Vetenskapsrådet

Delrapport från SKOLFORSK-projektet

KARTLÄGGNING AV FORSKNING OM FORMATIV BEDÖMNING, KLASSRUMSUNDERVISNING OCH LÄROMEDEL I MATEMATIK

VETENSKAPSRÅDETS RAPPORTER 2015

KARTLÄGGNING AV FORSKNING OM FORMATIV BEDÖMNING, KLASSRUMSUNDERVISNING OCH LÄROMEDEL I MATEMATIK

VETENSKAPSRÅDET
Box 1035
SE-101 38 Stockholm, SWEDEN

ISBN 978-91-7307-274-8

Vetenskapsrådet genomförde under 2014 ett projekt, SKOLFORSK, för att kartlägga befintlig utbildningsvetenskaplig forskning. Arbetet skedde på uppdrag av regeringen för att resultera i kartläggningar av svenska och internationella forskningsresultat med relevans för skolväsendet. Syftet var att skapa en plattform av kunskapsunderlag till det nybildade Skolforskningsinstitutet. Slutsatserna i denna delrapport är författarnas egna. Vetenskapsrådets sammanfattande rapport, Forskning och skola i samverkan, med en beskrivning av projektet och med de frågeställningar, resultat och rekommendationer som redovisats inom delprojekten kan liksom de övriga delrapporterna laddas ner från Vetenskapsrådets webbplats.

KARTLÄGGNING AV FORSKNING OM FORMATIV BEDÖMNING, KLASSRUMSUNDERVISNING OCH LÄROMEDEL I MATEMATIK

Andreas Ryve, Per Nilsson, Torulf Palm, Hendrik Van Steenbrugge, Catarina Andersson, Andreas Bergwall, Erika Boström, Maria Larsson och Lotta Vingsle

FÖRORD

Regeringen gav 2013-11-21 (U2013/6845/S) Vetenskapsrådet i uppdrag att svara för genomförandet av validerade kartläggningar av svenska och internationella forskningsresultat med relevans för skolväsendet. Kartläggningarna skulle utgå ifrån frågeställningar som är relevanta för, och framtagna i samråd med, verksamma i skolan och förskolan. Syftet med kartläggningarna var att utgöra underlag för systematiska sammanställningar av forskningsresultat med relevans för verksamhet inom skola och förskola som Skolforskningsinstitutet skulle få i uppdrag att genomföra. Uppdraget formulerades efter att huvudsekreteraren för Utbildningsvetenskapliga kommittén (UVK) vid Vetenskapsrådet utformat ett förslag till ett antal projekt som under ett år skulle arbeta fram ett underlag till Skolforskningsinstitutet.

Uppdraget från regeringen, med arbetsnamnet SKOLFORSK, har trots den korta tid som stått till buds, resulterat i sexton delprojekt där ett 40-tal forskare från femton olika universitet i Sverige, Norge och USA har medverkat. En välmeriterad forskare med expertkunskaper inom respektive område har varit ansvarig ledare för de olika projekten. Delprojekten, som alla har genomförts under 2014, varierar i tidsomfång - från fyra till elva månader. De kortare studierna syftar till att underlätta den nya myndighetens initiala arbete avseende processer och modeller för kunskapsbildning, och till att skapa gynnsamma förutsättningar för användning av forskningsbaserad kunskap i skolan. De längre projekten är exempel på olika typer av systematiska sammanställningar av forskningsresultat. De visar på olika modeller och metoder för hur forskning avseende lärande i skolan kan systematiseras och synliggöras.

Huvudsekreteraren för UVK, professor Eva Björck samt projektledaren, fil.dr. Cristina Robertson har varit ansvariga för projektet. SKOLFORSK har haft en referensgrupp med olika aktörer som arbetar med att befrämja praktiknära forskning och spridning av forskning. Projektet har haft nära kontakt med den grupp som planerat Skolforskningsinstitutet.

Ett varmt tack riktas till alla forskare som med kort varsel gjort det möjligt att genomföra detta projekt. Ni har berikat skolväsendet och Skolforskningsinstitutet med en gedigen bas att utgå ifrån i fortsatt arbete med skolans vetenskapliga förankring och uppbyggnad av den praktiknära skolforskningen i Sverige till gagn för förskolor, skolor och lärarutbildning.

Skolforskningsinstitutet önskas framgång och lycka med sitt fortsatta arbete!

Petter Aaasen, ordförande, Utbildningsvetenskapliga kommittén
Eva Björck, huvudsekreterare för utbildningsvetenskap, Vetenskapsrådet

Denna rapport presenterar kartläggningen av forskning om formativ bedömning, klassrumsundervisning och läromedel i matematik. Projektet är genomfört som tre delprojekt fokuserat på vart och ett av de ovan nämnda områdena. Introduktionen till projekten är skriven på svenska liksom delprojektet om formativ bedömning. Delprojekten om klassrumsundervisning och läromedel är skrivna på engelska.
Huvudansvarig för projektet är Andreas Ryve med Per Nilsson som biträdande vetenskapligt ansvarig. Torulf Palm är ansvarig för delprojektet om formativ bedömning där också Catarina Andersson, Erika Boström och Lotto Vingsle ingår i projektteamet. Per Nilsson är ansvarig för delprojektet om klassrumsundervisning där också Maria Larsson ingår i projektteamet. Hendrik van Steenbrugge är ansvarig för delprojektet om läromedel där också Andreas Bergwall ingår i projektteamet.

INNEHÅLL

FÖRFATTARPRESENTATION

Andreas Ryve är professor i matematikdidaktik vid Mälardalens högskola och gästprofessor på Umeå universitet. Andreas är projektledare och vetenskaplig ledare för 16Mkr-projektet ”Räkna med Västerås” (tillsammans med Kirsti Hemmi) och för 18Mkr-projekt ”Theorizing teacher use of curriculum material within mathematics classroom practice” finansierat av Vetenskapsrådet. Båda projekten syftar till att teoretisera läromedel, utbildningar, rutiner och organisationer som stöttar lärare att agera effektivt i matematikklassrummet.

Per Nilsson är professor i matematikdidaktik vid Örebro universitet. Pers forskning riktar sig mot att karaktärisera och förstå uppbyggnaden av matematiska resonemang och hur undervisning kan utformas för att stödja elevers förmåga att följa och föra matematiska resonemang. Han ledare VR-projektet ”Att utveckla experimentbaserad undervisning och didaktik i sannolikhet: Inferentialism, kontextualisering och variation i samverkan”. Per är också vetenskaplig ledare för Matematiklyftets moduler på temat sannolikhet och statistik.

Torulf Palm är docent i Pedagogiskt arbete vid Umeå universitet. Hans forskningsfokus är formativ bedömning, och han bedriver även lärarutbildning och lärarfortbildning inom detta område.

Hendrik Van Steenbrugge är postdoc vid Mälardalens högskola. Hans huvudsakliga forskningsintresse rör samspelet mellan lärare och läromedel i grundskolan. Forskningen bedrivs delvis i en tvärkulturell kontext. Delar av den är inriktad på faktorer som påverkar vilken karaktär som utmanande matematikuppgifter får när de används i gymnasieskolan.

Catarina Andersson är doktorand i pedagogiskt arbete med inriktning matematikdidaktik vid Umeå universitet. Hennes avhandling bygger på fyra delstudier som genomförts inom ett större forskningsprojekt om fortbildning av lärare i formativ bedömning. Catarina ansvarade i detta forskningsprojekt för studierna av hur fortbildningen om formativ bedömning i matematik påverkade klassrumspraktiken och elevers resultat i matematik för gruppen mellanstadielärare som deltog i fortbildningen.

Andreas Bergwall är adjunkt i matematik vid Örebro universitet och har mångårig erfarenhet av undervisnings- och utvecklingsarbete inom såväl lärarutbildning som ingenjörs- och naturvetenskaplig utbildning. Andreas är också doktorand i matematikdidaktik vid Mälardalens högskola. Hans avhandling fokuserar läromedels behandling av matematiska bevis och bevisrelaterade resonemang. Han har deltagit i arbetet med Matematiklyftets problemlösningsmodul för årskurs 7-9.

Erika Boström är doktorand i matematikdidaktik vid Umeå universitet. Hennes forskningsintresse är formativ bedömning och kompetensutveckling av matematiklärare inom detta område. Med sin forskning hoppas hon kunna bidra med kunskap om lärares implementering av formativ bedömning i klassrummet och även kunskap om hur man kan stödja lärare att utveckla sin formativa bedömningspraktik. Erika har tidigare arbetat som gymnasielärare i matematik och naturkunskap och hon har också arbetat som provutvecklare av gymnasieskolans nationella prov i matematik.

Maria Larsson är doktorand och adjunkt i matematikdidaktik vid Mälardalens högskola. Hennes avhandling fokuserar på lärarens roll i att planera och hålla givande helklassdiskussioner utifrån elevers olika sätt att tänka kring ett utmanande matematiskt problem. Maria har tagit fram övergripande struktur samt innehåll till Matematiklyftets problemlösningsmodul för årskurs 7-9. Hon är också författare till boken 32 rika problem i matematik.

Lotta Vingsle är doktorand vid Umeå Universitet. Hennes avhandling bygger på en studie av en lärares formativa bedömningspraktik under elevaktiva genomgångar i matematik. I studien identifieras de aktiviteter som läraren genomför under de tre faserna i den formativa bedömningen – samla in, tolka och använda information om hur eleverna förstår undervisningsinnehållet. Lotta är också matematiklärare och matematikutvecklare.

SAMMANFATTNING

Regeringen har gett Vetenskapsrådet uppdraget att genomföra ”kartläggningar av svenska och internationella forskningsresultat med relevans för skolväsendet” (Utbildningsdepartementet, U2013/6845/S). Föreliggande rapport presenterar det projekt inom uppdraget som fokuserar matematik. Projektet inom matematik är indelat i tre delprojekt som berör formativ bedömning, klassrumsundervisning, och läromedel. Bevekelsegrunden för att välja de tre ovan nämnda områdena är att de är viktiga komponenter i att stödja lärare i att etablera möjligheter för elever att utveckla matematiska förmågor. Det övergripande syftet med denna rapport är således att presentera resultaten från kartläggningar av litteratur rörande formativ bedömning, klassrumsundervisning, och läromedel i matematik.

Den metodologiska designen av alla tre delprojekt inspirerades av tio processer för systematiska litteraturöversikter introducerat av Gough, Oliver, and Thomas (2013). Vi inriktar oss särskilt på studier av artiklar publicerade i tidskrifter tillhörande Web of Science (WoS).

Vilken effekt har formativ bedömning på elevers prestationer i matematik och vilka egenskaper hos formativ bedömning är viktiga för att åstadkomma dessa effekter? Resultaten från det stickprov av artiklar som analyserats visar att alla strategier för formativ bedömning har ett positivt samband med elevers prestationer i matematik. Effektstorlekarna har oftast varit medelstora eller stora. Däremot är det svårare utifrån de ingående studierna att dra slutsatser om effekten av olika mer specifika egenskaper hos någon av de olika strategierna för genomförande av formativ bedömning. Till exempel indikerar studierna, i likhet med forskning från andra områden än ämnet matematik, att feedback har potential att påverka elevernas prestationer men att denna effekt är beroende av vilken typ av feedback eleverna får och på vilket sätt de får denna feedback. En annan fråga som blir relevant till följd av de positiva sambanden mellan formativ bedömning och elevers prestationer i matematik är hur svensk matematikundervisning ser ut med avseende på formativ bedömning. Några få publikationer har studerat effekter av kompetensutvecklingsprogram i formativ bedömning och kan ge en viss fingervisning om denna aspekt i några skolor och kommuner. Men när det gäller frågan om hur svenska klassrumspraktiker ser ut avseende formativ bedömning hittades inte en enda studie i databasen Swepub med fokus på att analysera nuvarande klassrumsundervisning i matematik.

Genom kartläggningen av klassrumsundervisning i matematik blev det tydligt hur studieobjekten bygger på ett analytiskt samband mellan en designvariabel och en effektvariabel. Designvariablerna grupperas enligt *interactional strategies, teaching approaches, learning material* and *background variables*. Effektvariablerna beskriver mot vad en design är riktad och de sorterar under *mathematical products* and *processes, student attributes* and *the establishment of a certain classroom practice.* Trots att den matematikdidaktiska forskningen under senaste decenniet vidgat synen på lärande i matematik mot matematiska förmågor och kompetenser, så visar kartläggningen att forskningen fortfarande har sin tyngdpunkt på att studera och utvärdera effekterna av olika undervisningsdesigner i relation till elevers lärande av matematiska produkter, dvs. i relation till elevers färdigheter att reproducera och lära sig hantera matematiska konventioner, symbolsystem, begrepp och algoritmer. Utifrån djupläsning av de artiklar som faktiskt fokuserar matematikundervisning relaterad till elevers lärande av matematiska förmågor urskiljs tre huvudteman: 1) olika aspekter av problemlösningsundervisning, 2) jämförelse som en effektiv undervisningsstrategi, samt 3) vikten av att läraren ställer frågor på ett sådant sätt att elever förklarar sina lösningar tydligt och detaljrikt. Forskningsdesignen för många av de kompetensorienterade studierna följer strukturen förtest – intervention – eftertest med användande av kontrollgrupp. Studierna visar på samband och korrelationer mellan undervisning och elevers lärande i termer av matematiska kompetenser, men endast några få av studierna gör anspråk på orsakssamband. Resultat pekar åt att problemlösningsorienterad undervisning kan ge positiva resultat för elevers utveckling av sina matematiska förmågor. Dock är det av stor vikt att läraren etablerar produktiva normer och förväntningar för att undervisning baserad på problemlösning ska lyckas. Viktigt att ta i beaktande är att nyanser av hur problemlösningsundervisning genomförs är centrala för hur framgångsrik den blir.

Ett centralt resultat gällande läromedel är den komplexitet som råder för att förstå hur läromedel kan stödja lärare att etablera klassrumsundervisning. Forskning visar på komplexa samband mellan läromedel, lärarkompetenser av olika slag, faktorer som kompetensutveckling och lärares syn på undervisning i olika

kulturella kontexter. Forskning börjar kartlägga aspekter av så kallade ”Educative curriculum material” - läromedel som syftar till att utbilda lärare - där lärare ges möjligheter att utveckla kunskaper om övergripande matematiska idéer, barns lärande i matematik och undervisningsmetoder. Centralt och intressant är också att studier börjar indikera att lärares kompetens att använda läromedel effektivt och fokusera på viktiga matematiska idéer i förberedelse av lektioner är korrelerat till effektiv klassrumsundervisning. Mycket mer forskning behövs för att definiera variabler och säkerställa samband och vi finner det intressant för den svenska kontexten att ingående undersöka hur läromedel kan utvecklas i syfte att stödja lärare att designa undervisning som aktivt engagerar elever i viktiga matematiska idéer.

Konkreta begränsningar av projektet är att vi (1) valt bort potentiellt användbara databaser och nästan uteslutande fokuserat på WoS, (2) inte hunnit koda mer än en relativt liten del av de artiklar från WoS som vi bedömer som relevanta, (3) inte i någon större utsträckning kunnat avgöra användbarheten av resultaten för den svenska kontexten och (4) därför inte kunnat engagera oss i en syntetisering av forskningsresultaten. Det finns således många delar att arbete vidare med men vi menar att denna rapport kan fungera som ett viktigt underlag för ett sådant arbete. I förhållande till den svenska kontexten (Hemmi & Ryve, 2014; Boesen, Helenius, Bergqvist, Bergqvist, Lithner, Palm, & Palmblad, 2014), internationell forskning (Hattie, 2009; Smith & Stein, 2011) och föreliggande projekt rekommenderar vi Skolforskningsinstitutet att beakta två aspekter av stor vikt för att utveckla elevers kunskap i matematik. För det första bör Skolforskningsinstitutet söka och sammanställa forskning som stöttar lärare, speciallärare, rektorer samt andra aktörer i skolverksamheterna att agera i undervisningssituationer. Skolan behöver kännetecknas av fokuserade diskussioner där yrkesgrupper får stöd i att aktivt agera och utveckla sin profession. I fallet matematik behöver exempelvis lärare konkret stöd för hur de ska förbereda lektioner, förutse elevers sätt att tänka, använda läromedel effektivt, i formativt syfte bedöma elevers kunskap, introducera matematiska problem, agera under grupparbeten, leda helklassdiskussioner och skapa progression mellan lektioner och terminer. Material från Skolforskningsinstitutet ska konkret stötta aktörer att agera och tänka i skolpraktiken. För det andra rekommenderar vi Skolforskningsinstitutet att söka och sammanställa forskning som stöttar skolhuvudmän att långsiktigt institutionalisera utvecklingsinsatser av klassrumsundervisningen. Konkret uttryckt, rektorer och huvudmän ska få stöd från forskningen i att inte enbart initiera och implementera utvecklingsinsatser utan också säkerställa att dessa institutionaliseras och därigenom blir en permanent del av skolans kärnverksamhet. Skolforskningsinstitutet bör således identifiera rutiner, material, roller och arbetssätt som stärker långsiktigheten i skolans utvecklingsarbete.

SUMMARY

The current project focuses on mathematics education, and is partitioned into three subprojects mapping research on formative assessment, classroom teaching, and curriculum programs in mathematics. The rationale for focusing on these three areas is that they are all highly relevant for understanding and improving Swedish mathematics education and students' knowing of mathematics. Therefore, the aim of the project is to map research on formative assessment, classroom teaching, and curriculum programs in mathematics education.

The methodology of the literature review has been inspired by Gough, Oliver, and Thomas (2013), and we have focused on the mapping on journal articles published on Web of Science (WoS).

The results from the sample of articles on formative assessment show that strategies of formative assessment in mathematics are positively correlated to students' performance in mathematics with medium and large effect sizes. However, based on the current mapping it is difficult to specify aspects of how the formative strategies are to be implemented in order to promote students' knowing of mathematics.

Despite the change in perspective of what constitutes knowledge in mathematics to also include reasoning, problem-solving and communication, the map shows that research is mainly focused on examining teaching methods and their effects on students' skills in mathematics. A closer examination of the studies that do focus on teaching for supporting students in developing competencies like reasoning and problem-solving shows that connections between and comparison of students' solutions, as well as teachers' ways of asking questions to support students in explaining their solutions clearly and in detail, are important for students' learning of these competencies.

A central finding stemming from this review of curriculum programs is the complexity involved in how the programs can support teachers in establishing classroom practices. Curriculum resources and teacher resources, as well as other influencing factors, impact the quality of instruction, and studies have begun to point out how curriculum resources and teacher resources uniquely and jointly impact classroom practices. Multiple research articles have expressed the need for teacher support in implementing curriculum programs, by means of professional development, teacher education and support provided by the curriculum programs themselves. Interesting in this regard is the state of the research field concerning the design of educative curriculum programs, and how teachers make use of such support. Studies have proposed design approaches, regarding both the actual development of educative curriculum programs as well as how to use them in teacher education to support prospective teachers' development of knowledge. Further, although research has revealed that it is important to prepare for teaching in certain ways, we found very little research that explicitly analyzed how teachers actually prepare for teaching a mathematics lesson.

Limitations of the project include: (1) the lack of searching in potentially relevant databases, (2) the fact that a relatively small proportion of articles found in the search have been coded, (3) that we have not engaged in deep considerations as to whether and in what ways results from international research are relevant in the Swedish context, and (4) that we therefore have not been able to synthesize the results of the study. In relation to the Swedish context (Hemmi & Ryve, 2014; Boesen et al., 2014), international research (Hattie, 2009; Smith & Stein, 2011), and the current project's findings, we recommend that *Skolforskningsinstitutet* focus on two aspects of great importance for developing students' knowing of mathematics. First, *Skolforskningsinstitutet* should synthesize research that supports actors, such as teachers and principals, in acting within school practices. In the case of teachers, support is needed to engage them in actively anticipating students' thinking, using curriculum programs effectively, introducing mathematical content, acting in group work, formatively assessing students' learning, and orchestrating whole-class mathematical discussions. Secondly, actors within school practices need support not only in initiating and implementing developments but also in institutionalizing such developments. *Skolforskningsinstitutet* should specify the kind of support needed in order to ensure that material, routines, competences, and organizations become integral and permanent features of Swedish school practice.

