) | **SU** ("Beginning teacher *" OR "New teacher*" OR "Novice*" OR "student teacher*" OR "pre$service teacher*" OR "advanced beginner*" OR "competent teacher*" OR "Expert teacher*" OR "experienced teacher*" OR "master teacher*" OR "Early career teacher*" OR "teacher professional" OR "Developmental Stage*" OR "Developmental Continuity" OR "Career ladder*" OR "Career development" OR "experience level" OR "teacher developmental stage* " OR "professional teacher development" OR "quali* N/3 teach*" OR "teacher professional development")<br><br>**AND**<br>**SU** ("Classroom practice*" OR "teacher* N/3 practice*" OR "teach* performance*" OR "Classroom management" OR "Educational practice*" OR "Classroom technique*" OR "Teach* skill*" OR "Teacher behavior$r" OR "teacher improvement" OR "instructional improvement*" OR "instructional development*")<br><br>**AND**<br>Research | **Search mode:**<br>Boolean Phrase<br><br>**Date Published:**<br>20060101-20160131<br><br>**Educational Level**:<br>Adult Education, Elementary Education, Elementary Secondary Education, Grade 1, Grade 2, Grade 3, Grade 4, Grade 5, Grade 6, Grade 7, Grade 8, Grade 9, Grade 10, Grade 11, Grade 12, High School Equivalency Programs, High Schools, Intermediate Grades, Junior High Schools, Middle Schools, Postsecondary Education, Primary Education, Secondary Education, Two Year Colleges<br><br>**Publication Type**: Book/Product Reviews, Books, Collected Works (All), Dissertations/Theses - Doctoral Dissertations, ERIC Digests in Full Text, ERIC Publications, Guides - Classroom - Teacher, Journal Articles, Reports (All), Speeches/Meeting Papers | 382 |
| 2 | PsycINFO<br><br>(EbscoHOST) | **SU** ("Beginning teacher *" OR "New teacher*" OR "Novice*" OR "student teacher*" OR "pre$service teacher*" OR "advanced beginner*" OR "competent teacher*" OR "Expert teacher*" OR "experienced teacher*" OR "master teacher*" OR "Early career teacher*" OR "teacher professional" OR "Developmental Stage*" OR "Developmental Continuity" OR "Career ladder*" OR "Career development" OR "experience level" OR "teacher developmental stage* " OR "professional teacher development" OR "quali* N/3 teach*" OR "teacher professional development")<br>**AND**<br>**SU** ("Classroom practice*" OR "teacher* N/3 practice*" OR "teach* performance*" OR "Classroom management" OR "Educational | **Date Published:**<br>20060101-20151231<br><br>**Search mode:**<br>Boolean Phrase<br><br>**Expanders:**<br>Apply equivalent subjects | 107 |

| | | | | |
|---|---|---|---|---|
| | | practice*” OR “Classroom technique*” OR “Teach* skill*” OR “Teacher behavior$r” OR “teacher improvement” OR “instructional improvement*” OR “instructional development*”) | | |
| 3 | SocINDEX<br><br>(EbscoHOST) | SU (“Beginning teacher *” OR “New teacher*” OR “Novice*” OR “student teacher*” OR “pre$service teacher*” OR “advanced beginner*” OR “competent teacher*” OR “Expert teacher*” OR “experienced teacher*” OR “master teacher*” OR “Early career teacher*” OR “teacher professional” OR “Developmental Stage*” OR “Developmental Continuity” OR “Career ladder*” OR “Career development” OR “experience level” OR “teacher developmental stage* ” OR “professional teacher development” OR “quali* N/3 teach*” OR “teacher professional development”)<br>**AND**<br>SU (“Classroom practice*” OR “teacher* N/3 practice*” OR “teach* performance*” OR “Classroom management” OR “Educational practice*” OR “Classroom technique*” OR “Teach* skill*” OR “Teacher behavior$r” OR “teacher improvement” OR “instructional improvement*” OR “instructional development*”) | **Date of publication:**<br>20060101-20151231<br><br>**Search mode:**<br>Boolean Phrase<br><br>**Expanders:**<br>Apply equivalent subjects | 5 |
| 4 | Web of Science | **TOPIC**: (“Beginning teacher *” OR “New teacher*” OR “Novice*” OR “student teacher*” OR “pre$service teacher*” OR “advanced beginner*” OR “competent teacher*” OR “Expert teacher” OR “experienced teacher*” OR “master teacher*” OR “Early career teacher*” OR “teacher professional” OR “Developmental Stage*” OR “Developmental Continuity” OR “Career ladder*” OR “Career development” OR “experience level” OR “teacher developmental stage* ” OR “professional teacher development” OR “quali* N/3 teach*”OR “teacher professional development”)<br><br>**AND**<br>**TOPIC**: (“Classroom practice*” OR “teacher* N/3 practice*” OR “teach* performance*” OR “Classroom management” OR “Educational practice*” OR “Classroom technique*” OR “Teach* skill*” OR “Teacher behavior$r” OR “educational strategie*” OR “teacher improvement” OR “instructional improvement*” OR “instructional development*”) | **Timespan:**<br>2006-2015<br><br>**Indexes:**<br>SSCI<br><br>**Excluding:**<br>**languages:**<br>( Turkish OR Spanish OR Portuguese OR Japanese)<br>**Web of science categories:**<br>(literature romance OR information science OR science library science OR nursing OR health care sciences services OR emergency medicine OR dentistry oral surgery medicine OR public environmental occupational health OR public administration OR hospitality leisure sport tourism OR psychology clinical) | 330 |
| 5 | SCOPUS | **TITLE-ABS-KEY** ("Beginning teacher OR "New teacher*" OR "Novice*" OR "student teacher*" OR "pre$service teacher*" OR "advanced beginner*" OR "competent teacher*" OR "Expert teacher" OR "experienced teacher*" OR "master teacher*" OR "Early career teacher*")<br><br>**AND**<br>**TITLE-ABS-KEY**("teacher professional" OR "Developmental Stage*" OR "Developmental Continuity" OR "Career ladder*" OR "Career development" OR "experience level" OR "teacher developmental stage* " OR "professional teacher development" OR "quali* N/3 teach*") | AND PUBYEAR > 2005<br><br>AND SUBJAREA<br>(mult OR arts OR busi OR deci OR econ OR psyc OR soci) | 36 |

| | | | | |
|---|---|---|---|---|
| | | **AND**<br>**TITLE-ABS-KEY**("Classroom practice*" OR "teacher* N/3 practice*" OR "teach* performance*" OR "Classroom management" OR "Educational practice*" OR "Classroom technique*" OR "Teach* skill*" OR "Teacher behavior$r" OR "educational strategie*" OR "teacher improvement")<br><br>without ("instructional improvement*" OR "instructional development) did not fit in the | | |
| 6 | ERIC, PsycINFO, SocINDEX<br><br>(EbscoHOST) | **SU** ("teacher education" OR "Communities of Practice" OR "Teacher Education Programs" OR "microteaching" OR "practicum*" OR "Mentor*" OR "Teacher Induction" OR "College School Cooperation" OR "Preservice Teacher Education" OR "professional communities" OR "Professional Networking" OR "practice teach*" OR "Teacher Educator*" OR "Teacher Certification" OR "Teacher Evaluation" OR "Teacher Researcher*" OR "Preservice Teacher Education" OR "Curriculum Design" OR "Teacher Evaluation" OR "Teach* internship" OR "Peer coaching" OR "supervision" OR "Teacher Certification" OR "Teacher retention" OR "Instructional Design" OR "Professional Development Schools" OR "Collaboration" OR "Co-regulation" OR "Community of inquiry" OR "Teacher Collaboration" OR "induction program*" OR "Qualification program*" OR "teacher observation*" OR "developmental workshop*" OR "collaborative curriculum design" OR "collaborative lesson design" OR "lesson study" OR "workplace learning" OR "network of teach*")<br><br>**AND**<br><br>**SU** ("Beginning teacher *" OR "New teacher*" OR "Novice*" OR "student teacher*" OR "pre$service teacher*" OR "advanced beginner*" OR "competent teacher*" OR "Expert teacher*" OR "experienced teacher*" OR "master teacher*" OR "Early career teacher*" OR "teacher professional" OR "Developmental Stage*" OR "Developmental Continuity" OR "Career ladder*" OR "Career development" OR "experience level" OR "teacher developmental stage* " OR "professional teacher development" OR "quali* N/3 teach*" OR "teacher professional development")<br><br>**AND**<br>**SU** ("Classroom practice*" OR "teacher* N/3 practice*" OR "teach* performance*" OR "Classroom management" OR "Educational practice*" OR "Classroom technique*" OR "Teach* skill*" OR "Teacher behavior$r" OR "teacher improvement" OR "instructional improvement*" OR "instructional development*") | **Peer Reviewed**<br><br>**Date Published:**<br>20060101-20160231<br><br>**Expanders –**<br>Apply equivalent subjects<br><br>**Search modes - Boolean/Phrase** | 464 |
| 7 | Web of Science | **TOPIC**<br>("teacher education" OR "Communities of Practice" OR "Teacher Education Programs" OR | **From:**<br>Web of Science Core Collection | 182 |