1. INLEDNING TILL DE TRE DELPROJEKTEN

Introduktion

Regeringen har gett Vetenskapsrådet uppdraget att genomföra ”kartläggningar av svenska och internationella forskningsresultat med relevans för skolväsendet” (Utbildningsdepartementet, U2013/6845/S). Utbildningsdepartementet (U2013/6845/S) preciserar vidare att ”Syftet med kartläggningarna är att dessa ska utgöra underlag för de systematiska sammanställningar av forskningsresultat med relevans för verksamhet inom skola och förskola som en planerad ny myndighet som regeringen avser att inrätta kommer att ha i uppdrag att genomföra”. Vi vill utifrån detta uppdrag således betona att uppdraget avser *kartläggningar* som ska *ligga till grund* för systematiska sammanställningar och att föreliggande kartläggning ska underlätta Skolforskningsinstitutets arbete att producera något med *relevans* för skola och förskola.

Föreliggande rapport presenterar det projekt inom uppdraget som fokuserar matematik. Projektet inom matematik är i sig indelat i tre delprojekt som fokuserar fenomenen formativ bedömning, klassrumsundervisning, och läromedel. Initiala och övergripande preciseringar av fenomenen i detta projekt är att formativ bedömning syftar till “…encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged.” (Black & Wiliam, 1998a, p.2), klassrumsundervisning syftar på ”classroom interaction among teachers and students around content directed toward facilitating students’ achievement of learning goals” (Hiebert & Grouws, 2007, p. 372) och läromedel syftar på läroböcker med tillhörande lärarhandledningar och övrigt material kopplade till dessa (Stein, Remillard, & Smith, 2007). Dessa initiala konceptualiseringar ger en övergripande ram för projektet men ett syfte med projektet är att utifrån sökning, kodning och kartläggning av forskningen ytterligare karakterisera dessa fenomen, delar inom dessa och relation till elevers lärande i matematik.

Bevekelsegrunden för att välja de tre ovan nämnda områdena är att projektet ska utgöra ett underlag för Skolforskningsinstitutet att etablera insatser med relevans för skola och förskola. Mer konkret, utifrån projektgruppens omfattande kunskap om forskningsfält och skolpraktik har vi gjort bedömningen att formativ bedömning, klassrumsundervisning och läromedel är viktiga komponenter i att stödja lärare i att etablera möjligheter för elever att utveckla matematiska förmågor. Vi lägger speciell vikt vid hur forskning från dessa områden kan stötta lärare att agera i matematikklassrummet. Lärares sätt att agera i klassrummet menar vi är synnerligen intressant i den svenska kontexten eftersom studier indikerar att svenska lärare i matematik typiskt behöver utveckla sin kompetens och praktik i att introducera matematiska problem, sätta mål för lektioner, genomföra och använda klassrumsaktiviteter i formativt syfte, agera under grupparbeten samt i att leda helklassdiskussioner (jmf, Hemmi & Ryve, 2014; Boesen et al., 2014). Vi kan således konstatera att internationell forskning lyfter lärarens agerande i klassrummet som centralt (Hattie, 2009), att nationell forskning indikerar att svenska klassrumsundervisning behöver utvecklas (Boesen et al., 2014), och att olika forskningsområden börjar intressera sig allt mer för forskning om klassrumsundervisning (Hiebert & Grouws, 2007) och hur den inte enbart kan beskriva och förklarar klassrumsundervisning i matematik utan också hur den kan stödja lärare att agera för att utveckla elevers kunskaper i matematik (Smith & Stein, 2011).

Mot denna bakgrund är det övergripande syftet med denna rapport att presentera resultaten från kartläggningar av litteratur rörande formativ bedömning, klassrumsundervisning med ett speciellt fokus på lärarnas agerande och från forskning om läromedel. Mer precisa syften och forskningsfrågor presenteras inom varje delprojekt.

Disposition

Rapporten är uppbyggd så att mer generella aspekter av projektet presenteras i denna inledande del. Där behandlas fokus, vägval och argument som är av relevans för alla tre delprojekt. Ovan berördes kort övergripande logik till varför vi har valt inriktningar på projektet och att vi betonar forskning som kan stödja lärare att agera. Nedan redogör vi för projektets övergripande metodologiska struktur och vägval inom denna

struktur följt av projektets styrkor och begränsningar. Efter denna gemensamma inledning presenteras varje enskilt delprojekt. Vi presenterar varje delprojekt så att dessa kan läsas separat. Varje delprojekt har således en inledning, specificering av syfte och forskningsfrågor, metodologi, resultat och en avslutande diskussion.

Metod

Den metodologiska designen av alla tre delprojekt inspirerades av de tio processer för systematiska litteraturöversikter introducerat av Gough, Oliver, and Thomas (2013). Vårt uppdrag var inte att genomföra en fullskalig systematisk litteraturöversikt men ramverket är till stora delar användbara för att strukturera arbetsgången inom det aktuella projektet. Inom projektet engagerade vi oss i processerna 1-7 samt i process 10. Gough et al. (2013) presenterar följande ramverk:

- Behov (eng. Need): Vem ställer frågorna och vad ska svaren användas till?
- Översiktens frågor (eng. Review questions): Vad avser översikten att uppnå och vilka antaganden bygger den på?
- Omfång (eng. Scope) Etablering av kriterier för att välja studier.
- Sökning (eng. Search) Design och genomförande av en sökstrategi för att finna potentiell forskning.
- Screen (eng. Screen) Genomgång av studiernas relevans och bortsortering.
- Kodning (eng. Code): Samla information för kartläggning av forskning och resultat samt för kvalitetskontroll.
- Kartläggning (eng. Map): Karakterisering och kartläggning av forskningsfältet
- Värdering (eng. Appraise) Bedömning av relevans, användbarhet och kvalitet av studierna.
- Syntetisering (eng. Synthesise): Sammanställning och syntes av forskningsresultat för att besvara översiktens frågor.
- Kommunicera (eng. Communicate): Beskrivning av genomförande av studien, resultaten och implikationer för framtida beslut.

Avseende behov och vem kartläggningen ska användas av är det viktigt att betona att detta projekt ska fungera som ett underlag för Skolforskningsinstitutet. Skolforskningsinstitutet ska i sin tur använda rapporten som ett av flera underlag för att stötta utvecklingsarbete i skolpraktiken. Rapporten ska således inte nödvändigtvis rikta sig direkt till aktörer inom skolväsenden. Eftersom vi har tolkat uppdraget för Skolforskningsinstitutet i termer av att institutet ska stödja aktörer i skolan har vi, enligt ovan, fokuserat på forskning som vi i framtiden menar är viktig för att utveckla den svenska skolpraktiken.

I varje enskilt delprojekt formuleras syften och forskningsfrågor som används som utgångspunkter för övriga steg i processen. De övergripande syftena för projektet har presenterats ovan.

Vilket omfång eller omfattning en kartläggning har bestäms också av vilka studier som inkluderas. En övergripande utgångspunkt för alla tre delprojekten har varit att fokusera på publikationer av hög vetenskaplig kvalitet. Vår bedömning, utifrån erfarenheter från det matematikdidaktiska fältet, är att journalartiklar typiskt håller högre vetenskapligt kvalitet än konferenspaper. Vidare är det inom fältet typiskt stor skillnad mellan olika tidskrifter. Utifrån denna utgångspunkt valde vi att fokusera på journalartiklar publicerade i Web of Science (WoS). Detta val kan naturligtvis diskuteras och utifrån vår erfarenhet finns även relevanta artiklar i tidskrifter utanför WoS. Vi menar dock att inom ramen för detta projekt är begräsningen av sökningen både relevant och nödvändigt. Vi har dock kompletterat läsningen med vissa källor publicerade utanför WoS och logiken för dessa val specificeras inom varje delprojekt. Vidare specificerar varje delprojekt vilka inklusionskriterier för studier som används rörande aspekter såsom studieobjekt, matematikens roll, ålder på elever/studenter och publiceringsår för studierna.

Typiskt har sökningsstrategin varit iterativ i det att initiala sökord har kompletterats med nya som vi upptäckt under sökprocessen. Vi har enligt ovan fokuserat på artiklar i WoS och också använt denna databas för sökning. Specificeringar av söksträngar, kompletterande sökningar i databaser, val i den iterativa processen och andra detaljer redogörs för inom varje delprojekt.

Vid processen att avgöra vilka artiklar som uppfyller inklusionskriterierna studerade vi titel, abstrakt, och nyckelord. I vissa fall krävdes läsning av hela artiklar för att avgöra om studien uppfyllde kriterierna för inkludering. Mer omfattande beskrivningar presenteras i anslutning till varje delprojekt.

Vi har i projektet kodad såväl abstrakt som fulltextartiklar. Kodningen av abstrakt och artiklar skiljer sig delvis åt och kodningsschema presenteras för varje delprojekt. I stora drag har vi fokuserat på studieobjekt, metod och data, teoretiska utgångspunkter, resultat, slutsatser och presentade implikationer för praktiken.

Kartläggning av abstrakt och artiklar syftar till att beskriva den forskning som bedrivs inom området. Vi kartlägger både abstrakt och artiklar och fokuserar här främst på vilka objekt och processer som studeras inom fältet och på vilka sätt som dessa konceptualiseras. Gough et al. (2013) menar att en kartläggning inte typiskt ska betona forskningsresultaten men i vårt fall har vi lyft fram även dessa eftersom vi inte engagerar oss i processerna 8 (Värdering) och 9 (Syntetisering) i detta projekt. Detaljer och specificeringar presenteras i varje enskild projektbeskrivning.

Enligt rekommendationer från Gough et al. (2013) har vi vid rapportering fokuserat på syfte och forskningsfrågor, metodologi, resultat och diskussion av resultaten. Vidare specificerar vi användbarhet, styrkor, svagheter, betoningar och begränsningar med rapporten med avseende på mottagaren. Enligt tidigare har vi skrivit rapporten så att varje delprojekt går att läsa separat. Återigen, detaljer och vägval presenteras i anslutning till varje delprojekt.

Styrkor, svagheter, betoningar och begränsningar

Projektets styrkor, svagheter, begränsningar och betoningar måste förstås i relation till dess syfte, projektbudget och tidsram. Enligt ovan så är det övergripande syftet att producera en kartläggning av forskning och forskningsresultat som ska vara användbar för Skolforskningsinstitutet i dess framtida arbete. Både budget och tidsram för projektet är relativt begränsat i förhållande till den mängd data som sökningar och screening har resulterat i. Detta innebär att vi inte har engagerat oss i en fullskalig litteraturöversikt. Konkret har vi därför (1) valt bort potentiellt användbara databaser och nästan uteslutande fokuserat på WoS, (2) inte hunnit koda mer än en relativt liten del av de artiklar från WoS som vi bedömer som relevanta, (3) inte i någon större utsträckning kunnat avgöra användbarheten av resultaten för den svenska kontexten och (4) därför inte kunnat engagera oss i en syntetisering av forskningsresultaten. Det finns således många delar att arbete vidare med och vi menar att denna rapport kan fungera som ett viktigt underlag.

Projektet främsta styrkor bedömer vi vara fokus och framväxten av konceptualiseringar av forskning som har stor potential att vara användbar för aktörer inom skolvärlden, att vi noggrant har genomfört och rapporterat metodologin samt betoningen på att rapportera studier publicerade från olika ämnesdiscipliner. Rapporten presenterar således en bild av forskning från olika ämnesdiscipliner som fokuserar på hur aktörer, främst lärare, kan agera för att skapa möjligheter för elever att utveckla matematiska förmågor. Utifrån Boesen et al., (2014) samt t ex Hemmi och Ryve (2014) menar vi att stöd för lärare att agera i och utanför matematikklassrummet är av stor vikt för att förbättra elevers kunskaper i matematik.

En potentiellt viktig aspekt som vi inom ramen för detta projekt endast kan diskutera på en abstrakt nivå är transfer av forskningsresultat till praktiken. Inom ramverket för metodologiska tillvägagångssättet (Gough et al., 2013) som presenterades ovan görs sådana ställningstagande i steg 8, värdera (appraise), där bland annat validiteten av forskning och forskningsresultat bedöms. Frågan om hur forskningsresultat om skolan från kulturella kontexter kan användas i andra kulturella kontexter är mycket komplex (Clarke, 2013). Att vi i förhållande till varje studie inte har hunnit värdera och föra resonemang om användbarheten i den svenska kontexten är en tydlig begränsning av denna rapport. Nedan diskuterar vi kort några aspekter som behöver klargöras i framtida arbete och som vi inom ramen för detta projekt inte har hunnit engagera oss tillräckligt i. Det är på intet sätt några utförliga diskussioner men de klargör dock viktiga begräsningar av föreliggande projekt och kan samtidigt ge inspel till framtida arbeten:

- **Bedömning och användbarhet av forskningsresultat från andra kulturella kontexter**. I och med stora internationella studier såsom PISA (Organisation for Economic Cooperation and Development, 2003; 2009) och TIMSS (Mullis, Martin, Foy, & Arora, 2012) har intresset för skola, undervisning och elevers kunskap i

andra länder ökat. Vad gör de länder som lyckas bra på dessa test? Vad mäter testen? Vad kan vi lära oss av länderna som lyckas bra och hur kan detta omsättas, implementeras eller etableras i Sverige? Vi menar att det är rimligt att den svenska skolan kan inspireras, lära sig och i vissa fall kanske direkt implementera arbetssätt, rutiner, och organisatoriska arrangemang från andra kulturella kontexter men att det är långt ifrån självklart vilka resultat som är användbara i den svenska kontexten. Låt oss ta Shanghai, Sydkorea och Japan som exempel. Alla länderna lyckas mycket bra inom både PISA- och TIMSS-studierna medan studier av klassrumsundervisningen i dessa länder visar på tydliga skillnader i central aspekter som lärare-elev interaktion, typer av matematiska problem som behandlas i undervisningen och styrande normer som reglerar synen på lärare och elever (Xu & Clarke, 2013). Det är således oklart både vad det är som gör att dessa länder lyckas och vad som kan vara användbart i den svenska kontexten. Ett sätt att förstå resultaten är att många olika sätt att undervisa är effektiva, att nuvarande forskningsmetodologier inte kan fånga de som förenar de framgångsrika länderna eller att klassrumsundervisning är kulturellt betingat i bemärkelsen att olika typer av klassrumsundervisning är effektiva i specifika kulturella kontexter. Oberoende vad som är förklaringen till att olika typer klassrumsundervisning är effektiva för att elever ska prestera bra på internationella test i matematik måste vi i relation till dessa frågor explicit klargöra hur vi ser på elever, lärare och vilka normer som vi anser är önskvärda i våra svenska skolor. Studier indikerar till exempel att elever i Sydkorea är mycket passiva i klassrummet och att lärarens auktoritet är stor både i termer av talartid och i termer av bärande idéer för klassrumsinteraktionen (Xu & Clarke, 2013). På ett övergripande plan behöver vi fråga oss vad det får för kort- respektive långsiktiga konsekvenser om vi i Sverige agerar för att etablera likartade normer och arbetssätt.

- **Relationen mellan forskningsresultat och lärare**: Relationen mellan forskningsresultat och lärare bör klargöras. Ska lärare förstås som en utförare som ska implementera forskningsresultat eller ska forskningsresultat användas av lärare i syfte att stärka deras professionella kunskap i att fatta genomtänka beslut i design och etablering av klassrumsundervisning (Remillard, Herbel-Eisenmann, & Lloyd, 2009)? I processen att bedöma och använda internationella forskningsresultat bör det således klargöras vilken typ av syn på lärare och dess arbete som är önskvärt och hur insatser bäst ska utformas för att stötta en sådan lärarroll.

- **Olika typer av lärare**: Vi har inom detta projekt inte hunnit bedöma studier och forskningsresultat utifrån utgångspunkten att olika typer av resultat kan vara av skiftande användbarhet för olika typer av lärare. Det är rimligt att anta att olika typer av lärare behöver olika typer av stöd. Utifrån ett sådant perspektiv kan lärare till exempel kategoriseras utifrån utbildningsnivå, erfarenhet, och matematiska kunskaper. Vi bedömer att lärares formella och reella kunskaper i matematik och matematikdidaktik i stor utsträckning kan avgöra vilket typ av stöd och material de kan finna användbara. Vårt projekt indikerar även att just kompetensen hos lärare att använda skriftligt material såsom lärarhandledningar i sig är en viktig aspekt för att förstå klassrumsundervisning. Här behöver mycket mer forskning och resultat kan vara av direkt relevans för hur Skolforskningsinstitutet ska designa material.

- **Olika kontexter inom Sverige**: En ytterligare aspekt som detta projekt och denna rapport inte behandlar ingående är diskussioner av i vilka olika typer av kontexter inom Sverige som forskningsresultat kan vara användbara. I vilken utsträckning är forskningsresultat relevanta för specifika årskurser, elever från olika socioekonomiska förhållanden eller för olika typer av matematiskt innehåll? Framtida arbete bör rimligen nyansera den initiala bild som detta projekt påbörjat att beskriva.

Avslutningsvis, vi bedömer att styrkorna i projektet är de valda fokusområdena och dess potentiella relevans för skolväsendet, den metodologiska noggrannhet med vilken vi har behandlat litteraturen, kvaliteten på kartläggningen samt identifiering och teoretiska diskussionerna av aspekter av relevans för framtida arbete. Begränsningarna i projektet och i denna rapport är många enligt ovan dels beroende av ramarna för projektet men också beroende på vilken typ av forskning och forskningsresultat som finns tillgänglig för att engagera sig sammanställningar av tydlig relevans för det svenska skolväsendet.

2.1 FORMATIV BEDÖMNING: INTRODUKTION

Begreppet formativ bedömning

Formativ bedöming är ett sätt att undervisa och lära, och ska inte förväxlas med bedömning som utförs i betygs- och urvalssyfte (vanligen benämnt summativ bedömning). En svårighet vid slutsatser och sammanställningar av forskningsresultat om formativ bedömning är begreppets innebörd och definition. Under åren har begreppen *formativ bedömning,* och *bedömning för lärande,* använts delvis med lite olika innebörd men ofta i samma grundbetydelse. Det exakta definitionen har också varierat lite mellan författare och har utvecklats med tiden. Den huvudsakliga andemeningen i definitionerna har dock oftast varit densamma (se nästa stycke). Vissa skillnader i innebörd kan dock skapa olika resultat i forskningsstudier. En annan svårighet i forskningssammanställningar inom området är att forskning inom området inte alltid gjorts under samma namn. Olika strategier för formativ bedömning, som t ex feedback och självreglerat lärande, har ofta gjorts under dessa benämningar, och inte under namnet formativ bedömning. En annan svårighet är att effekter av t ex feedback inte bara beror på huruvida en elev får feedback eller inte utan kan också bero på vilken typ av feedback eleven får och på vilket sätt den ges. Forskningssammanställningar av effekter av feedback kan därför ge mått på effektstorlekar av given feedback men det är viktigt att också särskilja olika typer av feedback. Liknande gäller även för forskning där termen som formativ bedömning använts. Den genomförda undervisningen som studerats kan, även om samma term använts, då vara av olika slag och dessutom av olika kvalitet. Sökningen efter publikationer till denna forskningsöversikt kommer därför att inkludera flera olika termer vars innebörd ligger inom den definition på formativ bedömning som ges nedan. De typer av formativ bedömning som studerats i publikationerna kommer också att beskrivas.

I denna forskningsöversikt konceptualiseras formativ bedömning som "…encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged." (Black & Wiliam, 1998a, p.2). Denna beskrivning kan utvecklas till följande mer detaljerade beskrivning:

> "Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about next steps in instruction that are likely to be better, or be better founded, than the decisions they would have taken in the absence of evidence that was elicited." (Black & Wiliam, 2009, p. 9).

Dessa definitioner betyder att formativ bedömning kan utföras på flera olika sätt så som feedback, självreglerat lärande, självbedömning och kamratbedömning. Dessa strategier för formativ bedömning är också inkluderade i tidigare forskningssammanställningar av Black & Wiliam (1998a) och Hattie (2009) där forskning kring formativ bedömning i olika ämnen och åldrar inkluderats (se nedan).

Bakgrund

Tidigare forskningsöversikter (Black & Wiliam, 1998a) och metaanalyser (Hattie, 2009) har dragit slutsatsen att formativ bedömning är ett av de mest effektiva undervisningssätt vi känner till för att höja elevers prestationer. Både Black & Wiliam och Hattie drar slutsatserna att detta gäller både för olika åldrar och ämnen. Black & Wiliams forskningsöversikt inkluderar bara studier om formativ bedömning, men de drar sina slutsatser baserat på en uppskattning av att vanliga effektstorlekar för formativ bedömning ligger i intervallet 0,4-0,7 (Black & Wiliam, 1998b). Hatties metaanalys är en jämförelse mellan en mängd olika variablers korrelationer med elevprestationer. Bland de variabler som har högst korrelation till elevers prestationer återfinns flera strategier för formativ bedömning. Ett exempel är "formativ utvärdering" (formative evaluation) som är den undervisningsvariabel som har högst korrelation till elevprestation. Det är därför inte överraskande att när Hattie i boken, baserat på de funna korrelationerna för olika variabler, presenterar en modell för

framgångsrik undervisning och lärande är den grundläggande hörnstenen i modellen just formativ bedömning (Hattie kallar denna undervisning för synligt lärande).

Både studierna av Black & Wiliam och Hattie har också fått stor genomslagskraft både internationellt och i den svenska diskussionen om undervisning. Hatties analys publicerades i en bok, medan Black & Wiliams analys publicerades i en vetenskaplig tidskrift. Den senare har därför varit föremål för granskning både vid publicering och i många andra forskares artiklar. Det är svårt att hitta publikationer som ifrågasätter forskningsunderlaget för slutsatsen att formativ bedömning verkligen har visat sig kunna vara effektivt och har stor potential. Det har under senare år förekommit en del artiklar som ifrågasatt effektstorlekarna, kvaliteten i några av de studier som exemplifieras i Black & Wiliam, och att underlaget inte är tillräckligt starkt för långtgående slutsatser om alla ämnen och åldrar (Dunn & Mulvenon, 2009; Kingston & Nash, 2011). Men även dessa kritiker tillerkänner potentialen forskningen indikerar. Det som argumenteras för är att det behövs mer forskning för att kunna underbygga mer långtgående slutsatser om de stora effekterna på alla åldrar och ämnen. Både Dunn & Mulvenon och Kingston & Nash baserar deras kritik på deras analys av artiklar ingående i studien av Black & Wiliam (1998a) och några andra utvalda studier. Kingston & Nash (2011) presenterade också en egen metaanalys av effekterna av formativ bedömning, bland annat för ämnet matematik, men den innehöll bara ett begränsat urval av studier där stora delar av viktiga områden inom formativ bedömning, som t ex självreglerat lärande, inte inkluderats på grund av de söktermer de använt. Denna metaanalys fick i sin tur kritik av bland annat metodologiska skäl som t ex vilka studier som inkluderats (Briggs, Ruiz-Primo, Furtak, Shepard & Yin, 2012; McMillan, Venable & Varier, 2013). En forskningsöversikt från det amerikanska utbildningsdepartementet (National Mathematics Advisory Panel, 2008) behandlar, som en del av rapporten, effekter av formativ bedömning på elevers lärande i specifikt matematik. De kommer fram till att formativ bedömning i matematik har stor potential för att höja elevers kunskaper, men att det fattas tillräckligt med forskning av hög kvalitet för mer tillförlitliga slutsatser om effekten av formativ bedömning. På grund av sökkriterierna innehåller översikten dock endast 10 studier. Dessa studier har också mest fokuserat de lägre skolåren och inkluderar oftast speciallärare. Översikten innehåller inte heller några artiklar publicerade efter år 2003.

Det finns därför rent allmänt ett behov av en kunskapsöversikt om formativ bedömning och dess effekter på lärande. Men det finns också ett speciellt behov av en kunskapsöversikt om formativ bedömning i enskilda ämnen som matematik. Trots att formativ bedömning i olika ämnen har mycket gemensamt så har ämnena också olika karaktär och synen på hur formativ bedömning konkret ska operationaliseras kan på grund av denna olikhet skilja sig mellan företrädare för olika ämnen som t ex matematik och språk (Bennet, 2011; Hodgen & Marshall, 2005). Det gör att egenskaperna för genomförande av formativ bedömning delvis kan skilja sig mellan ämnen. Empiriska studier har också visat att effekter av liknande interventioner i olika ämnen, som t ex matematik och språk, givit olika stora utfall (t ex Dignath & Büttner, 2008; Yeh, 2009).

Baserat på ovanstående inledning, som indikerar en stor potential för effekten av formativ bedömning på elevers prestationer i matematik samtidigt som det saknas tillräckligt med forskning på området, finns det anledning att göra en kartläggning över vilka belägg som finns tillgängliga för formativ bedömning och hur starka dessa belägg är. Detta gäller specifikt för ämnet matematik eftersom en ouppklarad fråga i befintlig forskningslitteratur handlar om i vilken utsträckning slutsatser från tidigare forskning kan beläggas för enskilda ämnen. Eftersom forskningsresultat kan vara kontextberoende finns också ett nationellt behov av en kartläggning med ett speciellt delsyfte att identifiera svensk forskning på området eftersom inga sådana studier finns inkluderade i de tidigare forskningsöversikter som beskrivits ovan. Specifikt kommer vi att undersöka forskning som studerat effekterna av formativ bedömning på elevers prestationer i matematik. För dessa studier kommer vi att analysera effekterna av olika strategier för formativ bedömning var för sig, och även som en helhet. Vi redogörar även för forskning om lärarfortbildning i formativ bedömning i matematik. Vi kommer att beskriva internationell forskning men även ha ett speciellt fokus på de forskningresultat som presenterats för den svenska skolkontexten. I den svenska kontexten kommer vi också att analysera studier i syfte att kunna besvara forskningsfrågan i vilken utsträckning och på vilket sätt den svenska matematikundervisningen idag kan betraktas innehålla inslag av formativ bedömning. I samband med att vi beskriver de forskningsresultat som finns i relation till nyss nämnda forskningsområden kommer vi också att beskriva vilken typ av forskning det är som genererar dessa resultat. Ramarna för arbetet med denna kartläggning medger inte en analys av

samtliga publikationer inom områdena, men ambitionen är att forskningsöversikten, utifrån tillgängliga ramar för projektet, ska kunna indikera vilken kunskap vi idag har kring ovanstående fokus och vilken kunskap som fortfarande saknas och behövs.

Forskningsfrågor

Baserat på ovanstående identifierade behov fokuserar denna forskningsöversikt följande forskningsfrågor:

Forskningsfråga 1: Vilket samband finns mellan formativ bedömning och elevers prestationer i matematik?
Forskningsfråga 2: Vilka egenskaper hos formativ bedömning är centrala i detta samband?
Forskningsfråga 3: Vilka egenskaper hos fortbildningar är centrala för att lärare ska vilja och kunna använda formativ bedömning på ett framgångsrikt sätt i sin undervisning?
Forskningsfråga 4**:** Vilken typ av forskning bedrivs för att besvara forskningsfrågorna?
Forskningsfråga 5: Hur ser svensk klassrumspraktik ut idag med avseende på formativ bedömning?

2.2 FORMATIV BEDÖMNING: METOD

För att kunna besvara forskningsfrågorna har ett antal val gjorts och beslut tagits kring utformning och begränsningar för litteratursökningen. I avsnittet ”Sökmetod och sökresultat” nedan beskrivs den metod som använts för litteratursökningen, och resultatet av sökningen i termer av antal publikationer som via denna sökning inkluderats i forskningsöversikten. I det senare avsnittet ”underlag för val av söktermer” beskrivs den process som använts för att ta fram ett underlag för besluten om vilka söktermer som skulle ingå i sökningen.

Sökmetod och sökresultat

Databaser och sökfilter

Med sökfilter menas de inställningar som kan göras vid sökning i databaser. Olika databaser har olika möjligheter att filtrera. Vid filtrering kan exempelvis typ av publikation och publikationer inom vissa årtal eller ämnesområden väljas. Vid sökning av internationella publikationer med relevans för forskningsöversiktens forskningsfrågor användes databasen Web of Science (WoS). För att få en noggrannare bild av hur forskningsområdet ser ut i Sverige så användes, i tillägg till WoS, även den nationella databasen SwePub för att hitta svensk forskning inom området. SwePub innehåller referenser från ett trettiotal lärosäten i Sverige. Sökningen gjordes inom alla forskningsområden och forskningsämnen inom både WoS (core collection) och SwePub då formativ bedömning är av intresse för många ämnesområden. Sökningen omfattar artiklar publicerade 2005-2014. För internationella publikationer söktes endast artiklar i vetenskapliga tidskrifter, Böcker, konferensbidrag och avhandlingar har filtrerats bort. Argument för detta val är att viktiga och intressanta resultat som presenteras i dessa former har i de flesta fall även publicerats i artikelform. För att kunna hitta all publicerad forskning som pågår inom området i Sverige har ingen begränsning av typ av publikation gjorts i sökningen i SwePub. Att kunna inkludera alla publikationsformer bedömdes vara möjligt baserat på förkunskaper om trolig mängd svensk forskning inom området formativ bedömning i matematik. SwePub har dock vissa begränsningar då databasen baseras på olika publiceringsdatabaser (t.ex. digitala vetenskapliga arkivet, DIVA). I dessa databaser är det forskarna själva som bokför vad som skrivs in och det kan medföra att information ibland saknas. För att säkerställa att svensk forskning (forskning i den svenska kontexten) som publicerats internationellt, och återfinns i WoS men inte i tidskrifter, inkluderas i forskningsöversikten så genomfördes ytterligare en sökning i WoS. I denna sökning användes samma filtrering som för SwePub (inte bara artiklar) men med specifika söktermer som indikerar svensk forskning (använda söktermer beskrivs nedan). Sökresultaten från sökningen i WoS och i SwePub jämfördes för att säkerställa att inga svenska publikationer missats.

Söktermer

Vid sökningen användes de söktermer som i forskningslitteratur används för olika strategier för formativ bedömning. Sökningen genomfördes med både engelska och svenska söktermer. För att sökningen ska ge utfall i relevant litteratur för forskningsfrågorna användes även söktermer för matematik, elever, fortbildning, och klassrumspraktik. Flera olika termer kan vara relevanta för innebörden av t ex formativ bedömning eller fortbildning. I denna forskningsöversikt kallas dessa termer för *söktermer* och de söktermer som beskriver samma innebörd kallas för en *sökgrupp*. T ex inkluderar den sökgrupp som kallas för formativ bedömning bland annat söktermerna formativ bedömning, bedömning för lärande och formativ utvärdering. Formativ bedömning kommer därför att ibland användas i betydelsen sökgrupp, och ibland i betydelsen sökterm. När ordet används som sökgrupp är det kursiverat. De söktermer, och de sökgrupper de tillhör, som används i sökningen återfinns i Tabell 1. Olika avstavningar har använts för att inkludera olika varianter av en sökterm i sökningen. Exempelvis så har söktermen self-assessment sökts med varianterna self-ass*, selfass och self ass*. Söktermen instruction har sökts med varianten instruc*, och söktermen Sweden har sökts med varianten Swe*. Symbolen * har funktionen att alla ord som kan bildas med alla möjliga bokstavskombinationer som följer

bokstäverna före symbolen kommer att användas i sökningen. En sökning med Swe* kommer därför att inkludera både Swedish och Sweden.

I databaserna kan sökning ske med söktermerna angivna för titel eller i andra delar av publikationerna. Vid sökningen i både WoS och SwePub användes söktermerna för sökgruppen *formativ bedömning* för sökning i titeln för publikationen. Det betyder att sökningen identifierar de publikationer som har någon av söktermerna för sökgruppen *formativ bedömning* i titeln. För att en publikation skulle inkluderas i forskningsöversikten krävdes också att en sökterm för sökgruppen *matematik* och en sökterm för antingen sökgruppen *elever,* sökgruppen *fortbildning,* eller sökgruppen *klassrumspraktik* skulle finnas i publikationens topic (titel, sammanfattning och nyckelord). Detta för att begränsa antalet publikationer till de som är mest relevanta för de ställda forskningsfrågorna, och inte t ex inkludera studier om feedback i andra miljöer än skolan. Sökgruppen klassrumspraktik användes för att identifiera studier om egenskaper i svensk undervisning med avseende på formativ bedömning. Söktermen Sweden inkluderades när sökningen med sökgruppen *klassrumspraktik* skedde i WoS.

Tabell 1. Sökgrupper och Söktermer. Sökgruppen formativ bedömning har i sökningen kombinerats med matematik och elever samt fortbildning och klassrumspraktik.

Sökgrupp	**Engelska söktermer**	**Svenska söktermer**	**Placering vid sökning**
formativ bedömning	formative assessment, assessment for learning, formative evaluation, self-assessment, self-regulated learning, peer assessment, feedback, response-system	formativ bedömning, bedömning för lärande, formativ utvärdering, självbedömning, självreglerat lärande, kamratbedömning, kamratrespons, feedback, återkoppling, responssystem	Titel
matematik	mathematics	matematik	Topic
elever	pupil, student	elev, student	Topic
fortbildning	in-service training, professional development, program, implement,	fortbildning, kompetensutveckling, program implementering	Topic
klassrumspraktik (användes enbart för sökning efter nationell forskning	instruction, teaching, teacher, pedagog, classroom, practice (+Sweden i WoS)	undervisning, lärare, pedagog, klassrum, praktik	Topic

Urval

De publikationer som identifierats med ovanstående sökning (104 internationella och 39 nationella publikationer) kodades för att kunna fatta beslut om dess relevans för forskningsöversiktens forskningsfrågor och kontext (t ex elever, lärare, skola), och därför om beslut för dess slutliga inkludering i forskningsöversikten. Urvalet av publikationer skedde i två steg. Publikationerna kodades för titel, typ av publikation (t ex tidskrift), typ av studie, elevers ålder, elevernas årskurs, skolämne, studiens syfte samt typ av resultat. När kodningen gav positivt svar för alla kriterier som beskrivs nedan behölls publikationen till nästa urvalsomgång.

- Publikationen innehöll relevant innebörd av söktermerna (när till exempel den engelska termen ”pupil” står för innebörden av det svenska ordet ”pupill” och inte ”elev” sorterades publikationen bort)
- Publikationen behandlade skolämnet matematik (publikationer som inte behandlade matematik men ändå dykt upp i sökningen sorterades bort)
- Publikationen behandlade åk 1-9 eller gymnasiet

- Publikationen innehöll en studie som var empirisk (eller forskningsöversikter av empiriska studier)
- Publikationer som i sin helhet var möjliga att på ett relativt enkelt sätt hitta (konferensabstrakt i SwePub innehöll inte alltid information om t ex konferensnamn och årtal)
- Publikationen beskriver forskningsmetod och forskningsresultat (konferensabstrakt i SwePub innehöll inte alltid denna information)

De publikationer vars kodning givit ett positivt svar på något av ovanstående kriterier (60 publikationer) behölls och kodades i ett steg 2. Urvalskriterierna i detta urvalets andra steg utgick också från denna forskningsöversikts forskningsfrågor. Publikationen skulle motsvara något av kriterierna:

- Studiens syfte behandlar formativ bedömnings effekt på elevers matematikprestationer
- Studiens syfte beskriver samband mellan den formativa bedömningspraktiken och prestationer i matematik
- (Typ av) resultat visar att studien helt eller delvis beskriver effekt på elevers prestationer eller egenskaper hos den formativa praktiken som visar effekt på elevers prestationer
- Resultaten behandlar egenskaper hos en formativ bedömningspraktik, relaterat till matematik
- Resultaten behandlar effekter eller egenskaper hos en kompentensutveckling i formativ bedömning

Efter ovanstående urval återstod 39 (30 internationella och 9 nationella) publikationer. Sammanfattningsvis identifierade sökningen i databaserna Web of Science och SwePub 140 publikationer varav 39 publikationer kvarstod för djupanalys för att besvara forskningsöversiktens forskningsfrågor. I Tabell 2 redovisas resultaten av de olika sökningarna i termer av antal publikationer.

Tabell 2. Resultat av sökning (söktermer, antal sökträffar samt antal sökträffar efter urval). Söktermerna i tabellen är översatta till svenska. En publikation kan ha hittats i sökningen på flera söktermer. Antalet publikationer är därför mindre än antalet sökträffar.

Resultat av sökning för internationell forskning i WoS			
Sökgrupp i titel	**Sökgrupp i Topic (Titel, keywords och abstrakt)**	**Antal sökträffar**	**Antal sökträffar efter urval**
Formativ bedömning	matematik AND elev	104	34
Formativ bedömning	matematik AND fortbildning	22	6
Totalt för internationell sökning i WoS		126 träffar bestående av 104 artiklar	42 träffar bestående av 30 artiklar
Resultat av sökning för nationell forskning i SwePub			
Sökgrupp i titel	**Sökgrupp i Topic (Titel, keywords och abstrakt)**	**Antal sökträffar**	**Antal sökträffar efter urval**
Formativ bedömning	matematik AND (elev OR klassrumspraktik)	36	12
Formativ bedömning	matematik AND fortbildning	13	7
Totalt för nationell sökning i SwePub		55 träffar bestående av 39 publikationer	16 träffar bestående av 9 publikationer
Resultat av sökning för nationell forskning i WoS			
Sökgrupp i titel	**Sökgrupp i Topic (Titel, keywords och abstrakt)**	**Antal sökträffar**	**Antal sökträffar efter urval**
Formativ bedömning	matematik AND Sweden	0	0
Totalt för nationell sökning i WoS		0	0

Underlag för val av söktermer

I detta avsnitt beskrivs den process som ledde fram till de val av söktermer som gjordes och val av var i publikationen som söktermerna skulle återfinnas (titel eller topic). Utgångspunkten för val av söktermer var att de artiklar som söktes skulle behandla formativ bedömning i skolämnet matematik och i klassrumspraktiken, dvs. elevers lärande eller lärares undervisning. Det innebar att först identifiera termer som handlar om formativ bedömning, och därefter undersöka hur specifika övriga söktermer behövde vara för att antalet relevanta publikationer som hittades skulle vara möjliga att analysera inom ramen för projektet. Ett femtiotal publikationer bedömdes vara möjliga att hantera för djupanalys.

Val av söktermer och placering, internationellt

För databasen Web of Science gäller att placering av söktermer kan väljas i titel eller i topic (i topic kan söktermerna återfinnas i publikationernas titel, sammanfattning eller nyckelord). Citationstecken kan användas kring sökord som innehåller fler än en term för att sökningen enbart ska välja ordföljden inom citationstecknen. Asterixer (*) används för att få med alla böjningar av en term. Sökordet lärar* motsvarar sökning för lärare, lärarna, lärares, lärarhandledning etc. Operatorerna AND och OR används för att kombinera sökord och sökgrupper i sökningar. När flera söktermer kombineras i sökningen förekommer det att databasen tar med ett antal träffar som inte helt motsvarar träffar för alla söktermer i sökningen.

Vid val av söktermer och dess placering i sökningen valdes, baserat på kartläggningens inriktning, att utgå från att söktermen matematik (mathematics) skulle placeras i topic för alla sökningar. För att identifiera lämpliga söktermer för formativ bedömning (formative assessment) användes Black & Wiliam (1998) samt Wiliam &Thompson (2007) då dessa båda publikationer är tongivande inom området. Den första av de två är en mycket inflytelserik forskningsöversikt inom området formativ bedömning (inte fokuserad på något specifikt ämne), och den andra innehåller ett förslag på hur olika strategier för formativ bedömning kan ses som en integrerad helhet. Genom att använda flera söktermer för formativ bedömning (formative assessment, assessment for learning, formative evaluation, feedback, response-system, self-assessment, self-regulated learning, and peer assessment) beaktas olika aspekter av formativ bedömning i sökningen.

Flera termer som mer eller mindre kan knytas till formativ bedömning kan användas som söktermer. För att hitta söktermer som ger ett hanterligt urval av publikationer gjordes flera testsökningar med olika val av söktermer. Baserat på dessa testsökningar minskades antalet söktermer till de termer som bedömdes vara vanligast för att beskriva de strategier som ofta kopplas ihop med formativ bedömning och som inkluderats i forskningsöversikten av Black & Wiliam (1998). Dessa söktermer återfinns i Tabell 1 och utgör sökgruppen *formativ bedömning*. I Tabell 2 ovan presenteras antal sökträffar för den slutgiltiga sökningen baserat på de söktermer som valts ut efter testsökningarna. Citationstecken kring sökord innehållande flera ord användes inte då ordföljden i orden inte skulle vara exkluderande.