| | | | | |
|---|---|---|---|---|
| | | "microteaching" OR "practicum*" OR "Mentor*" OR "Teacher Induction" OR "College School Cooperation" OR "Preservice Teacher Education" OR "professional communities" OR "Professional Networking" OR "practice teach*" OR "Teacher Educator*" OR "Teacher Certification" OR "Teacher Evaluation" OR "Teacher Researcher*" OR "Preservice Teacher Education" OR "Curriculum Design" OR "Teacher Evaluation" OR "Teach* internship" OR "Peer coaching" OR "supervision" OR "Teacher Certification" OR "Teacher retention" OR "Instructional Design" OR "Professional Development Schools" OR "Collaboration" OR "Co-regulation" OR "Community of inquiry" OR "Teacher Collaboration" OR "induction program*" OR "Qualification program*" OR "teacher observation*" OR "developmental workshop*" OR "collaborative curriculum design" OR "collaborative lesson design" OR "lesson study" OR "workplace learning" OR "network of teach*")<br><br>**AND**<br><br>**TOPIC** ("Beginning teacher *" OR "New teacher*" OR "Novice*" OR "student teacher*" OR "pre$service teacher*" OR "advanced beginner*" OR "competent teacher*" OR "Expert teacher*" OR "experienced teacher*" OR "master teacher*" OR "Early career teacher*" OR "teacher professional" OR "Developmental Stage*" OR "Developmental Continuity" OR "Career ladder*" OR "Career development" OR "experience level" OR "teacher developmental stage* " OR "professional teacher development" OR "quali* N/3 teach*" OR "teacher professional development")<br><br>**AND**<br>**SU** ("Classroom practice*" OR "teacher* N/3 practice*" OR "teach* performance*" OR "Classroom management" OR "Educational practice*" OR "Classroom technique*" OR "Teach* skill*" OR "Teacher behavior$r" OR "teacher improvement" OR "instructional improvement*" OR "instructional development*") | **Timespan:**<br>2006-2016<br><br>**Indexes:**<br>SSCI<br><br>**WEB OF SCIENCE CATEGORIES:**<br>Education educational research | |

# Appendix 3: Review summaries regarding the support of ECT

Lopez, Lash, Schaffner, Shields, and Wagner (2004) of the SRI, which is part of the Centre for Education Policy Reports conducted a thorough review of the empirical evidence on the effectiveness of induction programs in 2004. The review studied whether induction programs affect teacher quality (particularly in terms of student achievement) and which components of induction programs are the most promising for teacher quality (again, particularly in terms of student achievement). Twelve experimental and quasi-experimental works were selected for their review. Four studies reported a positive relationship between participation in an induction program/activity and early career teacher effectiveness in terms of student achievement gains or otherwise (one of which lacked significant tests), four studies indicated mixed results and two found no impact. There was no conclusive evidence overall about the impact of mentoring. None of the studies reviewed included effect sizes. Limits to the generalizability of the findings include problematic definitions of early career teachers participants, lack of information about the methods used to select early career teachers for the study, and insufficient descriptions of the participants and on how well the study participants represented the population of interest. Taking all limitations together, the authors conclude that although research includes some positive findings, the reviewed studies are not strong enough to conclude that induction works in terms of improving teacher retention or teacher effectiveness. The lack of high-quality (quasi-)experimental research precludes pinpointing the most effective induction components.