Beslut behövde också tas om söktermerna för sökgruppen *formativ bedömning* skulle sökas inom topic, tillsammans med sökgruppen *matematik*, eller bara i publikationernas titel. I Tabell 3 syns resultatet av fyra olika sökningar som exemplifierar beslutsunderlaget. Ett beslut togs att bara söka i publikationernas titel eftersom sökning i topic gav för många träffar för att vara hanterligt. En sökning med termerna feedback och matematik i topic gav 3115 träffar (Sökning 1 i Tabell 3). Av dessa identifierades en mängd artiklar som använde termen feedback för andra områden än de avsedda, exempelvis för området teknik. En utökning av antalet söktermer i topic, med t ex elev så att att endast publikationer där också denna term finns med identifieras (Sökning 2), ökade andelen artiklar med relevans, men antalet identifierade publikationer var fortfarande för högt med tanke på att en slutlig sökning även ska göras för de andra söktermerna i sökgruppen *formativ bedömning*. Den slutliga bedömningen var att vi skulle få en större andel relevanta artiklar som besvarade våra forskningsfrågor och ett möjligt antal artiklar att hantera genom att placera sökgruppen *formativ bedömning* i titel och övriga sökgrupper (*matematik*, *elev*, *fortbildning*) i topic, vilket exemplifieras av Sökning 3 och 4 i Tabell 3.

Tabell 3. Exempel på sökningar som underlag för beslut om hur sökning skulle ske i Web of Science

	Sökväg, titel eller topic **Beteckningen ”TI= ” betyder att söktermen inom parentes återfinns i titel. Beteckningen ”TS=” betyder att söktermen inom parentes återfinns i titel, sammanfattning eller nyckelord.**	**Antal Träffar**
1	TS= (feedback) AND TS = (math*)	3115
2	TS= (feedback) AND TS = (math* AND student*)	307
3	TI= (feedback) AND TS = (math* AND student*)	22
4	TI= (peerass* OR peer ass* OR peer-ass*) AND TS = (math* AND student*)	9

För att kunna besvara Forskningsfråga 3, ”Vilka egenskaper hos fortbildningar är centrala för att lärare på ett framgångsrikt sätt ska använda formativ bedömning i sin undervisning?”, valdes sökgruppen *fortbildning.* Till den gruppen valdes söktermerna in-service training, professional development, program, samt implement som alla är termer som använts för fortbildning (Villegas-Reimers, 2003). Eftersom det är fortbildning av en specifik yrkesgrupp som är vårt intresse så lade vi till söktermen lärare till sökgruppen. Sökgruppen *fortbildning* placerades i topic. De slutliga sökgrupperna och söktermerna som använts för kartläggningen presenteras i Tabell 1.

Val av söktermer och placering, nationellt

För databasen SwePub gäller samma förutsättningar som för Web of Science (WoS). SwePub är dock en enklare sökmotor som inte hanterar lika många operatorer (AND, OR) i varje sökning. Den nationella sökningen syftade till att belysa den svenska forskningen av relevans för forskningsfrågorna. Av den anledningen användes samma söktermer som för den internationella sökningen, vilket motsvarar en sökning av publikationer för att kunna besvara Forskningsfrågorna 1-4 presenterade i Avsnitt 2.1 ovan. Men sökningen efter svenska publikationer gjordes också med syfte att kunna besvara Forskningsfråga 5 som handlar om hur svensk klassrumspraktik ser ut idag med avseende på formativ bedömning. För sökning av publikationer av relevans för denna fråga använde ytterligare söktermer (se nedan).

Motsvarande testsökningar som gjordes för den internationella sökningen gjordes även för den nationella sökningen. Sådana testsökningar exemplifieras i Tabell 4 genom söktermerna formative assessment, formative evaluation, assessment for learning samt feedback. Detta utgör 4 av 8 söktermer i sökgruppen *formativ bedömning* (formative assessment, assessment for learning, formative evaluation, feedback, response-system, self-assessment, self-regulated learning, and peer assessment). Då vi placerade dessa fyra sökord i topic tillsammans med sökgrupperna *matematik* och *elev* blev det 279 träffar (Sökning 1 och 2). När formativ bedömning och feedback placerades i titel, med matematik i topic får vi 60 träffar totalt (Sökning 3 & 4). Det indikerar fortfarande ett alltför omfattande sökområde. Av den anledningen valde vi att göra sökningar på samma sätt som i WoS (Sökning 5 & 6). Resultatet blev då 11 träffar och då bör även beaktas att motsvarade sökning skulle göras med svenska sökord.

Tabell 4. Exempel på sökningar som underlag för beslut om hur sökning skulle ske i SwePub

	TI (Titel)	**TS (titel, nyckelord eller abstrakt)**	**Antal träffar**
1		(formative assessment OR formative evaluation OR assessment for learning) math* AND (stud* OR pup*)	175
2		(feedback) math* (stud* OR pup*)	104 träffar
3	(formative assessment OR formative evaluation OR assessment for learning)	Math*	13 träffar

	TI (Titel)	TS (titel, nyckelord eller abstrakt)	Antal träffar
4	Feedback	Math*	48 träffar
5	(formative assessment OR formative evaluation OR assessment for learning)	Math* AND student*	7 träffar
6	Feedback	Math* AND student*	4 träffar

De engelska termerna översattes till svenska och hänsyn togs till hur Wiliam (2013) och Lundahl (2014) använder termerna. Vi hade redan söktermen elev, så för Forskningsfråga 5, om den formativa klassrumspraktiken, utökades sökningen till att omfatta lärare och hela undervisningsmiljön (pedagog, lärare, undervisning, klassrum, praktik). Dessa termer översattes även till engelska och användes vid sökning i både SwePub och WoS (se Tabell 1).

Vid sökning i WoS avseende Forskningsfråga 5 användes samma sökfilter som för SwePub, dvs. även andra publikationer än artiklar inkluderades i sökningen. Sökningen gjordes med tillägget Sweden (Swe*) i topic då det var svensk forskning vi var intresserade av. Den sökningen gav inga träffar. Därmed behövdes inte ytterligare sökningar göras för söktermerna elev, lärare och klassrum.

Sammanfattningsvis så kan vi konstatera att några av begreppen inom formativ bedömning används inom helt andra områden. Feedback används exempelvis inom flera tekniska områden. Detta medförde att vi valde att lägga sökgruppen formativ bedömning i titeln för vår sökning i både den internationella och den nationella sökningen. Övriga sökgrupper har placerats i topic (titel, nyckelord eller sammanfattning). Söktermen matematik användes i alla sökningar.

2.3 FORMATIV BEDÖMNING: RESULTAT, INTERNATIONELLA PUBLIKATIONER

Detta avsnitt behandlar internationella (inklusive svenska) publikationer publicerade i tidskrifter registrerade i Web of Science. Forskningsfrågorna 1-4 besvaras nedan för varje strategi för genomförande av formativ bedömning. Dessa strategier är formativ bedömning, feedback och självreglerat lärande. Inga artiklar som motsvarade sökkriterierna fokuserade på begreppen kamrat- och självbedömning. Forskningsfråga 5 om svensk klassrumspraktik i relation till formativ bedömning besvaras i Avsnitt 2.4 som handlar om nationella publikationer. Alla resultaten baserar sig på utvalda artiklar enligt sökmetod beskriven i Avsnitt 2.2.

Formativ bedömning, bedömning för lärande och formativ utvärdering

I avsnittet presenteras först en kort beskrivning av innebörden i begreppen formativ bedömning och bedömning för lärande som är de termer som använts i de artiklar som inkluderats i detta avsnitt. Därefter besvaras forskningsfrågorna. Dessa svar baseras på artiklarna som inkluderats i forskningsöversikten och som sammanfattas i det påföljande avsnittet.

Definition av formativ bedömning

Formativ bedömning har i forskningslitteraturen definierats på lite olika sätt, vilket kommer att framgå av de nedanstående artiklarna. Den bärande idén i innebörden av formativ bedömning, bedömning för lärande respektive formativ utvärdering, som är begreppen som behandlas i detta avsnitt, är dock i alla inkluderade artiklar att information om elevernas kunskaper samlas in för att sedan användas till att anpassa den fortsatta undervisningen (Burns, Klingbeil, and Ysseldyke, 2010) Det kan innebära både en läraranpassning i termer av att läraren använder informationen för att ta beslut om övergripande strategier i klassrummet och att eleven individuellt leds vidare i sitt lärande genom t ex. hintar och stödjande frågor (Koedinger, Mclaughlin, & Heffernan, 2010). Läraren behöver i båda fallen, i rätt tid, ge återkoppling för att bekräfta korrekta elevsvar och stödja en förändring av mindre funktionella begreppsuppfattningar lösningsmetoder (Clarke, Doabler, Cary, Kosty, Baker, Fien, & Smolkowski, 2014). Denna innebörd ligger inom den mer omfångsrika definition av formativ bedömning av Black & Wiliam (2009) som gavs i inledningen av denna rapport och som också rymmer andra fokus som t ex stödjande av eleverna som agenter i utförandet av den formativa bedömningen (se avsnittet om självreglerat lärande nedan).

Forskningsfråga 1: Vilket samband finns mellan formativ bedömning och elevers prestationer i matematik?

Fem studier, varav en reviewstudie (Yeh, 2009) som i sin tur innehöll två studier, undersökte effekter av formativ bedömning på elevers matematikprestationer. De två studierna som analyserats av Yeh (2009) (en experimentell och en kvasi-experimentell studie) fann båda signifikanta prestationsökningar som följd av formativ bedömning med en medeleffektstorlek på 0,4. I jämförelse med effekterna av minskade klasstorlekar var det senare 124 gånger mindre kostnadseffektivt. Även i den kvasiexperimentella studien av Burns et al. (2010) var formativ bedömning effektivt då ökningen av elevprestationerna för gruppen som använt datorprogrammet TEFE jämfört med kontrollgruppen var statistiskt signifikant ($p<0,001$). Någon effektstorlek angavs inte. Även i den tredje studien (Koedinger et al., 2010) som inkluderade datorprogrammet ASSIST ökade elevernas prestationer för alla elever ($p<0,001$, $d=0,23$) men speciellt för elever i matematiksvårigheter ($p<0,001$, $d=0,50$). I studien av Clarke et al. (2014) studerades endast elever i matematiksvårigheter. Interventionsgruppen ökade sina prestationer mer än kontrollgruppen på ett av testen på begreppskunskap ($p<0,0015$, $g=0,8$). Dock fanns ingen signifikant skillnad på ett annat av deltesten på begreppsförståelse och

inte heller på automatisering av procedurkunskaper. I studien av Phelan et al. (2011) användes bedömningsmaterialet Powersource. Materialet med den korta utbildningen var inte tillräckligt för att ge lärarna sådant stöd att det gav signifikant ökade elevresultat jämfört med kontrollgruppen totalt sett. Däremot fanns en signifikant effekt för eleverna med höga resultat på förtestet ($p<0,01$, $d=0,5$).

Effekten av formativ bedömning på elevernas prestationer var signifikant för den totala elevgruppen i alla studier utom den av Phelan (2011) som innefattade en mycket kort utbildning för lärarna. Effektstorlekar angavs inte men i de fyra studier som de angavs låg effektstorleken mellan 0,4 och 0,8. Den högsta effektstorleken kom dock när elevurvalet var elever i matematiksvårigheter (Clarke et al., 2014). I den enda studien där signifikanta effekter inte identifierades för hela gruppen elever, identifierades dock signifikanta effekter på elever med höga resultat på förtestet (Phelan et al., 2011). Det betyder att i en studie som specifikt studerade elever med låga provresultat visades speciellt höga effekter på dem, medan i en annan studie var det de elever med höga resultat där effekten var som störst. Generellt visar studierna på en positiv effekt av formativ bedömning på elevers matematikresultat, men det finns en variation i effektstorlekar och vilka elever som gynnas mest. Detta kan ha flera orsaker men en kan vara att den typ av formativ bedömning som används i studierna är olika, vilket diskuteras i Forskningsfråga 2.

Forskningsfråga 2. Vilka egenskaper hos formativ bedömning är centrala i detta samband?

I studierna som undersökts i Yeh (2009) används datorbaserade adaptiva tester 2-5 gånger i veckan där sedan individuell återkoppling skedde till eleverna. Lärarna fick också sammanfattande information kring detaljer om klassens och individuella elevers prestationer på proven. Informationen används för att anpassa undervisningen efter klassens och individuella elevers identifierade lärandebehov. Undervisningen blev mer individualiserad och effektiv, och eleverna upplevde kontroll då de fick feedback om sin progression i lärandet. Den frekvent återkommande informationen om elevernas kunskaper, och att denna information används för de anpassningar av undervisning och lärande som görs i klassrummet, verkar därför vara centrala för effekten på elevers matematikprestationer som efter ungefär ett halvår var statistiskt signifikant med en medeleffektstorlek på 0,4 över de båda studierna som fanns i denna review. Lärarna hade också fått kortare träning i hur informationen från proven kan användas för beslut om de kommande stegen i undervisningen. Däremot verkar ingen substantiell träning ingått kring hur eleverna kan reglera sitt lärande, och det datorbaseade systemet är begränsat till de uppgifter som ingår i databasen och att användas i sin helhet vid specifika tillfällen. I de två studierna av Burns et al. (2010) respektive Koedinger et al. (2010) användes andra datorprogram som stöd för lärarens undervisning. I den första av studierna användes programmet TEFE som genererar arbetsblad till eleverna, rättar dem och meddelar när det är dags för test för den individuella eleven. Programmet rättar även testet och genererar nya arbetsblad vid behov. Läraren får information om alla dessa steg för varje elev och har då underlag för att ta beslut om sin undervisning. I studien av Koedinger et al. (2010) användes programmet ASSIST som också är tänkt att frekvent generera små tester. Programmet gör dock mer än bara generera och rätta test. De ger också korrigerande feedback och leder eleven vidare med hintar och stödjande frågor. I dessa studier mättes effekterna för ett eller flera års implementering av undervisningen. Båda dessa stöd för undervisningen gav signifikanta effekter på elevernas prestationer men inga effektstorlekar angavs i någon av studierna. I den femte studien (Clarke et al., 2014) fick eleverna i matematiksvårigheter undervisning i smågrupper där en av flera komponenter var att frekvent ge återkoppling till eleverna för att bekräfta deras korrekta svar och rätta till eventuella missuppfattningar. Denna undervisning bedrevs i snitt två gånger i veckan under ett halvår. I den enda studien som inte gav signifikanta effekter på hela den grupp som studerades (men dock på de duktigare eleverna) var studien av Phelan et al. (2011). I den användes ett stödmaterial, som inte var datorbaserat, som skulle hjälpa läraren att återkommande samla in information om elevernas kunnande och stödja lärarnas uppföljande undervisning baserat på denna information. Den skiljer sig också från de andra i det att interventionen genomfördes under en mycket kortare tid, 8 lektioner jämfört med de andra studiernas genomförandetid på ett halvår upp till över 5 år. Den tid lärarna fick till fortbildning av att undervisa med hjälp av Powersource var också endast 9 timmar. Det var visserligen inte mindre än de som genomförde undervisningen i studien av Clarke et al., men i den studien var det inte de ordinarie lärarna som genomförde

interventionen utan utomstående som också bara hade eleverna i smågrupper. I de två datorbaserade interventionerna i studierna av Burns et al. (2010) och Koedinger et al. (2010) står inget om fortbildning men där sköter datorprogrammen en stor del av interventionen. Även interventionerna i Yeh (2009) baserades på datorprogram, men innehöll också en kortare fortbildning för lärararna och kontinuerlig support om datorprogrammet..

Sammanfattningsvis genomfördes formativ bedömning på olika sätt i de olika studierna, men positiva effekter på elevernas lärande identifierades i alla studier. Den komponent som fanns i alla de olika sätten att genomföra den formativa bedömningen är att information ofta insamlas om elevernas kunnande och att läraren (och möjligen eleverna), delvis med hjälp ett datorprogram, anpassar undervisning och lärande efter elevernas kunnande och lärandebehov. Den studie där effekterna inte kunde identifieras för hela den undersökta elevgruppen (Clarke et al., 2014) var den studie där läraren själv skulle ansvara för utvecklingen av undervisningen, med stöd av ett befintligt material, men där kompetensutvecklingen var mycket begränsad i tid.

Forskningsfråga 3: Vilka egenskaper hos fortbildningar är centrala för att lärare ska vilja och kunna använda formativ bedömning på ett framgångsrikt sätt i sin undervisning?

Tre av studierna innehöll fortbildning i vid bemärkelse (Phelan et al, 2011, Clarke et al, 2014 och Lee et al 2012) men det är enbart Phelan som både undersöker om fortbildningen har effekt på elevernas matematikkunskaper och beskriver fortbildningens egenskaper. Fortbildningen i användandet av ett bedömningsmaterial (Powersource), på totalt 9 timmar och 4 träffar, innehöll information om de teoretiska grunderna för projektet samt undervisning om hur man analyserar elevresultat och anpassar undervisningen. Vid dessa tillfällen fick lärarna söka mönster i elevsvar och leta vilka begreppsmissuppfattningar de kunde identifiera. Kontrollgruppens lärare deltog i en alternativ fortbildning med liknande struktur som den för försöksgruppen, men med ett annat innehåll. Fortbildning i formativ bedömning (med hjälp av Powersource) gav signifikant effekt för högpresterande elever och svårare algebrauppgifter medan kontrollgruppens elever presterade bättre inom delområdet ”egenskaper hos aritmetik”

I artikeln av Clarke et al. beskrivs begränsade effekter av en 9 timmar lång fortbildning när det gäller elevernas prestationsökningar. Fortbildningen beskrivs dock inte. I artikeln av Lee et al så beskrivs fortbildningen och svårigheter för lärarna att använda dess innehåll i sin undervisning, men man undersöker inte effekten på elevers prestationer.

Forskningsfråga 4: Vilken typ av forskning bedrivs för att besvara ovanstående forskningsfrågor?

Studierna som analyserats har en jämn spridning över åldrar för studerade elever från förskoleklass upp till högstadiet. De flesta av studierna inkluderar stora elevurval med tusentals elever. Några av studierna är experimentella medan flera är kvasiexperimentella vilket gör dessa resultat mer osäkra, då grupperna som jämförs inte nödvändigtvis är jämförbara. Samtliga studier var från USA. Fem av de sex studierna undersökte Forskningsfråga 1, och en av studierna (Lee et al., 2012) fokuserade på Forskningsfråga 3 som handlar om vilka egenskaper som är central för att lära ska vilja och kunna använda formativ bedömning på ett framgångsrikt sätt. Två av studierna inkluderade studier på elever i särskilda behov. Den studie som bara studerade elever i särskilda behov var den studie som hade ett litet mindre urval av elever. I de fem artiklar som studerade Forskningsfråga 1 användes olika statistiska metoder där p-värden angavs, medan effektstorlekar angavs i några fall. Interventionerna som studerades varierade från 8 lektioner upp till 5 år, där de flesta var under minst ett halvt år. Alla studier som fokuserade Forskningsfråga 1 var interventionsstudier där undervisningen skulle bygga på formativ bedömning och effekterna av denna undervisning i form av elevers prestationer på prov jämfördes med prestationerna av elever i kontrollgrupper där undervisningen inte påverkats av tankar om formativ bedömning. Typen av intervention och lärarmedverkan varierade mellan studierna, men handlade om lärarens (eller datorns) formativa bedömningspraktik via bedömning, feedback och anpassning av undervisning (till skillnad från ett fokus på eleven som agenten för den formativa bedömningen i form av självreglerat lärande). I två av studierna var dataprogram huvudfokus i den formativa bedömningen, i

en av studierna var det utomstående som kom in och bedrev undervisningen (för smågrupperna med elever i matematiksvårigheter) och i en av studierna var det de vanliga lärarna som fick fortbildning och som sedan skulle genomföra de nya idéerna om formativ bedömning. Studien som behandlade Forskningsfråga 3 fokuserade på lärarna och innehöll enkäter, intervjuer, loggböcker med mera.