Wang, Odell, and Schwille (2008) conducted a critical review of the literature from 1997 till 2003 to explore the effects of teacher induction on early career teachers' conceptions and practice of teaching. Their literature search was conducted in ERIC with "teacher induction" and "early career teacher" and "teacher mentoring" as keywords. They used three approaches to understand effects found in the literature: 1. addressing the assumed effects of teacher induction components on ECTs' teaching using theoretical assumptions as a base, 2. analyzing the effects through teachers' self-reports, and 3. exploring the effects of using multiple data sources. The authors define 3 major components of induction programs: 1. Teacher mentoring relationships, as being the major supporting structure for early career teachers 2. Different forms of collaboration amongst early career teachers and colleagues; and 3. PD-activities designed to affect teaching and student achievement (Moir & Gless, 2001). They conclude that although teacher induction affects early career teacher's ideas about teaching, few studies capture effects on teaching practice and student achievement. Not one of the reviewed studies identified any actual effects of mentoring behaviours and processes on what early career teachers think and do in their classrooms or on effects on student learning. Furthermore, they identified that the quality of conceptions (and the match) of teaching and learning to teach that early career teachers and their mentors hold, plays an important role in the effects of induction on early career teaching practice. They also found that the quality of pre-service teacher education can have lasting influences on what and how early career teachers learn to teach.

Ingersoll & Strong (2011) conducted a critical review of the literature since the mid-1980s to provide a reliable and current assessment of what is known and not known about the effectiveness of ECT induction and mentoring programs. Their literature search began by

contacting leading researchers in the field and analysts in state governmental agencies in the United States. They examined existing systematic, narrative, or traditional reviews of such research, and searched several online databases, using a combination of three keywords: early career teacher induction, mentoring programs, and teacher mentors in combination with program evaluation, teacher improvement, effectiveness, retention, student achievement, and teaching practice. They only included (un)published empirical studies that met three criteria: 1) studies that sought to evaluate the effects of induction using one or more outcomes, 2) studies of induction that compared outcome data from both participants and nonparticipants in particular induction components, activities, or programs, and 3) studies that contained explicit descriptions of their data sources, sample sizes, research methods, and outcomes. Despite the fact that all of the studies reviewed have limitations and weaknesses of some kind (e.g. no comparison of ECTs who participated in induction to teachers who did not participate), the authors conclude that most of the studies reviewed provide empirical support for the claim that induction for ECTs and teacher mentoring programs in particular have a positive impact on three sets of outcomes: teacher commitment and retention, teacher classroom instructional practices, and student achievement. Most of the studies reviewed showed that ECTs who participated in some kind of induction performed better at various aspects of teaching, such as keeping students on task, developing workable lesson plans, using effective student questioning practices, adjusting classroom activities to meet students' interests, maintaining a positive classroom atmosphere, and demonstrating successful classroom management. Almost all of the studies on student achievement show that students of ECTs who participated in induction have higher scores, or gains, on academic achievement tests. There were exceptions to this overall pattern, in particular the study of (Glazerman et al., 2010). Some of the evidence suggests that the quantity of induction is important; programs that are more comprehensive, or longer, or include more depth of support appear to be better. The authors notice that much of the existing empirical research on the effects of induction lacks theory. Studies examine what works, but not why or why not.

Long et al. (2012) reviewed literature on the relationship between induction and mentoring and early career attrition and retention. The authors searched several databases using the terms: beginn*, novice, new, early career, newly qualified teacher, NQT, teacher, mentor, socialization, socialisation and intern. They limited their search to articles that were: empirical studies, written in English, in refereed publications, published from years 2000 – 2011, from Canada, the US, New Zealand, Australia, and the UK, and included material linked to teacher attrition or retention. The literature was organized using six of the Wood and Stanulis (2009) criteria of quality induction programs: educated mentors, reflective inquiry and teaching processes, systemic and structured observations, formative teacher assessment, administrators' involvement, and school culture supports. The authors conclude that the effect of induction (including mentoring) programs is unclear because of the multiple factors that influence a teachers decision to stay in or leave the profession. Studies were found showing the quality of teaching may be impacted with induction (including mentoring) but links to retention were often not made or were tenuous. School culture and context seems to be of significance. School cultures which are highly collaborative, value all teachers' knowledge, which focus on what is most educative for students, and which see students as the responsibility of the whole school, appeared most successful in retaining ECTs. Principals seem