Mångfalden av förhållanden som ingått i studierna är en fördel ur perspektivet att studierna belyser flera olika aspekter av sambandet mellan formativ bedömning och elevers prestationer. Men eftersom antalet studier är få har denna mångfald också en stor nackdel i det att resultaten från studierna sällan behandlar exakt samma sätt att genomföra formativ bedömning, vilket betyder att de inte kan ha funktionen som upprepande studier för att slutsatserna ska bli säkrare. Däremot innehåller de alla samma grundingrediens och stödjer därmed tillsammans slutsatsen om effekterna av den grundläggande underliggande idén i formativ bedömning som innebär att frekvent insamling av information om elevernas kunnande används för anpassning av undervisningen efter dessa identifierade lärandebehov. De effektstorlekar som funnits i dessa studier ska beaktas i ljuset av den information och anpassningar av undervisning och lärande som möjliggörs men också begränsas av de datorbaserade test som är grunden för den formativa bedömningen i dessa studier. Lärarna får hjälp av dessa program, men begränsas av att inte ha fått fortbildning och stöd för att kontinuerligt själva driva den formativa bedömningen genom andra informella bedömningar och skapandet av lämpliga anpassningar av undervisningen.

Sammanfattning av artiklarna

En av artiklarna om formativ bedömning är en amerikansk reviewartikel som jämför kostnadseffektiviteten av minskade klasstorlekar med användning av ”snabb” formativ bedömning (Yeh, 2009). Snabb formativ bedömning definieras som ett system som ger formativ teståterkoppling till elever och lärare om elevernas resultat i matematik och läsning 2-5 gånger i veckan. Undervisningen bygger på en använding av ett datorbaserat program för adaptiva prov. Eleverna förtestas genom ett 15 minuter långt prov. Baserat på resultaten genereras sedan uppgifter på en lämplig nivå från en stor datadabas av uppgifter. Eleverna arbetar med uppgifterna med papper och penna, men får dem bedömda av datorprogrammet genom att scanna sina lösningar. Datorprogrammet ger sedan omedelbar feedback till lärare och elever. Läraren får, för varje elev, sammanfattande information om detaljer kring deras kunskaper i ämnet. Informationen används för att anpassa undervisningen efter klassens och individuella elevers identifierade lärandebehov. Undervisningen i studierna blev mer individualiserad och effektiv, och eleverna upplevde kontroll då de fick feedback om sin progression i lärandet.

Effektstorleken från formativ bedömning baseras på två experimentella studier i läsning samt en experimentell och en quasi-experimentell studie i matematik (för elever i årskurs 2-8 respektive 3-10). En medel-effektstorlek på 0,3 uppmättes efter 7 månader i den experimentella studien i matematik. I den studien ingick 1880 elever. Medel-effektstorleken 0,4 uppmättes efter 18 veckor i den quasi-experimentella studien med 2202 elever. Medeleffektstorleken från båda studierna i matematik var 0,4. För att beräkna kostnadseffektiviteten för de båda typerna av intervention (reducerad klasstorlek och formativ bedömning) använder sig Yeh av effektstorlekar på lärande och kostnadsberäkningar från tidigare studier. I de tidigare studierna mättes elevernas lärande utifrån standardiserade test. Effektstorleken divideras med kostnaden per elev. Effektstorleken från reducerad klasstorlek baseras på två metaanalyser och tre delstatsbaserade studier. Reduktion av klasstorlek visade sig vara 124 gånger mindre kostnadseffektivt än införandet av system som innebär snabb formativ bedömning av elevernas framsteg i matte och läsning 2-5 gånger per vecka.

Reformer i USA för att öka elevers lärande har inneburit att lärarna förväntas använda sig mer av formativ bedömning. I två amerikanska storskaliga studier infördes användning av datorprogram för att underlätta mer frekvent formativ bedömning Båda studierna mätte effekten av användningen av datorprogrammet på elevernas lärande. I den ena studien användes programmet TEFE (Technology-enhanced formative evaluation) i grundskolor (Burns, Klingbeil, and Ysseldyke, 2010). Programmet TEFE är ett ramverk för ett datadrivet beslutsfattande om elevens fortsatta arbete. Datorprogrammet genererar anpassade tester, rättar elevernas svar, fastställa lämpliga undervisningsmål baserat på elevernas testsvar och övervakar elevernas framsteg. TEFE förväntas vara ett komplement till ordinarie undervisning som ska underlätta för läraren. Läraren får stöd för att

ta beslut om den fortsatta helklassundervisningen och för individualiserad undervisning. Lärarens arbetsbörda underlättas generellt så att läraren kan använda sin tid på bästa sätt. Datorprogrammet generar arbetsblad till eleverna. Programmet rättar arbetsbladen och meddelar såväl eleven som läraren när eleven är redo för ett test. Även testet genereras och rättas av programmet. Efter testet får eleven vid behov extra övningsuppgifter eller så går eleven vidare mot ett nytt lärandemål (med andra övningsuppgifter). Läraren får information om hur det går för alla elever i klassen och för varje enskild elev för att kunna ta beslut om undervisningen.

Studien var quasi-experimentell utifrån att det slumpmässiga urvalet byggde på huruvida skolorna själva köpt in datorprogrammet TEFE. Fyra delstater valdes strategiskt ut. Totalt ingick 360 skolor med i genomsnitt 523 elever i de tidigare skolåren. För varje delstat valdes slumpvis 30 skolor till varje grupp: (A) skolor som använt TEFE i 1-4 år och 11 månader, (B) skolor som använt TEFE i mer än 5 år och (C) skolor som inte använt TEFE. Varje delstats slutliga examens-test utgjorde data om elevernas prestationer i matematik och läsning.

I skolor som använde programmet TEFE presterade en högre procentuell andel av eleverna på en hög prestationsnivå i matematik än eleverna i kontrollgruppen, (även sedan resultaten från lästest använts som kovariat). Skillnaden var statistiskt signifikant, $p < 0.001$. Skolor som använt TEFE i mer än 5 år hade något högre andel högpresterande elever än eleverna i de skolor som använt TEFE i 1-4 år och 11 månader. Eftersom resultaten stödjs av tidigare utvärderingsstudier av TEFE föreslår författarna att detta är preliminära belägg för att användning av TEFE ökar den procentuella andelen högpresterande elever på delstatsprov. Forskarnas försiktighet beror på att studien var en quasi-experimentell studie som inte kunde kontrollera för alla variabler. Ett annat resultat var att prestationerna inte skiljde sig mellan vita och icke-vita minoritetsgrupper i interventionsgruppen, men väl i kontrollgruppen.

I den andra amerikanska datorbaserade interventionen definieras formativ bedömning som en mekanism för att individualisera och möta elevernas behov (Koedinger, Mclaughlin, & Heffernan, 2010). Författarna menar att formativ bedömning är speciellt viktigt för elever i behov av stöd och att det kan vara extra svårt för läraren att kunna möta dessa elevers behov. I denna studie användes datorprogrammet ASSIST. Författarna menar att mer frekvent testande (benchmark testing) i sig inte kan få någon effekt på elevers lärande om inte informationen från testen används. En snabb och tillförlitlig återkoppling med ASSIST möjliggör två sätt att använda resultat från testerna: (I) Läraranpassning: Läraren använder information från datorprogrammet för att ta beslut om strategier i klassrummet, till exempel genom att genomföra en helklassgenomgång om många elever har samma svårighet. ASSIST ger läraren detaljerade rapporter om elevers svårigheter. Det är inte bara information baserat på rätt/fel utan även om vilket stöd eleven behöver, hur många försök eleven gjort, hintar som eleven efterfrågat, responstid och antalet övningstillfällen. (II) Elevers lärande: Eleverna får ”samtidig intelligent handledning”, vilket innebär att datorprogrammet på ett avancerat sätt ger direkt korrigerande feedback och leder eleven vidare med hintar och stödjande frågor. Tanken är att undervisningen ska bli mer kostnadseffektiv då datorprogrammet kan korta ner tiden för uppföljning, dvs innan eleven kan ta nästa steg i sitt lärande, samt att läraren ska frigöras från administrativa uppgifter för att få tid för andra aktiviteter som genomförande av samarbetsaktiviteter eller att hjälpa elever i svårigheter.

I studien ingick 1240 åk 7 elever i tre undersökningsskolor och en jämförelseskola. Andelen elever i behov av särskilt stöd var 21% (255 elever). Ett förtestresultat och ett eftertestresultat användes för att jämföra undersökningsgrupperna med jämförelsegruppen. Förtestet var det slutliga testet i slutet av åk 6 och eftertestet var det slutliga testet efter åk 7. Eftertestet var svårare, men innehöll endast ett begränsat antal uppgifter på den nivå som undervisas i åk 7. Ett 2xANCOVA användes för att testa skillnaderna mellan grupperna på eftertestet efter att ha beaktat förtestresultaten (som kovariat).

Eleverna i de tre skolorna där ASSIST användes presterade signifikant bättre ($p < 0{,}001$, $d=0{,}23$) än eleverna på den skola där ASSIST inte användes. Skillnaden var speciellt stor för elever i matematiksvårigheter ($p<0{,}001$; $d=0{,}5$). Ytterligare resultat visade att ju mer eleverna använde ASSIST ju större blev fördelarna för deras lärande. Forskarna påvisade även att lärarna använde sig av informationen om elevernas lärande för att ta beslut om undervisningen av klassen. Detta undersöktes inte direkt utan slutsatsen drogs baserat på att lärandet hos elever som själva inte använde ASSIST ändå påverkades positivt av att ASSIST användes

I en amerikansk pilotstudie undersöktes effekten av en intervention, som de kallar Fusion, på elever i åk 1 som riskerar att hamna i svårigheter i matematikundervisningen (Clarke, Doabler, Cary, Kosty, Baker, Fien, &

Smolkowski, 2014). Interventionen Fusion innehåller två huvudkomponenter: (I) förståelse för innehåll och färdigheter gällande heltal och (II) explicita och systematiska principer för undervisningsdesign. Dessa principer bygger på (a) att när ett nytt och komplicerat matematiskt innehåll introduceras sker det genom tydliga förklaringar och demonstrationer, (b) att ge frekventa möjligheter för studenter att praktisera viktigt matematiskt innehåll, och (c) att i rätt tid ge återkoppling för att bekräfta korrekta elevsvar och rätta till eventuella missuppfattningar.

Alla åk 1 elever testades en månad innan interventionen. De 10 elever med lägsta poäng på testet valdes ut som riskelever (dock valdes inte elever ut som inte kunde identifiera eller skriva siffror eller de som inte kunde tillräckligt bra engelska). Totalt 89 riskelever från 10 olika skolor fördelades slumpvis till interventionsgrupp (44 elever) eller kontrollgrupp (45 elever). Interventionen riktades enbart mot riskeleverna och undervisningen genomfördes av nio interventionister. Undervisningens ca 60 lektioner genomfördes i liten grupp med ca 5 elever. En 30 minuters-lektion per dag, tre gånger per vecka, genomfördes över en period på ca 20 veckor (i praktiken blev det 29 veckor och 51 lektioner i snitt). Interventionisterna hade fått en utbildning (workshops) vid två tillfällen à 3 timmar. Den första workshopen gavs en månad innan interventionen och den andra efter ca 5 veckor efter det att interventionen påbörjats. Den matematik som tränades och sedan mättes jämförs med ”number sense”, vilket innebär både begreppsförståelse och automatisering av procedurkunskaper. Eleverna i studien gjorde tre olika test. Samma tre test användes både som förtest två veckor innan interventionen och som eftertest inom två veckor efter interventionen. Undervisningssituationen observerades vi tre tillfällen, i början, mitten och slutet av interventionen för att mäta fideliteten, dvs i vilken utsträckning interventionistens undervisning ligger i linje med interventionens underliggande principer (the fidelity of the intervention), och för en övergripande utvärdering av undervisningskvaliteten. Fideliteten till interventionen mättes av observatörer som markerade på en 10-gradig skala graden av fideliteten på de tre första genomförda huvudaktiviteterna i en lektion. Observatören gav även en sammanfattande bedömning av implementeringen på en 7-gradig skala. Till observationen av den övergripande undervisningskvaliteten användes ett holistiskt bedömningsverktyg (Ratings of Classroom Management and Instructional Support) för att mäta interaktionen mellan elev och lärare. Verktyget har 14 delar som bedöms utifrån en 4-gradig skala. Flertalet modeller användes i analysen för att undersöka samband mellan lärandeökningar (skillnader mellan för- och eftertest), fidelitet och kvalitet i undervisningen.

Ingen skillnad fanns mellan interventionsgruppen och kontrollgruppen på förtestet. På ett av eftertesten om begreppskunskaper visade resultatet att interventionsgruppens elever hade signifikant ($p = 0,015$) bättre kunskaper. Effektstorleken var stor (Hedges’s $g = 0,8$). Ingen signifikant skillnad fanns på automatisering av procedurkunskaper och ett annat test av begreppskunskaper, men små effektstorlekar fanns (vilket författarna menar ger ett signifikant övergripande resultat enligt What Works Clearinghouse). Inget signifikant samband kunde uppmätas mellan fideliteten till interventionen i undervisningssituationen och elevernas resultat. Här påpekar författarna att detta kan bero på den höga graden av fidelitet, som i snitt var högre än 9 på den 10-gradiga skalan.

I en stor experimentell studie av 85 åk 6 lärare och 4091 elever i USA undersöktes om lärares användning av ett bedömningsmaterial (Powersource) tillsammans med en fortbildning hade effekt på elevernas lärande i matematik (Phelan, et al., 2011). Förutom effekten av lärarnas användning av Powersource undersöktes om effekten varierade utifrån uppgiftstyp eller vilket skoldistrikt skolan tillhörde. Materialet förväntades hjälpa läraren att få bra information om hur undervisningen fungerat, samt att stödja lärarens uppföljande undervisning utifrån denna information. Med uppföljning avses anpassad undervisning och att läraren ger eleverna feedback som stödjer deras djupare förståelse för grundläggande idéer i matematiken. Lärare använder materialets kontrollmaterial ”checks for understandings” och stödmaterialet för uppföljning i fem olika steg i undervisningen, för varje delområde (i studien fyra delområden i algebra, se a, b, c, d nedan):

1. Kontroll och analys av elevernas förförståelse av grundläggande idéer och tillämpningar i matematiken (här algebra).
2. Undervisning. Vid behov ingår uppföljning av identifierade missuppfattningar från steg 1.
3. En andra kontroll och analys av elevernas förståelse och (vid behov) uppföljning.

4. Undervisning som innehåller presentation av huvudidéernas användning i problemlösning, symbolisk representation och räkneuppgifter.
5. En tredje kontroll och analys av elevernas förståelse och (vid behov) uppföljning.

Studien hade en kluster-randomiserad kontrollerad design med slumpvis fördelning av lärare inom och mellan skolor. I tre av skoldistrikten fördelades lärarna inom varje skola till försöksgrupp eller kontrollgrupp. I fyra skoldistrikt fördelades skolor slumpvis till att vara försöksskola eller kontrollskola. Interventionen genomfördes i genomsnitt över 8 lektioner. Fyra olika moduler i algebra behandlades under denna period: (a) rationella tals likvärdighet, (b) egenskaper hos aritmetik, (c) principer för att lösa linjära ekvationer, och (d) tillämpning av grundläggande principer på dessa områden till andra viktiga områden av matematiken, såsom geometri och sannolikhet. Fortbildningen, på totalt 9 timmar, innehöll ett inledande möte om målet med forskningsstudien och de teoretiska grunderna för projektet. Det ansågs viktigt att lärarna skulle få förståelse för de matematiska grunderna och den formativa bedömningsprocessen. Lärarna fick även råd om hur de skulle analysera elevresultat och anpassa undervisningen. I fortbildningen ingick också tre uppföljande möten efter var och en av de tre första undervisningsmodulerna. Vid dessa tillfällen fick lärarna titta på elevresultat från distriktet, vilket gav dem möjligheter att se mönster i elevsvar och vilka de vanligaste felsvaren och begreppsmissuppfattningarna var. De flesta av kontrollgruppens lärare deltog i en alternativ fortbildning med liknande struktur som den för försöksgruppen, men med ett annat innehåll. Kontrollgruppens lärare hade inte tillgång till något material. Båda gruppernas elever testades (algebrakunskaper) i ett förtest i början av läsåret och ett eftertest i slutet av läsåret. De två testen var olika. Uppgifterna testades med en Rasch modell. Faktoranalys genomfördes för att mäta variation förklarad av huvudkonstrukt. I analysen användes en två-nivå Hierarkisk modell, med läraren som analysenhet. Interventionen innehöll tre delar: material för bedömning och stöd för uppföljande undervisning, samt fortbildning.

Studien visar på att materialet, en så pass kort fortbildning och kort intervention inte var tillräckligt stöd till lärarna för att ge signifikant effekt på elevernas lärande totalt sett. Däremot uppmättes en signifikant effekt ($p<0,01$) för elever med högre resultat på förtestet. För de högpresterande eleverna var effektstorleken upp till 0,5 (pooled) standardavvikelser. Effekten var också större på svårare (algebra)uppgifter, speciellt uppgifter om distributiva lagen. Försöksgruppens elever överträffade, med statistisk signifikans ($p=0,002$) kontrollgruppens elever på uppgifter inom delområdet ”egenskaper hos aritmetik”. Denna effekt ökade ju bättre resultat på förtestet eleverna hade ($p<0,001$).

Implementeringen av undervisningsmetoden ”Technology-Enhanced Formative Assessment” (TEFA) och användandet av ”Classroom Response Systems” (CRSs) undersöktes i en amerikansk studie (Lee, Feldman, & Beatty, 2012). Syftet med forskningsstudien var att (1) undersöka vilka faktorer som hindrar implementering av TEFA-pedagogik; (2) identifiera vilka av dessa faktorer som är vanligast och starkast; och (3) att undersöka den generella processen för hur lärare initialt implementerar TEFA. Det som benämns formativ bedömning i studien är lärarens analyser av vad eleverna tänker och orsaken till att de tänker så, samt lärarens påföljande revision av lektioner utifrån analysen. TEFA baseras på fyra principer: frågedriven undervisning, dialogisk diskurs, formativ bedömning och kommunikation på metanivå. I undervisningsmetodens frågecykel ingår även att läraren ställer frågor, eleverna tänker själva eller med kamrater och sedan svarar på frågorna via CRS. Efter det att eleverna svarat följer en helklassdiskussion som läraren leder och summerar innan en ny frågecykel av TEFA påbörjas. Huvuddelarna i frågecykeln benämns: ”pose, think/talk, answer, histogram, share, och discuss”, vilket ger en bild av genomförandet av en frågecykel.

Studien i denna artikel är en delstudie i ett större forskningsprojekt. De 38 högstadielärarna i matematik och naturvetenskap från sex olika skolor deltog frivilligt i studien. Lärarna var indelade i fyra kohorter som påbörjade studien vid tre olika tidpunkter med ett års mellanrum. Resultaten i denna delstudie baseras på de första två kohorterna och första två åren. De hade då genomfört två respektive ett år av den treårig fortbildning som ingick. Fortbildningen började med tre eller fyra dagars (sommar)workshop där CRS-teknik och TEFA-pedagogik introducerades. Efterföljande veckovisa, och senare varannan veckas, eftermiddagsträffar på skolan under första året fokuserade på att hjälpa deltagarna att utveckla färdigheter för de fyra principerna i TEFA. Under det andra och tredje året genomfördes aktionsforskningsträffar var tredje till fjärde vecka. På dessa träffar valde lärarna aspekter av TEFA som de vill fokusera och utveckla genom experimenterande, reflektion

och diskussion. Vid alla typer av träffar delgav lärarna varandra erfarenheter och reflekterade över användandet av CRS och TEFA. Projektgruppen erbjöd även teknisk support.