to have a very important role in the success of induction programs, setting a tone for collegiality amongst all staff. School cultures supportive of an integrated approach rather than those oriented toward supporting veteran or ECTs were most successful in retaining ECTs. Only two studies were found that focused on systemic and structured observations, these studies provoke questions around the common assumption in mentoring and induction programs that observation is always valuable. Based on the limited research the authors found, they conclude that there appears to be a need for more exploration of the links between formative teacher assessment and induction programs. The authors note that several lines of research have focused on the lives of ECTs themselves and that there is an interest in ECTs' developing identities as teachers. Several researchers suggest it is vital that ECTs' voices are heard in designing induction programs.

Bullough (2012) reviewed an important aspect of induction: mentoring and tutoring (narrow interventions possibly embedded in broader interventions). He describes mentoring practices across four USA. He concludes that there is a great variation in the kind and quality of induction offered to early career teachers across the USA. They distinguish 3 levels: (a). *Basic induction,* having a mentor and supportive communication with school administrators; (b) basic induction plus collaboration, where the early career teacher reported having a mentor in their own field and regular and supportive communication with administrators or department chairs, a common planning period or regularly scheduled collaboration with other teachers in their area, and participating in a ECT seminar; and (c) all the above plus participating in an external teacher network and receiving extra resources, -reduced instructional load, fewer preparations, and a classroom aid. Helman (2006) identified three mentoring stances: efforts designed to extend ECT thinking, teaching specific content and practices, and promoting accountability by clarifying expectations of teaching and learning. Hobson, Ashby, Malderez, and Tomlinson (2009) concluded their review with that "little is known about the cost-effectiveness of mentoring, how to increase 'mentees willingness to be mentored ', how to impact on learning outcomes, how to increase mentor retention, how to influence mentors of various types of training" (p. 203). Feiman-Nemser (2001) warned that mentoring may reinforce ineffective practice: "sometimes {mentoring} reinforced traditional norms and practices rather that promoting more powerful teaching" (p. 1031). Reiman and Thies-Sprinthall, (1998) identified 5 conditions for adult learning: roletaking (not role playing), reflection, balance (between roletaking and reflection), continuity (paced practice, and continuous reflection), support and challenge. Bullough (2012) concludes that much of the studies of mentoring are under-theorized. He states that the work of Reiman (1999) is a welcome exception to the rule, being 'provocative and surprisingly underappreciated'. Reiman offers 'adult learning theory-based practical guidelines' for mentoring. He presents theoretical tenets of Piaget and Vygotsky for the theoretical underpinning of a 'taxonomy for reflection' to be used in differentiated mentoring of ECTs. Further gaps in the mentoring literature are: how mentoring affects mentors (Achinstein, 2006).

# Appendix 4: Article summaries with regard to ECT studies with less strong research designs

A number of studies were found that focus on specific aspects of interventions for early career teachers. Stanulis et al. (2012) were inspired to investigate the "zero effect on classroom practices" found by Glazerman et al. (2010). They hypothesized that targeted mentoring, in contrast to generic mentoring, might lead to the desired effects. In a quasi-experimental pre-post-test design they collected data of 42 elementary teachers of high-poverty schools with students with low achievement scores and high teacher attrition rates. The targeted mentoring intervention was executed by full-release mentors (not school-based), who were experienced teachers, forming a learning community, and was aimed to support teachers to lead high-level discussions that promote critical thinking. The intervention entailed intensive mentor support (no other intervention) consisting of co-planning, co-teaching, observing, debriefing, mentoring on the move, videotaping teaching, modelling, assisting the selection of teaching materials, working on school culture and using student work. The effect measure was the observation of ECT practices, more specifically their ability to facilitate high-level discussions that promote critical thinking. Rubrics were designed for the observation with 2 dimensions: *accountable talk* with 6 rubrics (the teachers' guiding complex thinking, relating students ideas to each other, pressing for accuracy and reasoning, building on prior knowledge, pressing for evidence to support ideas, prompting and modelling the creation of meaning from text) and *academic rigor* with 2 rubrics (actively engaging students, analysing and evaluating challenging content), rigor of the discussion, rigor of the text and rigor of the lesson activity. In the control condition novices relied on school-based full-time teacher mentors. These control schools with 40 ECTs had no criteria for selecting mentors. Comparative results (pre-post-test) reveal that the treatment group outperformed the control group on accountable talk and academic rigor with effect sizes: .30 to .50.