Den huvudsakliga datainsamlingen för att identifiera svårigheter och barriärer i implementeringsprocessen gjordes via sju månatliga enkätundersökningar från oktober till maj. Analysen av data från enkätsvaren jämfördes med data från fältanteckningar från observationer, intervjuer, dagbok/loggbok, och transskript från fortbildningsträffar som samlats in i flera delstudier i det stora forskningsprojektet. På detta sätt triangulerades data från flera datakällor. För att undersöka hinder för implementeringen (Syfte 1) användes en analys av de öppna frågorna på enkäten baserad på ”Grounded theory”. Denna analys gjordes i fyra steg med programmen HyperResearch och NVivo. För att identifiera de vanligaste och starkaste faktorerna (Syfte 2) gjordes en matris för varje lärare med kodfrekvens i ena riktningen och tid/antal i den andra riktningen. Frågor som besvarats på en Likert-skala analyserades och grafer skapades för varje individuell lärare för att visa styrkan av varje faktor över tid. För att klassificera typer av lärarerfarenheter av implementeringen (Syfte 3) samlades en lärares alla grafer och sedan gjordes en kvalitativ helhetsbeskrivning av lärarens implementering av TEFA och mötta problem. Profilerna diskuterades med forskare och triangulerades med övrig data och analyser samt erfarenheter från fortbildningen. För att förstå mönstret i gruppen placerades enkätfrågor som besvarats på Likert-skalor in i de kategorier som identifierats från öppna frågor (Cronbach’s alpha beräknades för mått på internal consistency). Utifrån denna skapades en graf som visar vanliga svårigheter och mönster för utveckling.

Resultaten baseras därmed på lärarnas uttalanden och rapporter om implementeringen av TEFA och användning av CRS. Tio framträdande hinder identifierades, varav tidsbegränsningar och framställande av frågor rapporterades som mest problematiska. De övriga åtta identifierade hindren är: svårigheter med hårdvara och mjukvara; använda teknik; integrera TEFA i kursplanen; underlätta klassrumsdiskussioner; eleverna; kontextuella faktorer; sätt att vara lärare; och formativ bedömning. I resultaten finns fem vignetter, karaktärsteckningar, som illustration för olika implementeringsätt av TEFA och hur olika typer av faktorer påverkade lärare. Fyra generella kategorier beskriver lärares initiala implementering av TEFA, samt vanliga svårigheter och mönster i implementeringen: misslyckande i initial teknikintegrering, ingen pedagogisk implementering, passiv pedagogisk implementering, eller aktiv pedagogisk implementering.

I de mer detaljerade resultaten delas de försvårande faktorerna upp i yttre och inre faktorer som båda har delmängder. Yttre faktorer delas in i Typ-I med direkt koppling till TEFA, och Typ-0 som är faktorer utan koppling till TEFA. Exempel på faktorer av Typ I (med koppling till TEFA) är: *svårigheter med hårdvara och mjukvara* (tekniska fel, resurstillgång, begränsningar av programvara, och otillräcklig teknisk support); *tidsbrist* (planeringstid, tid i klass); påtryckning från läroplanen (vidd kontra djupet av täckning i ett område, standardiserade tester); och elevernas beteenden, attityder och förmågor. Typ-0 (ej specifikt TEFA): distraktioner och krav på lärarnas personliga liv; motstridiga prioriteringar i skolan; händelser och avbrott i skolan, och faktorer som påverkar elevers och lärares humör, energi och fokus, som väder, årstid, tid på dygnet; och fortskridande utveckling av terminen. Det fanns även en Typ-II kategori, som inte fokuserades i denna artikel, som handlade om faktorer som pedagogisk och administrativ policy eller sociala och kulturella frågor. De inre faktorerna delades in i Typ-I med koppling till lärarens kunskaper och förmågor och Typ II med koppling till lärares djupare sätt att vara lärare. Inre Typ-I (kunskaper och förmågor) är: användning av teknik; utveckla TEFA frågor; integrera TEFA i kursplanen; underlätta klassdiskussion; och formativ bedömning. Typ II (lärares djupare sätt att vara lärare) är: tveksamhet och osäkerhet om värdet av TEFA; motstånd mot förändringar i undervisningen; perspektiv, föreställningar och filosofier om undervisning, lärande och studenter; personlighet, egenskaper och självförtroende; tillfredsställelse med TEFA; och personlig biografi, etc.

Feedback

I avsnittet presenteras först en kort allmän definition av feedback. Därefter besvaras forskningsfrågorna. Dessa svar baseras på artiklarna som inkluderats i forskningsöversikten och som sammanfattas i det påföljande avsnittet.

Definition av feedback

Feedback är en viktig komponent och strategi i formativ bedömning. Det är en komponent som ofta undersökts i de studier som påvisar effekt av formativ bedömning på elevers lärande. Feedback från läraren förväntas initiera kognitiva och/eller känslomässiga processer hos eleverna, som i sin tur påverkar elevernas motivation och prestationer (Rakoczy, 2008). I Hattie och Timperleys (2007) forskningsöversikt sammanfattas resultat från ett stort antal sådana studier. Det finns olika definitioner av begreppet feedback i litteraturen. En vanlig definition beskriver feedback som information från en person eller ett material angående någons prestation eller förståelse (Hattie & Timperley, 2007; Rakoczy et al., 2008). Feedback är då alltid en respons på en föregående handling. Det går att skilja på olika typer av feedback, samt olika former för hur feedback ges och tas emot. Exempel på olika typer av feedback är omedelbar eller fördröjd feedback, feedback som talar om huruvida något är korrekt eller ej och feedback som ger ytterligare information. Formen för hur feedback ges kan variera på olika sätt. Den kan ges t ex muntligt eller skriftligt, av en vuxen, kamrat, bok, av sig själv eller av ett datorprogram. Feedback kan också tas emot på olika sätt, t ex individuellt eller i grupp. Studier har visat att både typ av feedback och hur den ges kan påverka lärandet (se t ex. Rochelle, 2010).

Forskningsfråga 1: Vilket samband finns mellan feedback och elevers prestationer i matematik?

Av de sju studierna var det tre stycken som jämförde effekten av feedback jämfört med ingen feedback, eller med mindre frekvent feedback, på elevers matematikprestationer. Resultaten varierade mellan studierna. I två av dessa studier gjordes en jämförelse med kontrollgrupper (Brosvic et al., 2006; Labuhn et al., 2010), och i den tredje studien jämfördes effekten av feedback som funktion av hur frekvent eleverna fick feedback av olika typer (Rakoczy, 2008). I studien av Brosvic var elevernas ökning av matematikprestationer statistiskt signifikant större för gruppen elever som fick omedelbar feedback än för grupperna som fick fördröjd feedback eller ingen feedback alls. Någon sådan skillnad fanns inte i studierna av Labuhn et al. (2010) och Rakoczy et al. (2008). Här kan dock konstateras att den typ av feedback som gavs i Labuhnstudien var sådan som i de andra studierna visat sig vara mindre effektiv för lärande. Feedbacken gavs inte på varje uppgift utan enda som ett sammanlagt resultat på en uppsättning uppgifter. Det torde därför vara svårt att kognitivt processa denna information på ett sätt som är användbart för lärande. Tiden för feedbacken var också väldigt kort. Eleverna fick feedback vid ett enda tillfälle och fick då 10 minuter att processa feedbacken. De effekter feedback kan ha på elevers motivation kan möjligen vara begränsade med en så kort intervention. I studien av Rakoczy et al. (2008) togs ingen hänsyn till kvaliteten i feedbacken som gavs och varje elev kan ha fått relativt lite feedback eftersom studien genomfördes i ordinarie klassrumsmiljö under endast under tre lektioner där läraren inte fått några instruktioner om givande av feedback. Eftersom en elev inte är ensam i klassen, och lärarna inte fått stöd i utvecklandet av tekniker för hur individuell feedback kan ges till många elever under en lektion, kan många elever ha fått väldigt lite feedback vilket torde påverka effekten på elevernas lärande.

Studien av Gursel et al. (2006) mätte om den enda eleven i studien lärde sig mer än den kunde förut med ny information från feedbacken. Detta var fallet men resultatet säger inte så mycket om storleken på lärandet eller viktiga medierande processer.

Forskningsfråga 2. Vilka egenskaper hos feedback är centrala i detta samband?

Fyra av studierna jämförde effekten av olika typer av feedback, och på vilket sätt den gavs. Dessa faktorer visade sig påverka effekten av feedback på elevernas prestationer. Studien av Cates (2005) på fyra elever visade att när det gäller feedback på uppgiftsnivå (återkoppling som anger om ett svar är rätt eller fel) kan det för vissa elever vara fördelaktigare med att få feedbacken från en dator jämfört med från en klasskamrat, medan det för andra elever kan vara tvärtom. I studien av Brosvic et al. (2006) på ett större urval elever syntes ingen statistiskt signifikant skillnad mellan de grupper av elever som fick datorgenererad feedback på uppgiftsnivå och de som fick samma typ av feedback från läraren. Däremot lärde eleverna sig mer när de fick feedback (på uppgiftsnivå) efter varje uppgift jämfört med när de fick feedback efter de svarat på alla uppgifter under ett pass (Brosvic et al., 2006). Rakoczy et al. (2013) jämförde effekterna av skriftlig processorienterad feedback

och feedback i form av betyg på en uppsättning uppgifter. Ingen direkt statistiskt signifikant skillnad mellan grupperna fanns, men däremot fanns en statistiskt signifikant indirekt effekt på elevernas kunskapsutveckling av den processorienterade feedbacken jämfört med feedback i form av betyg) via elevernas upplevda användningsmöjligheter av feedbacken. Även elevers lärandemålsorientering medierade denna process. I studien av Roschelle et al. (2010) lärde sig eleverna som fått feedback på smågruppsnivå (och skulle processa den tillsammans) mer än de som fick feedback på individnivå.

Forskningsfråga 3: Vilka egenskaper hos fortbildningar är centrala för att lärare ska vilja och kunna använda feedback på ett framgångsrikt sätt i sin undervisning?

Ingen av studierna innehåller någon form av lärarfortbildning där effekten av fortbildning på lärares feedback och elevers prestationer undersökts.

Forskningsfråga 4: Vilken typ av forskning bedrivs för att besvara ovanstående forskningsfrågor?

Studierna som analyserats har en jämn spridning av åldern för de studerade eleverna från årskurs 3 upp till årskurs 9. Däremot är den geografiska spridningen begränsad då sex av de sju studierna kommer från Tyskland eller USA. Den matematik som ingått i de test som använts för mätning av elevernas prestationer har varit faktakunskaper och procedurhantering men även begreppsförståelse. Den typ av feedback som ingått i studierna är feedback på uppgiftsnivå (rätt eller fel), processorienterad/informativ feedback, poäng, betyg och extra information. De olika sätt som feedbacken har givits på är omedelbar och fördröjd feedback, feedback om en viss uppgiftslösning eller totalt på en uppsättning lösningar, feedback given av en dator eller kamrat eller lärare, feedback för processande i grupp eller individuellt, och feedback given i form av grafer. Naturligt, med tanke på forskningsfrågorna som styrt urvalet av publikationer, så innehåller de flesta studier (fem av sju) ett stort antal elever i urvalet. Statistiska forskningsmetoder är därför vanligt förekommande i dessa studier. Interventionen i studierna i form av feedback varierar från 10 minuter vid ett tillfälle till uppemot 20 sessioner under sex veckor. Fem av de sju studierna är från kontrollerade förhållanden där eleverna fått feedback individuellt eller i smågrupper. Endast två av studierna är från normala klassrumsmiljöer. Ingen av studierna innehåller någon form av lärarfortbildning där effekten av fortbildning på lärares feedback och elevers prestationer undersökts.

Sammanfattning av artiklarna

Effekten på elevers lärande i matematik av olika typer och former för feedback undersöktes i en amerikansk interventionsstudie som inkluderade 40 åk 3 elever i matematiksvårigheter och 40 normalpresterande åk 3 elever (Brosvic et al. 2006). Det matematiska innehållet eleverna skulle lära sig i interventionen var talfakta för talen 0-9 i de fyra räknesätten (t ex 2+3). Lärandet skedde under trettio sessioner som var och en bestod av en genomgång av ämnesinnehållet och ett antal uppgifter eleverna skulle lösa. Uppgiftsformatet var flervalsfrågor med fyra svarsalternativ. Resultaten på de fem första och de fem sista av de 30 sessionerna under 5-6 veckor utgjorde data för jämförelse av prestation innan och efter interventionen (för varje räknesätt). De 20 mittersta sessionerna, interventionssessionerna, kunde ges en eller två gånger per dag.

Båda grupperna elever (elever med matematiksvårigheter respektive normalpresterande i matematik) fördelades slumpvis till ett av flera försöksförhållanden. Eleverna erhöll antingen (i) ingen feedback, (ii) fördröjd feedback eller (iii) omedelbar feedback när de löste uppgifter under interventionssessionerna. Den omedelbara feedbacken gavs antingen (iiia) av läraren eller (iiib) av datorprogrammet IF AT (Immediate Feedback Assessment Technique) som bygger på principen ”svara till du svarat rätt”. Den feedback som gavs var att uppgiften lösts korrekt eller att den inte lösts korrekt. Om den inte lösts korrekt uppmanades eleverna att göra ett nytt försök. Den omedelbara feedbacken gavs efter varje uppgift. Den fördröjda feedbacken gavs efter eleven svarat på alla uppgifter under en session, och eleven hade då 30 minuter avsatt tid för att se över uppgifterna, feedbacken och svaren. Medelvärdet av antal korrekta svar beräknades för varje grupp (10 elever) och varje räknesätt före och efter interventionen.

Resultaten visar att eleverna lärde sig talfakta för talen 0-9 i de fyra räknesätten bättre då de fick omedelbar feedback, jämfört med både fördröjd feedback och ingen feedback alls. Studien visade en statistiskt signifikant effekt ($p<0,005$) av interventionen omedelbar feedback på matematikprestationer för både elever i matematiksvårigheter och normalpresterande elever. Effekten på elevernas resultat påverkades inte av om den omedelbara feedbacken kom från läraren eller från datorprogrammet. Några mått på effektstorlekarna som är möjliga att använda för jämförelse med andra studier (t ex Cohens' d) gavs inte i artikeln. I en uppföljningsstudie i samma artikel jämfördes utfallet från omedelbar feedback under tre försöksförhållanden (i) feedback från läraren, (ii) feedback från datorprogrammet IF AT, och (iii) feedback från läraren kombinerat med att när eleven får rätt svar efter att först ha svarat fel skulle eleven både muntligt och skriftligt uttrycka uppgiften och dess korrekta svar fem gånger. Ingen statistiskt signifikant skillnad i elevernas antal korrekta svar fanns mellan de tre försöksförhållandena. Uppföljningsstudier visade att eleverna i matematiksvårigheter bibehållit kunskaperna efter 6 månader (inga statistiska beräkningar visades för detta) men att de, till skillnad från de normalpresterande eleverna, inte hade utvecklat sina strategier för att använda dessa talfakta. Dessa elever fortsatte i stor utsträckning att använda sig av fingerräkning.

I en turkisk studie undersöktes om eleverna tillägnade sig kunskaper och färdigheter utöver målet med undervisningen då läraren vid korrekt elevsvar gav kort informativ feedback i form av information om kunskaper som inte direkt efterfrågades i uppgiften (Gursel, Tekin-Iftar, and Bozkurt 2006). I studien ingick en matematikelev tillsammans med fyra geografistuderande mellanstadieelever, alla med lindrig utvecklingsstörning. Matematikeleven skulle lära sig namn på nio matematematiska symboler t ex namnet kilo för symbolen kg. Eleverna deltog i träningssessioner där läraren använde sig av en undervisningsmetod med ”samtidig uppmaning”. I samband med denna undervisningsmetod gav läraren kort informativ feedback i det fall att eleven svarade rätt. Läraren kunde till exempel ge informationen ”It is used when measuring weight” i tillägg till att bekräfta att elevens svar ”kilo” var korrekt på frågan om namnet för förkortningen ”kg”. Lärandemålet var faktakunskaper i form av namnet på symbolen och denna ytterligare information innehöll ytterligare faktakunskaper. Läraren undervisade matematikeleverna i par vid tre tillfällen i veckan (för matematikeleven totalt 15 tillfällen, ca. 2 ½ timme).

Medelvärdet för andelen korrekta responser utgör underlaget för resultaten. Resultaten visar att eleven i viss utsträckning tillägnade sig den informativa feedbacken och därmed lärde sig fakta utöver det som var det ordinarie lärandemålet. Då undervisningen innehöll informativ feedback klarade eleven i slutsessionerna 67 % av uppgifterna som krävde den fakta som presenterats i den informativa feedbacken jämfört med 0 % under de första sessionerna. Ingen kontrollgrupp fanns med i studien, utan den förvärvade kunskapsnivån mättes gentemot den ursprungliga kunskapsnivån, och ingen signifikansnivå eller effektstorlek redovisas.

I en tysk interventionsstudie (Labuhn, Zimmerman, and Hasselhorn 2010) användes grafer för att visa eleverna deras resultat från flera på varandra följande tester i matematik. Syftet med studien var att undersöka ett antaget samband mellan elevers självregleringsprocesser och elevers lärande av ett nytt matematiskt innehåll. Antagandet var att feedback i form av denna typ av grafer kunde hjälpa eleverna att övervaka sin prestation, fokusera på utvecklingsprocessen och reflektera över orsakerna till utvecklingen. Utifrån teorier om självreglering anses elevens förmåga att avgöra vad han/hon kan eller inte kan (performance calibration) vara en viktig process i självreglering. Författarna menar också att lågpresterande elever anses vara en riskgrupp som tenderar att övervärdera sina kunskaper mer än andra elever så resultat från dessa elever utgjorde även en speciell delstudie.

Nittio åk 5 (10-åriga) elever blev slumpvis utvalda till en av nio typer av försöks- eller kontrollförhållanden. En variation mellan de olika grupperna, med vardera 10 elever, innebar att eleven antingen gavs feedback i form av en graf över enbart elevens egen serie av resultat på uppgifter (individuell feedback) eller om även andra elevers resultat markerades i grafen (jämförande feedback). En ytterligare variation var om eleverna gavs lärandemål innan deras uppgiftslösande (mastery learning standards) eller jämförandemål (social comparison standards). Interventionen genomfördes i en-till-en situation med varje elev (ej klassrumsmiljö). Fyra uppsättningar av vardera fem uppgifter gavs under en 40 minuter lång lektion. Det matematiska innehållet i interventionen var prioriteringsregler, ett för eleverna nytt område. Lektionen inkluderade även instruktioner samt förtest och eftertest, som båda innehöll åtta matematikuppgifter (som testade procedurkunskaper på prioriteringsregler) och ett antal enkätfrågor. Enkätfrågorna i för- och eftertestfaserna mätte elevernas

förutsägelse av förmåga, uppgiftslösningsförmåga och elevens uppfattning och värdering av sin problemlösning. De elever som övervärderat sin förutsägelse av resultat på förtestet klassificerades som riskgruppselever (dessa elever bekräftades också som lågpresterande på kunskapstestet i matematik).

I analysen användes ANOVA (3x3 analysis of variance) för att mäta effekten av feedback (individuell, jämförande eller ingen feedback) och självutvärderande mål (mastery learning, social comparison standards) på elevernas förmåga att avgöra vad han/hon kan eller inte kan (performance calibration) och prestation (uppgiftslösning). Ett ytterligare statistiskt test (Kruskal-Wallis, icke-parametriskt test) användes för mätning av effekten av feedback för riskgruppen (övervärderande elever).

Resultaten visar ingen signifikant effekt (på nivån $p<0,05$) av någon del av interventionen på elevernas prestationer i matematik. Värt att notera är att eleverna inte fick feedback på varje uppgift utan bara efter en uppsättning uppgifter. Den feedback de fick verkar också vara att de fick veta antal rätt på en uppsättning uppgifter. De fick alltså inte veta vilka uppgifter de löste korrekt och inte heller tips på vad i deras lösning som var bra och vad som kan förbättras. Hela interventionen var också mycket kort (40 minuter för allt följand tillsammans: instruktionsfas, förtest, undervisning och arbeta med övningsuppgifterna, feedback och eftertest). Eleverna satt också en och en, så t ex feedbacken i form av jämförelser med andra elevers resultat gavs inte i en klassrumssituation där andra elever kunde uppfatta elevens resultat.