As a follow-up of the Stanulis et al. (2012) study, Stanulis, Brondyk, Little and Wibbens (2014) investigated the practice of one mentor who seemed to be effective in shifting elementary ECTs teaching practice to include discussion-based teaching. The mentor participated in an intensive two-year university-led professional development intervention to prepare her to help ECTs learn to lead classroom discussions for high-level thinking and incorporating the idea of academic rigor. Support for the mentor provided by the university included monthly study groups, monthly coaching, frequent email conversations, and e-mail communications and agenda exchange to help mentors as teacher leaders as they led monthly ECT study groups in their school. All three ECTs selected for this study were African-American, in their early twenties and graduated from a four-year teacher education program. They each made gains in their teaching effectiveness on a pre-post observational instrument (Stanulis et al., 2012). The authors attempted to increase credibility and ensure construct and internal validity of their longitudinal descriptive case study by utilizing multiple sources of data, including interviews, direct observations, and documentation from both mentor and ECTs and engaging in an iterative process of explanation building. A descriptive coding process was used to organize the data across all participants to characterize how the ECTs and mentor approached their practice. Having a targeted practice as a mentor, taking a stance as teacher leader, and seeking support to become an effective mentor characterized the studied mentor. She met each ECT where they were in their development, and took on the authority to move each of them

forward as they embraced features of discussion based teaching. She provided an image of an exemplar and of the possible, instead of facilitating learning ordinary practices, helping ECTS learn the power of local knowledge from teachers who took on teacher educator roles, who pushed back against institutionalized norms of learning to teach alone or learning to teach the scripted curriculum.

Luft et al. (2011) provided support for the the importance of pedagogical content knowledge (PCK) when supporting instructional practice of science teachers. Using an embedded concurrent mixed methods design they studied the changes in beliefs, pedagogical content knowledge (PCK), and instructional practices of early career secondary science teachers during their first 2 years in the classroom as they participated in one of four different induction programs. Two programs focused on teaching in general, two others were science specific. An abbreviated overview of the programs is provided in the article. After purposeful sampling to select ECTs (n=98) they were interviewed and observed several times during the schoolyear. Triangulation of data and potential threats to the validity of the design were addressed. Using SPSS all of the collected beliefs and PCK data were analyzed comparatively with an analysis of variance and the least significant difference (LSD) follow-up test. Qualitative data were coded. Analysis revealed that first year teachers who participated in science specific induction programs strengthened instructional practices. They continued to enact more interactive learning environments that had more investigations and laboratories than did their peers in the general induction programs. ECTs were consistent across groups in the instructional methodologies they used in their classrooms and by the end of the second year, regardless of program, most of the teachers shared similar beliefs and PCK. However, over 2 years, the ECTs in science specific induction programs created more interactive classrooms and enacted more investigations and laboratories than their peers in the general induction programs. Beliefs were more likely to be impacted by induction programs than by working with students. Analysis also revealed mentors and colleagues played important, but different roles over time in socializing ECTs into the school and science education community. Initially ECTs relied on those closest to their classroom and who could offer some form of assistance related to instruction, working within the school, or just an emphatic ear. In the second year, ECTs became more purposefully engaged with different groups of educators in order to acquire knowledge or to find new ways of teaching science. ECTs participating in science specific induction programs were more likely to move from opportunistic support to strategically selected forms of support that were focused on teaching science over time. ECTs knew what forms of support would help build their capacity as science teachers, and they sought them out in their second year. ECTs benefited the most from the combination of a science mentor and an ongoing emphasis of teaching science. When ECTs were consistently supported to think about teaching science, they were more purposeful as they sought support and as a result were more satisfied.