En tysk interventionsstudie av 146 elever från 54 olika klasser i årskurs nio undersöktes hur skriftlig process-orienterad feedback på test påverkade elevernas prestationer och elevernas upplevelse av denna typ av feedback i jämförelse med feedback i form av betyg på test (feedback för social jämförelse) (Rakoczy et al. 2013). Eleverna valdes slumpvis till en av två försöksgrupper där en av de två typerna av feedback gavs i respektive grupp (skriftlig processorienterad eller social jämförelse/betyg). Urvalet var stratifierat då max 3 elever per klass skulle ingå i respektive grupp. Hela interventionen, som tog 100 minuter per elev, genomfördes av en utbildad försöksledare i en experimentell miljö med eleverna en och en. I början mättes elevernas lärandemålsorientering och intresse för det påföljande testet med enkäter. Detta följdes av ett förtest i matematik som inkluderade uppgifter om Pythagoras sats eller linjära ekvationer. Under en 25 minuters rast för eleven rättades testet och skriftlig feedback genererades. Under 10 minuter fick sedan eleven se sitt resultat och feedback som också förklarades för eleven. Deras upplevda kompetenssupport, användbarhet av feedbacken och intresse för det kommande eftertestet i matematik mättes sedan med enkäter. Interventionen avslutades med eftertestet. Data analyserades sedan med hjälp av stiganalys (path analysis).

Den positiva totala effekten på elevernas lärande från processorienterad feedback jämfört med betyg var inte signifikant. Resultaten visade dock att process-orienterad feedback hade statistiskt signifikant indirekt effekt på elevernas kunskapsutveckling, jmf med betyg, via upplevd användning av feedbacken. Lärandemålsorientering med inriktning mot att uppnå målen (Mastery approach goal orientation) medierade också påverkan av feedback på upplevd användbarhet av feedbacken.

I en amerikansk experimentell studie jämfördes effekten av feedback som gavs till en grupp elever (grupp-feedback) och feedback som gavs individuellt till elever (Roschelle et al. 2010). Författarna var intresserade av hur feedback kan processas i en social situation. Därför undersöktes både effekter på elevers beteenden och på prestationer i matematik. Tre skolor med vardera två åk 4 klasser ingick i studien. De totalt 173 eleverna undervisades som vanligt i bråk av sin lärare i den första halvan av arbetsperioden på totalt ca 24 dagar. Under denna del var läraren mer aktiv, dvs hade fler genomgångar. Inför den andra halvan av undervisningsperioden, då eleverna mer aktivt arbetade själva med innehållet bråk, fördelades eleverna slumpvis till en försöksgrupp eller kontrollgrupp. Förhållandena i de båda grupperna liknade varandra så mycket som möjligt förutom att i försöksgruppen gavs feedbacken till elevgrupper bestående av tre elever och i kontrollgruppen fick eleverna individuell feedback. I båda grupperna erhöll eleverna feedback via IT. I försöksgruppen fick eleverna feedback via en datorbaserad intervention kallad ”Technology-mediated, Peer-Assisted Learning”, (TechPALS) och i kontrollgruppen gavs feedback via ett datorprogram (iSucceed Math). Under den andra halvan av perioden (12 dagar), då experimentet genomfördes, undervisade lärarna slumpmässigt i någon av de båda grupperna dagligen. Lärarna bytte också undervisningsgrupp efter halva denna tid (6 dagar) för att lärareffekt skulle undvikas.

Interventionen (TechPALS) innehåller en databas med innehåll specifikt för bråk, en mjukvara anpassad till bärbara datorenheter och träningsmoduler för både lärare och elever. I båda grupperna löste eleverna

jämförbara uppgifter. I försöksgruppen valde eleven först individuellt en av flera givna lösningar. De tre eleverna behövde sedan som grupp komma överens om en gemensam lösning innan de kunde få feedback från datorprogrammet. Feedbacken gavs då på gruppnivå och talade om ifall de svarat rätt eller behövde försöka en gång till. I en annan typ av uppgift i försöksgruppen skulle eleverna para ihop olika representationsformer av bråk. De behövde utbyta representationer och finna icke-överensstämmande representationer med varandra för att lösa uppgiften. Ytterligare en typ av uppgift består av två steg: att finna ett unikt bråk inom ett givet intervall och att värdera ett bråk på en tallinje (0-1). Då eleverna funnit olika bråk inom intervallet tillåts de gå till steg två, där de sedan bedömer om allas unika bråk ryms inom det givna intervallet. Återigen ska eleverna nå konsensus inom gruppen och datorprogrammet talar om huruvida gruppens svar är korrekt eller ej. Försöksgruppens elever arbetade varje dag i grupper med tre elever. Gruppernas sammansättning slumpades varje dag av datorprogrammet. I ett tillhörande material tränades elevernas samarbetsförmåga, t ex uppmanas att fråga och förklara ”hur och varför” i stället för att bara fråga och ge svar. Kontrollgruppens elever fick via datorprogrammets tre olika delar demonstration, vägledd praktik och utmaning. I den vägledda praktiken fick eleverna individuell feedback som bekräftade om lösningen var korrekt eller om ett nytt försök av eleven behövdes. I den sista utmanande delen behövde eleverna lösa 80 % av problemuppgifterna innan de erhöll feedback. För att få gå över till den utmanande delen behöver eleverna ge fyra korrekta lösningar. I försöksgruppen stöttade läraren grupparbetet (t ex genom att hänvisa till samarbetsmaterialet) och i kontrollgruppen hjälpte läraren eleverna individuellt då de hade frågor.

Samtliga 173 elevers kunskapsutveckling mättes med identiska förtest och eftertest (29 uppgifter om bråkbegreppet och procedurberäkningar med bråk). Data om elevbeteenden och elevers uttalanden samlades in via klassrumsobservationer. Eftertestet gjordes 12 dagar efter förtestet. Mann-Whitney U-test användes för statistisk analys av påverkan av de två olika feedbackförhållandena på elevers beteenden gällande social bearbetning av feedback, och ANOVA för analyser av påverkan på elevernas kunskaper i bråk.

Resultaten visade att feedback på gruppnivå signifikant ökade elevernas sociala processande av feedback, t ex att förklara begrepp och färdigheter (bråk), ställa frågor och diskutera det man är oense om. Eleverna i denna grupp lärde sig också signifikant mer om bråk ($p<0.05$) än eleverna i kontrollgruppen. Effektstorleken varierade mellan $d = 0.14$ och $d = 0.44$ (medel 0.22) i de tre olika skolorna. I en av skolorna hade man redan innan försöket undervisat i bråk och eleverna på denna skola presterade bättre på förtestet och effektstorleken på kunskapsutvecklingen var mindre än för de andra två skolorna. Både elever med låga och höga resultat på förtestet drog fördel av att få gruppfeedback. Baserat på ovanstående resultat drar författarna slutsatsen att om eleverna på ett bra sätt fungerar som resurser för varandra ökar elevernas möjligheter att lära. Feedback som bekräftar när en lösning är korrekt var i detta fall tillräckligt.

I en amerikansk småskalig studie jämfördes effekten av kamratfeedback med effekten av feedback given från dator (Cates 2005). Fyra elever (8-11 år) i matematiksvårigheter delades slumpmässigt in i två par. Interventionen innebar att paren (drill-)tränade på additionsfakta för talen 0-12. Under varje träningssession på 3 minuter fick eleverna omedelbar feedback från endera parkamraten eller via dator. Dessa båda förhållanden liknade varandra strukturmässigt. Vid kamratfeedback (A) presenterade en elev så kallade flashcards med talfakta (t ex 3+4) för kamraten som svarade muntligt. Antalet korrekta svar räknades (utifrån att korrekta svar generade en egen hög med flashcards) och rapporterades till forskaren. Vid datorfeedback (B) presenterades på liknande sätt flashcards för varje elev. Här svarade eleven på tangentbord och datorprogrammet rapporterade antal rätt. Eleverna övade i 18 respektive 21 sessioner.

Elevernas resultat från varje session utgjorde data på elevernas kunskapsutveckling. Resultaten visade att det förhållande som var mest effektivt var olika för eleverna, dvs. feedback från kamrat var mer fördelaktigt för några elever, medan feedback via dator ökade lärandet för de andra eleverna. Författarna diskuterar möjliga bakomliggande förklaringar, t ex att åldern på eleverna kunde påverka vilken feedback som var mest gynnsam, utan att dra några slutsatser eftersom undersökningen var begränsad vad det gäller urvalet av elever.

Sambandet mellan tre olika typer av feedback och elevernas inre motivation och prestation i matematik undersöktes i en tysk studie (Rakoczy et al., 2008). De tre typerna av feedback var: (a) utvärderande positiv (svaret är rätt) eller (b) utvärderande negativ (svaret är fel), samt (c) informativ feedback (ger information som leder eleven vidare vid både rätt och fel svar). Forskarna ville även undersöka om sambandet mellan informativ feedback och elevernas motivation och prestation kunde förklaras av medierande processer, vilket detta fall var

kognitivt processande av feedbackinformation, och en upplevelse av positiva känslor under lärandet. I studien deltog 10 tyska klasser (åk 8 eller åk 9) med totalt 240 elever som arbetade med Pythagoras sats. Detta var inte en interventionsstudie, utan studien utfördes under tre ordinarie lektioner i elevernas klassrumsmiljö.

I början av skolåret mättes elevernas allmänna intresse för matematik i en enkät. Precis innan lektionerna mättes elevernas förförståelse om Pythagoras sats i ett förtest som testade viktiga förkunskaper för djup begreppsförståelse. Forskarna filmade sedan klasserna under de tre lektionerna, men bara den del av lektionerna där eleverna arbetade själva eller i grupp analyserades. Lektionerna innehöll en introduktion om Pythagoras sats där lärarna ombads gå igenom ett matematiskt bevis, men i övrigt erhöll inte lärarna några instruktioner för undervisningen. Under lektionen efter de tre videoinspelade lektionerna som ingick i studien tillfrågades studenterna i en enkät om deras kognitiva processande, känslomässiga erfarenheter och inre motivation under lektionerna. Inre motivation konceptualiserades som ”working interest” så måttet på inre motivation kontrollerades därför för ”dispositional interest” som mättes i början av året. Elevernas kunskaper om Pythagoras sats efter de tre lektionerna mättes med ett nytt test (eftertest) som innehöll uppgifter som krävde djup begreppsförståelse och enkla tillämpningsuppgifter. Data från de två testen analyserades baserat på Rasch modellen. För att undersöka vilken typ av feedback som eleverna erhöll analyserades den feedback som gavs under det att eleverna arbetade, ensam eller med kamrat, under dessa tre lektioner. Lärare och elevers feedback kodades sedan som någon av tre typerna (a, b eller c, se ovan) av feedback och en frekvens av förekomst beräknades inför kommande analyser. Regressionsanalyser genomfördes för att svara på forskningsfrågorna om sambandet mellan olika typer av feedback och elevers prestationer och motivation. I modellerna för motivation kontrollerades för eleverna intresse i början av året. I modellerna för elevernas kunskapsutveckling kontrollerades för elevernas förkunskaper om Pythagoras sats. Ytterligare analys gjordes för att kontrollera hur sambandet mellan informativ feedback och de båda beroende variablerna motivation och kunskaper om Pythagoras sats medierades av kognitiv aktivitet och känslomässiga upplevelser.

Resultaten visar att informativ feedback korrelerar med elevernas inre motivation ($p<0,01$). Även positiv utvärderande feedback (vid rätt svar) i klassrummet är kopplat till ökad inre motivation ($p<0,00$). En sådan korrelation finns dock inte för negativ utvärderande feedback (vid fel svar). En statistiskt signifikant indirekt relation fanns också mellan informativ feedback och inre motivation, medierat av både upplevelser av positiva känslor under lärandet och kognitivt processande av feedbacken. Informativ feedback verkar alltså fostra motivation via känslomässiga erfarenheter och kognitivt processande. Författarna tolkar det som att då eleverna får ledtrådar för fortsatt arbete upplever de lektionen positiv, de känner sig stimulerade och bearbetar innehållet i större utsträckning, vilket leder till högre motivation. Däremot verkade inte informationen i den informativa feedbacken i klassrummen ha varit tillräcklig för att fostra känslomässiga erfarenheter och djup bearbetning av innehåll på ett sätt som leder till bättre prestationer. Någon signifikant effekt på elevernas prestation kunde inte påvisas för någon av de tre olika typerna av feedback. Forskarna menar dock att det är svårare att uppnå signifikans på effekter av feedback vid klassrumsexperiment där feedback i den ordinarie undervisningen undersöks. I denna studie var läraren inte anvisad till att använda en viss typ av feedback, vilket är fallet i många andra kontrollerade interventionsstudier om feedback. I denna studie hade den informativa feedback som gavs stor variation. Det är inte säkert att den typ av informativ feedback som givit positiv påverkan på prestation i tidigare laboratoriestudier ingått i den feedback som identifierats i dessa klassrum.

Självreglerat lärande

I avsnittet presenteras först en kort allmän definition av självreglerat lärande. Därefter besvaras forskningsfrågorna. Dessa svar baseras på artiklarna som inkluderats i forskningsöversikten och som sammanfattas i det påföljande avsnittet.

Definition av självreglerat lärande

Den formativa klassrumspraktiken behöver inte utföras enbart av läraren, även eleven kan vara en aktiv deltagare i denna praktik. Läraren kan stödja eleverna att reglera sitt eget lärande. Sådant självreglerat lärande innebär att planera, övervaka, reflektera över och modifiera sitt lärande. Självreglering (SR), eller självreglerat

lärande (SRL), definieras och operationaliseras i forskningen på olika sätt. I artiklarna som inkluderats i denna forskningsöversikt refereras oftast till definitioner av Zimmerman (1999) eller Boekaerts (1999). I dessa breda definitioner av självreglering ingår kognition, metakognition och komponenter för motivation (vissa forskare använder metakognition som det övergripande begreppet, och andra särskiljer motivation från självreglering). Begreppet kognition syftar till de olika kognitiva strategier som används för att lära ett innehåll och lösa uppgifter. Med metakognition menas de metakognitiva strategier som används för att kontrollera och reglera kognition. Komponenter för motivation som är vanligt förekommande i definitionen för självreglerat lärande är hur eleven uppfattar sig själv i förhållande till en uppgift, till exempel tron på att man ska klara uppgiften (self-efficacy) samt intresse för eller affektiva reaktioner på uppgiften (Boekaerts, 1999).

Forskningsfråga 1. Vilket samband finns mellan självreglerat lärande och elevers prestationer i matematik?

Femton av de sexton studierna, som för denna forskningsöversikt identifierats i Web of Science, undersökte sambandet mellan elevers självreglering och deras prestationer i matematik, och i samtliga fall identifierades ett positivt och statistiskt signifikant samband. Sambandsmåtten i studierna varierade mellan effektstorlekar (t ex Cohen's d), korrelationskoefficienter, förklarad varians, ökning i procent och standardiserade effekter i multivariata strukturella ekvationsmodeller. I metanaalysen av Dignath & Büttner (2008) där interventionsstudier med fokus på SRL analyserats var medelvärdet för effektstorlekarna mycket höga 0,96 för årskurs 1-6, och lägre 0,23 för årskurserna 7-10. I de övriga studierna i urvalet där effektstorlekar angavs varierade dessa mellan 0,44 och 0,94. I korrelationsstudierna varierar korrelationerna mellan 0,4 och 0,8. Den enda studien med lägre korrelationer var studien av Marchis (2012) med korrelationer i intervallet 0,2-0,4. I de studier som angav den andel av variansen i prestationen som kan förklaras av SRL skilde det sig mellan studierna från 2,4% (Cleary & Gregory, 2013; Lazakidou & Retalis, 2010) till 41 % (Ocak &Yamaç, 2013).

Forskningsfråga 2. Vilka egenskaper hos självreglerat lärande är centrala i detta samband?

De olika studierna visar alla på ett samband mellan självreglerat lärande och elevers prestationer i matematik. Styrkan i sambandet varierar dock mellan studierna. Det kan finnas många anledningar till det, som till exempel olika ålder och nationalitet på eleverna i studien, eller den matematik, typ av test eller betyg som används som mått på elevernas matematikprestationer. Andra skäl kan vara de mått som används på elevernas självreglering (t ex elevers självskattning i elevenkäter eller lärares bedömningar), och metoder för beräkning av detta samband (t ex korrelationer för en grupp elever, eller interventionsstudier där signifikansanalys använts för att undersöka sambandet mellan interventionsgrupp och kontrollgrupp). Ytterligare skäl till den varierande storleken på sambandet kan vara att de olika studierna använt olika delar av självreglerat lärande som mått på detta konstrukt.

Det är svårt att dra säkra slutsatser om vilka aspekter av SRL, och interventioner som stöder elevers SRL, som är viktigast för ökade prestationer i matematik. I stället verkar flera olika aspekter av SRL vara centrala. I en metastudie av Dignath & Büttner (2008) analyserades 74 interventionsstudier som fokuserar på att stödja elevers självreglerade lärande i matematik, läsning eller skrivning. Av dessa innehöll 28 studier ämnet matematik. Deras analys av studierna som innehöll ämnet matematik visade att ju fler undervisnings- och träningssessioner som ingick i interventionen desto högre elevprestationer i matematik. De jämförde också effekterna av interventioner med olika teoretisk bakgrund och strategiinstruktion. För de tidigare skolåren fanns också ett positivt samband mellan interventioner som fokuserar på kognitiv strategiinstruktion jämfört med interventioner som fokuserar på metakognitiv reflektion. För senare grundskoleår var effekten av interventionerna på elevers matematikprestationer högre om interventionenernas teoretiska bakgrund fokuserade på motivationsteorier i jämförelse med metakognitiva och socialkognitiva teorier om självreglerat lärande. Metaanalysen inkluderade också andra ämnen och interventionernas effekt på andra konstrukt än prestationer. Exempelvis var interventionernas effekt på elevernas strategianvändning i deras självreglerade lärande större i den tidigare skolåren också om interventionerna baserades på en socialkognitiv teoretisk

bakgrund, och om de innehöll kognitiv och metakognitiv strategiinstruktion och metakognitiv reflektion. I detta sammanhang kan det vara värt att nämna att även sådana samband kan vara värda att beakta då inte alla interventioner som studerat det direkta sambandet mellan SRL-stödd undervisning och elevernas prestationer skett över en längre tidsperiod, och även effekter på elevers SRL-strategianvänding kan påverka prestationerna i ämnet på sikt. Det går inte heller att från övriga studier i denna rapport dra säkra slutsatser om exakt vilka aspekter av SRL som är viktigast för höga prestationer i matematik eller vilket SRL-innehåll interventionsstudier bör ha. En analys kan dock göras av de aspekter av SRL som ingår i de studier som visade sig ha högst samband med matematikprestationer. I korrelationsstudien av Throndsen (2011) fanns höga korrelationer mellanlågstadieelevernas matematiska prestationer och både deras metakognitiva kunskap och metakognitiva strategier. Höga korrelationer identifierades också i studien av Metallidou & Valchou (2010) mellan å ena sidan åk 5 elevers matematikprestationer och å andra sidan deras motivation att genomföra uppgifter, sätta mål, planering och självutvärderande förmågor, metakognitiva kunskaper (som att veta vad, hur och när och varför strategier ska användas) samt motivationsfaktorer. I interventionsstudien av Kramarski & Mizrachi (2006), med en effektstorlek på 0,94, fick åk 7 eleverna i interventionsgruppen träna sig på att använda sig av fyra olika kategorier av metakognitiva frågeställningar vid problemlösning.

Forskningsfråga 3: Vilka egenskaper hos fortbildningar är centrala för att lärare på ett framgångsrikt sätt ska använda självreglerat lärande i sin undervisning?

Både den review (Montauge, 2008) och den meta-analys (Dignath & Büttner, 2008) som ingår i denna kartläggning behandlar interventioner men inte i något fall beskrivs egenskaper hos fortbildning och fortbildningens effekter på elevers prestationer. I flera fall beskrivs istället de undervisningsmetoder som implementeras, antingen av lärare eller med hjälp av datorprogram. Samma sak gäller för de fem enskilda artiklar som innehåller interventionsstudier. Korrelationsstudierna innehåller inga fortbildningar. Dignath & Büttner (2008) konstaterar också denna brist på studier där fortbildningen beskrivs, och menar att sådana skulle vara värdefulla för att få en bild av svårigheterna med implementering av en undervisning som stöder elevernas självreglerade lärande. Detta är speciellt viktigt eftersom deras metaanalysstudie visar att de interventioner för att stödja elevers självreglerade lärande som sköttes av forskare hade större påverkan på elevernas prestationer än de som genomfördes av lärare. De efterlyser därför fler studier på effekter av lärarfortbildning på lärarnas implementering av en undervisning som stödjer elevers SRL, och denna undervisnings effekter på elevers prestationer.

Forskningsfråga 4: Vilken typ av forskning bedrivs för att besvara ovanstående forskningsfrågor?