Haigh and Anthony (2012) explored how reported efficacy of a cohort of early career secondary science teachers (N = 20) in New Zealand changed over the first 18 months of their teaching career and how the nature of their induction program and the specific science teaching context influenced their teacher self-efficacy. New Zealand has a mandated and government-funded two-year induction program for all secondary provisionally registered

teachers (PRTs), with three essential features; educative mentoring, professional learning opportunities, and formative and summative evaluations of professional practice. Using data from questionnaires and interviews (6, 12 and 18 months into teaching) efficacy scores were analyzed using both ANOVA and independent means t test procedures. Reliability of analysis was sought through inter-researcher discussion, and content- and face validity of the used scales were checked. Scrutiny of interview transcripts and efficacy profiles led the authors to select transcripts from three of the participants for further detailed analysis in order to develop biographical cases representing different efficacy profiles and different contextual features. Results show that while the ECTs overall reported relatively constant ratings of self-efficacy, they demonstrated different patterns of declared efficacy across the 18-month period. Interview transcripts revealed stories of ECTs who were strongly focused on making a difference to their students' learning. The three case teachers all gained knowledge of their subject matter, became surer of their pedagogy and classroom management strategies, and developed a greater understanding of their schools' environment to very different degrees and consequently different levels of efficacy. Reflection on their approach to and effectiveness of behaviour management appeared to have some impact on their overall efficacy beliefs. Visiting other teachers' classrooms resulted in an increase in self-efficacy if observation was followed by critical reflection and successful try-outs of new approaches. The three cases appreciated encouraging comments made by administrators, teachers and students and indicated these comments boosted their confidence as teachers. Formal and informal feedback from school leaders, teachers, students and parents positively influenced efficacy. The findings from this study concerning the link between contextual factors and ECTs' self-efficacy beliefs are mixed. The teachers illustrated various ways in which the school contexts might affect the teachers' professional lives and therefore their efficacy beliefs, but they also provided examples where contextual factors appeared to have little influence. School leadership, school management systems and the level of formal professional support influenced the ECTs' reported efficacy differently. Based on the authors consideration of the findings of this study and their reading of academic and professional literature in the area they present a model for induction to support professional learning that takes into account the individual needs of the ECT and is targeted to build efficacy. It suggests how induction practices can influence identified sources of efficacy with recognition of the ECT's professional strengths and identifies the school's socio-political context as the learning environment. Similar results were found by Simon (2011).

Fresko and Nasser-Abu Alhija (2015) explored the perceived value by ECTs of induction seminars, within which ECTs are expected to share their experiences with others in a non-threatening and supportive environment, in the context of the Israeli induction program. The three main components of the induction program were in-school mentoring, seminars, and evaluation/assessments of ECTs. A mixed method approach including questionnaires, semi-structured interviews, and observations of seminar meetings was used and seminars and participants (378 elementary and secondary ECTs and 29 seminar leaders) were randomly selected. For the qualitative component five teacher training institutions were purposively selected. Results show that ECTs responses were highly diverse (reflected in relatively large standard deviation values) and the contribution of the seminars was rated, at best, as only moderate. ECTs reported feelings of satisfaction from being able to share with each other.

Emotional support could be sought and provided in a non-threatening environment. Participants recommended bringing experts to lecture on specific topics of interest to everyone.

*Bringing early career intervention studies together*

From our review of studies with a broad intervention perspective, we have learned that induction in general has positive effects on teaching practice, self-efficacy and retention. It has also been revealed that features of induction arrangements have different impact weights on certain outcome measures (Helms-Lorenz et al., 2015; Ingersoll et al., 2004; Kang et al., 2012). From the focussed studies it becomes clear that the quality of individual features of induction can be improved. This implies that the general impact of induction arrangements will improve when mentors support ECTs more effectively (Stanulis et al., 2014). Being supported by a science teacher mentor boasts the instruction practice of ECTs teaching science (Luft et al., 2011). Induction arrangements can therefore be improved by introducing a stronger PCK component in ECT mentoring. Reflection on their approach to and effectiveness of behavior management appeared to have some impact on their overall efficacy beliefs. Visiting other teachers' classrooms resulted in an increase in self-efficacy if observation was followed by critical reflection and successful try-outs of new approaches. Colleagues should be aware that ECTs appreciate encouraging comments made by administrators, teachers and students and these comments boost their confidence as teachers (Haigh and Anthony (2012). The quality of seminars can be improved by introducing more expertise (Fresko & Nasser-Abu Alhija, 2015). Finally, targeted mentoring can increase mentoring effects (Stanulis et al., 2012). Based on this study having a focused, theoretically based mentoring approach seems plausible and feasible.