Av studierna som sammanställts nedan finns en review (Montauge, 2008) och en meta-analys (Dignath & Büttner, 2008). En empirisk studie undersöker undervisningens innehåll av möjligheter för självreglering. Nio av de övriga totalt 13 empiriska studierna på elevnivå är genomförda i Europa, tre i USA och två från Asien (Israel). Hälften riktar sig mot elever i årskurs 1-6 och övriga till elever i årskurs 7-12. I två studier (varav en review) undersöks specifikt elever i särskilt undervisningsbehov i matematik och dessa studier tenderar att ha ett fåtal elever med i sina studier (t ex Ness, 2012). Det är stor variation på antal elever som ingår i studier som omfattar alla elever, från en klass på 24 elever (Throndsen, 2011) till 756 elever (Rosario, 2013). I de studier där antal elever är färre än 30 elever har man använt metoder som tänka-högt protokoll eller intervjuer för att undersöka elevers självreglering medan i studier med fler än 30 elever har man använt elevenkäter och/eller lärarenkäter som datainsamlingsmetod för att undersöka elevers självreglering. Den matematik som använts för mätning av elevernas prestationer har varit elevers betyg i matematik (7 studier) eller test inkluderande standarduppgifter, proceduruppgifter, men också uppgifter som testar resonemangs- och problemlösningsförmåga som kräver mer komplexa kunskaper (t ex Kramarski & Gutman, 2006). Den typ av självreglering som mätts har varit både för användning vid uppgiftslösning (t ex självreglering vid problemlösning) och som process vid lärande mer allmänt (t ex reglering av beteende under lärandeprocessen). Vanligtvis ingår alla tre komponenter i självreglering i studierna; kognition, metakognition och motivation och

de uttrycks i elevers kunskap om, och användning av, olika strategier i självregleringen. Åtta av studierna är korrelationsstudier där man mäter elevers självreglering och prestation i matematik. I två av dessa mäter man förändring över tid. Ingen av studierna innehåller någon form av lärarfortbildning där effekten av fortbildning på lärares undervisning i självreglerat lärande och elevers prestationer undersökts.

Sammanfattning av artiklarna

I en norsk korrelationsstudie över tid (1 år) undersöktes samband mellan elevers självregleringsförmåga och prestation i matematik (Throndsen, 2011). Självreglering definieras i studien som bestående av tre komponenter; kognition, metakognitiv kompetens och motivationsfaktorer. Förmågan att koordinera och interagera dessa komponenter är enligt författarna kännetecknande för god självreglering. Ett exempel på koordination av dessa komponenter är elevers strategianvändning och denna (typ av strategi) behandlas i studien som ett mått på elevers självreglering. Kognition definieras inte. Metakognition definieras som elevens medvetna och avsiktliga tankar om sina tankar. Elevers metakognitiva kompetens kategoriseras som metakognitiv kunskap och metakognitiv process vid problemlösning. De motiverande faktorer som tillskrivs självreglering i studien var hur eleverna attribuerar framgång eller misslyckande (till ansträngning eller inneboende förmåga) samt tron att klara av en uppgift (self-efficacy). Begreppet matematisk kompetens definieras inte i studien, men bedömning av elevers matematikkunskaper gjordes med två standardiserade test som prövade baskunskaper i matematik. Vilka dessa baskunskaper är beskrivs inte närmare. Efter varje matematikprov delades klassen in i tre lika stora grupper beroende på resultat; hög, medel och lågpresterande elever. Elevernas matematik prestationer bedömdes med matematikproven utifrån deras baskunskaper i matematik. Baserat på dessa resultat, för varje prov, kategoriserades eleverna som hög- medel- och lågpresterande i matematik för analysen.

Studien genomfördes under skolår 2 och skolår 3 (n=27, en klass). Kriterier för val av deltagande skola, eller deltagande klass, beskrivs inte och ingen intervention gjordes. Elevers strategianvändning (ex. fingerräkning, räkna i huvudet och automatiserad beräkning) identifierades med tänka-högt-metod när de löste additions- och subtraktionsuppgifter vid tre tillfällen (ht år 2, vt år 2, ht år 3). De använda strategierna poängsattes mellan 1-10, där 1 poäng gavs för den minst avancerade strategianvändningen. Metakognitiv kompetens och motivationsfaktorer bedömdes under intervju vid två tillfällen (ht år 2, ht år 3). De metakognitiva strategierna kvalitetsbedömdes med en poängskala 1-3. Matematisk kompetens bedömdes med matematikprov vid tre tillfällen (ht år 2, vt år 2, ht år 3). Eleverna fick dock två tester vt år 2 eftersom man behövde byta talområde då elevernas kunskapsnivå ökade.

Resultaten visar att korrelationen mellan strategianvändning och prestation i matematik var hög. För de fyra matematiktesten och strategianvänding var korrelationen 0.60-0,67, alla med $p < 0,01$. Den mest använda strategin var för den högpresterande gruppen automatiserad beräkning (mellan 49-72 % för de 4 test-tillfällena). De lågpresterande eleverna använde oftast fingerräkning och räkna i huvudet-strategier (60-78% för de 4 testtillfällena).

Metakognitiv kompetens (metakognitiv kunskap och metakognitiv process) korrelerande också positivt med matematisk prestation både ht år 2 och ht år 3. Den metakognitiva kompetensen varierade mellan hög- medel- och låg-presterande grupperna redan i år 2 men det var först i år tre som skillnaderna blev signifikant ($r=0,87$, $p< 0,001$). De högpresterande elevernas metakognitiva kunskaper (vad de vet, och hur de tänker, om när och hur de ska använda strategier) utgjorde den största metakognitiva skillnaden mellan grupperna.

Motiverande komponenter visar samband med matematikkunskaper först i skolår 3. Attribuering för framgång korrelerar med prestation i matematik ($r= -0,42 < 0,05$) där högpresterande elever i matematik i större utsträckning attribuerar framgång till ansträngning och lågpresterande elever oftare attribuerar misslyckande till inneboende förmåga. Samband finns också mellan tron på att kunna klara en uppgift (self-efficacy) och prestation i matematik i år 3 ($r=0,60$, $p < 0,01$) där de lågpresterande elevernas tro på att de ska kunna klara uppgiften är lägre än de högpresterandes.

Sammanfattningsvis visar studien att god prestation i matematik är relaterad till användning av avancerade matematikstrategier i aritmetik, metakognitiv kompetens för ämnet samt motivationskaraktäristika som attribuerar framgång till ansträngning och tro på att kunna lösa matematikuppgifter. Skillnad mellan hög- och

lågpresterande elevers självreglering skiljer sig i skolår 2 när det gäller användning av problemlösningsstrategier, och även för metakognitiv kompetens och motiverande faktorer i skolår 3.

Lazakidou & Retalis (2010) genomförde en interventionsstudie i Grekland med syfte att undersöka vilken effekt en datorbaserad undervisningsmetod hade på elevers självreglering vid problemlösning. I studien särskiljs självregleringsförmåga och metakognition. Med självreglering menas elevers användning av strategier som hjälper dem att tänka (t ex att sätta mål och strategisk planering), handla (t ex fokusera uppmärksamhet) och motiveras för uppgiften (t ex intresse och tro på att klara uppgiften) under problemlösningsprocessen. Metakognitiva strategier definieras som den typ av handlingar som aktiverar processer som övervakar och kontrollerar ens tankar under de tre självregleringsfaserna förtanke, utförande och självreflektion. Metakognitiva frågeställningar börjar ofta med ”hur” eller ”varför”. Författaren påpekar att gränsen mellan begreppen självreglering och metakognition är otydlig och i viss mån överlappande. De har i sin operationalisering av begreppen använt sig av metakognitiva och självreglerande frågor som ställs till eleverna under sex identifierade steg i problemlösningsprocessen (definiera problemet, konstruera en strategi, organisera information, fördela resurser, övervaka och utvärdera lösningen). I artikeln hänvisar författarna till en tabell där skillnaden mellan självreglerande och metakognitiva frågor exemplifieras. Matematisk kompetens definieras i studiens intervention som problemlösningsförmåga (genuin problemlösning) enligt en sexstegs-modell, men beskrivs inte närmare för det matematiktest som utgör mätning av elevernas matematiska prestation.

Deltagande klass, bestående av 24 elever i skolår 4, valdes då de tidigare deltagit i ett kollaborativt projekt och därför förväntade man att till en del kunna undvika samarbetsproblem i grupperna. Det fanns ingen kontrollgrupp. Interventionen varade under två månader och innehöll 10 lektioner. I interventionen integrerades teorier om problemlösning och självreglering till en undervisningsmetod i tre faser. De tre faserna bestod av observation (läraren löser problem inför eleverna), samarbete (elever löser problem tillsammans), semistrukturerad guidning (individuell problemlösning med hjälpfrågor) och slutligen individuell problemlösning utan stöd. Dessa faser ska enligt författarna motsvara de tre faser som behövs för att bli självreglerande och slutligen autonom lärande. Vid lektionernas genomförande i datalaboratoriet var lärare och en observatör närvarande. De datorprogram som eleverna använde var Synergo och Moodle learning mangement system. Läraren instruerades att stödja undervisningsmodellen vid genomförandet. Ingen kompetensutveckling för detta stöd beskrivs i artikeln.

Metakognitiva strategier bedömdes med tänka-högt-protokoll. Självreglerande strategier och autonomi bedömdes med hjälp av loggprotokoll från elevers samarbetslektioner vid datorer. Elevernas matematikkunskap mättes med för- och eftertest. Testen bestod av fyra olika typer av problemlösningar (numerisk, icke numerisk, ej användbar numerisk data samt ett problem med kodad numerisk data) som rättades av forskare. Elevers matematikkunskaper mättes även vid varje interventionslektion och dessa uppgifter rättades av läraren. De senare resultaten användes tillsammans med variansanalys för att kunna avgöra om resultaten från för- och eftertest var signifikanta. Ingen kontrollgrupp fanns med i studien, utan den förvärvade kunskapsnivån mättes gentemot den ursprungliga kunskapsnivån.

Elevernas högre matematikförmåga efter interventionen förklarades med att eleverna förbättrade sin självreglering. ANOVA visade statistiskt signifikant ökning av problemlösningsförmåga över de tio sessionerna, $p< 0,01$. Den procentuella ökningen från för- till eftertest var 2,4 %. Loggprotokollen visade signifikant skillnad för elevernas användning av självreglerande strategier (främst självbedömning och slutsatser om påföljande ansträngning, $p < 0,01$). Tänka-högt protokollen visade signifikant ökning för användning av metakognitiva strategier från ett medel på 2,6 strategier före interventionen till ett medel av 4,1 strategier efter interventionen.

Dresel & Haugwitz (2008) har i en tysk interventionsstudie undersökt vilken effekt datorgenererad attribuerande feedback (intern attribuering vid framgång; instabil attribuering vid misslyckande) hade på elevers motivationskaraktäristika (befrämjande attribuering, självuppfattning och hjälplöshet) och prestation i matematik. Ett antagande var att denna attribuerande feedback skulle öka elevers motivation och prestation i matematik. Ett annat antagande var att elever som fick stöd i form av metakognitiva kontrollfrågor (sätta mål, planera, övervaka och reflektera över problemlösningsprocessen) i ännu högre grad skulle förbättras vad gäller motivation och prestation i matematik. Metakognition beskrivs i studien som en del av självreglerat lärande

(SRL) och motivation beskrivs som relaterad till SRL. Matematisk kompetens definieras inte i studien, men mäts med test innehållande den matematik som förväntas undervisas i årskursen.

I interventionsprogrammet som pågick under en termin deltog 151 elever i nio klasser i årskurs sex. Klasserna fördelades slumpvis till tre behandlingsgrupper. Placebogruppen (PC) erhöll enbart feedback om korrekt eller inkorrekt svar, gruppen (AC) erhöll attribuerande feedback samt en grupp (AMC) erhöll attribuerande feedback och metakognitiva kontrollfrågor. De metakognitiva kontrollfrågorna besvarades på ett arbetsblad. Dataprogrammet MathWarp tillhandahöll problemuppgifter (typ av problem beskrivs inte), samt gav feedback (korrekt/inkorrekt feedback till alla grupper samt attribuerande feedback till AC och AMC). Genom att använda datorgenererad feedback ansåg författarna att de dels fick kontroll över den feedback som gavs, dels fick eleverna frekvent feedback.

Data om elevers motivation och metakognitiva strategier samlades in via enkäter vid tre tillfällen, före och direkt efter interventionen samt 5 månader efter interventionen avslutats. Matematisk kunskap mättes vid samma tidpunkter. Analysen omfattade elever som arbetat med programmet under 5–9 lektioner.

Resultaten visar att elever i de grupper som fått attribuerade feedback (grupperna AC och AMC) hade signifikant bättre resultat på eftertest i matematik än placebogruppen PC ($d=0,54$, $p<0,01$). AMC-gruppen som även fått metakognitiva kontrollfrågor hade signifikant bättre matematikresultat än AC-gruppen ($d=0,52$, $p<0,05$). Vid det uppföljande provet 5 månader senare kvarstod dock inte skillnaderna mellan grupperna. Användning av metakognitiva strategier ökade signifikant ($d=0,32$, $p<0,05$) för elever i AMC-gruppen, de använde fler metakognitiva strategier (de satte mer omfattande mål, planerade och övervakade sin lärandeprocess) än elever i PC och AC grupperna från för- till eftertest. Fem månader senare var skillnaden fortfarande signifikant ($d=0,28$, $p<0,05$). Elever i AC/AMC-grupperna hade även signifikant högre motivation för matematikarbete, en skillnad som kvarstod vid uppföljningstest ($d = 0,48$, $p<0,01$vid eftertest, $d=0,35$ $p<0,05$ vid uppföljningstest). Det fanns ingen signifikant skillnad mellan AC och AMC grupperna avseende motivation.

I en israelisk interventionsstudie (Kramarski & Gutman, 2006) undersöktes effekterna av en datorbaserad undervisningsmetod för självreglerat lärande på elevers matematikprestationer, samt sambandet mellan egenskaper hos elevers självreglering och prestation i matematik. Elevernas matematikförmåga testades före och efter interventionen för tre olika typer av uppgifter; standarduppgifter, mer komplexa uppgifter som krävde problemlösning, samt uppgifter som testade elevernas förmåga att förklara matematiska resonemang. Självreglering mättes i studien med enkäter, före och efter interventionen, genom elevers självbedömning av två självreglerande processer; användning av problemlösningsstrategier och självövervakande strategier. Problemlösningsstrategierna är områdeskonkretiserade som t ex ”innan jag börjar lösa ett problem försöker jag förstå tillgängliga data i uppgiften; när jag ska beräkna lutningen börjar jag med att referera till specifika punkter på grafen”. Självövervakningsstrategier handlar om kontroll av lösningsprocessen (t ex ”efter jag har löst uppgiften kontrollerar jag om lösningen är rimlig”).

Interventionen integrerades som en 5 veckor lång del av ett treårigt statligt fortbildningsprojekt (fortbildningen beskrivs närmare i artikeln av Kramarski & Revach (2009)). Elever i åk 9 (n=65) blev slumpmässigt utvalda till två elevgrupper som undervisades om linjära funktioner. Den ena gruppen (EL+IMP) fick tillgång till E-lärandemiljö (datorbaserad undervisningsmiljö) och stöd för självreglerat lärande (SRL) genom IMPROVE. Metoden IMPROVE tillhandahåller fyra kategorier av metakognitiva frågeställningar (vad handlar problemet om, konstruera samband mellan tidigare och ny kunskap, vilka strategier bör användas och varför, reflektion om processen och lösningen). Den andra gruppen (EL) hade samma tillgång till E-lärandemiljö men fick inget stöd för SRL. Interventionens generella lektioner bestod av två stycken 90-minuters lektioner i veckan och behandlade linjära funktioner. Eleverna löste uppgifter och lärare och elever diskuterade metoder för att matematiskt förklara sin lösning. Två 90-minuterslektioner varje vecka bestod av undervisning i E-lärandemiljö. I E-lärandemiljön uppmanades eleverna att lösa problem i par och skicka sina lösningar till läraren. EL-eleverna fick enbart feedback i form av om svaren var korrekta eller inte. Gruppen EL+IMP fick även feedback enligt undervisningsmetoden IMPROVE som är utarbetad för att stödja SRL. EL+IMP-gruppens lärare modellerade också användning av de metakognitiva frågeställningarna och eleverna ombads ge skriftliga svar på frågeställningarna.

Studien visar att det inte fanns någon statistiskt signifikant skillnad i resultat på förtestets proceduruppgifter mellan de två elevgrupperna. I eftertestet presterade dock elevgruppen som fått stöd för självreglering (EL+IMP) statistiskt signifikant bättre än gruppen som inte fått SRL-stöd ($p<0,001$, Effektstorlek = 0,44 (typ av effektstorlek är inte angivet)). En statistiskt signifikant skillnad i prestationer mellan förtest och eftertest när det gäller uppgifter som kräver transfer respektive förklaring av matematiska resonemang hittades också ($p<0,001$). Skillnad mellan förtest och eftertest för gruppen som fått stöd för självreglering hade effektstorlek på 1,75 för transferuppgifterna och 2,24 för uppgifter som krävde matematiska förklaringar medan motsvarande effektstorlekar var 0,77 och 0,46 för den andra gruppen. Eleverna i den SRL-stödda gruppen ökade också sina matematiska argument mer än den andra gruppen ($p<0,05$), men däremot inte de procedurella argumenten.

Resultaten visar också att elevernas självövervakningsstrategier ökade signifikant i den SRL-stödda gruppen, jämfört med den icke SRL-stödda gruppen ($p<0,05$), medan ingen signifikant skillnad fanns när det gäller den andra aspekten av SRL som var problemlösningsstrategier. Signifikanta korrelationer fanns också både i den totala gruppen och i den SRL-stödda gruppen mellan självövervakningsstrategier och prestationer på uppgifter som kräver transfer och matematiska förklaringar. I EL-gruppen fanns motsvarande korrelation endast för uppgifter som krävde matematiska förklaringar. Inga statistiskt signifikanta korrelationer fanns mellan SRL och lösning av proceduruppgifter.

I ytterligare en israelisk interventionsstudie (Kramarski & Mizrachi, 2006) undersöktes effekterna av olika undervisningsmetoder för självreglerat lärande samt samband mellan elevers självreglering och prestation i matematik (problemlösning). Interventionen integrerades som en 4 veckor lång del av ett treårigt statligt fortbildningsprojekt (fortbildningen beskrivs närmare i artikeln Kramarski & Revach, 2009). Studien genomfördes med elever i årskurs 7 (4 klasser, n= 86) i den israeliska grundskolan. Klasserna valdes slumpmässigt till en av fyra undervisningsmetoder. Tre metoder kombinerades till de fyra undervisningsgrupperna; problemlösning under diskussioner online, problemlösning med diskussioner i par, och metakognitiv guidning under problemlösning. De fyra undervisningsgrupperna fick samma generella instruktioner för problemlösning och lektionerna hade samma form; läraren introducerade problemlösningsstrategier genom modellering (t ex användning av tabeller och algebraiska formler) och summerade i slutet av lektionen. Undervisningsgruppen Ftf (face-to-face) praktiserade dessa problemlösningsstrategier under pararbete. Metoden Online diskussioner (Online) bestod av att elever löste problem i par och i datormiljö 45 minuter per vecka. Deras lösningar var tillgängliga för andra par att kommentera. Läraren uppmuntrade eleverna att reflektera på lösningsprocessen och be andra par om hjälp om problem uppstod. Metoden metakognitiv guidning (+meta) baseras på IMPROVE, utvecklad av Kramarski & Mevarechs (2003). Metoden IMPROVE tillhandahåller fyra kategorier av metakognitiva frågeställningar (vad handlar problemet om, konstruera samband mellan tidigare och ny kunskap, vilka strategier bör användas och varför, reflektion om processen och lösningen). Denna metakognitiva guidning applicerades både i E-lärandemiljö (datorbaserad undervisningsmiljö) och i vanlig undervisningsmiljö. I datormiljön presenterades dessa frågor som hypertext och i den vanliga undervisningsmiljön fick eleverna ett blad med metakognitiva frågor att besvara under problemlösningsprocessen. De ovan beskrivna tre undervisningsmetoderna kombinerades till fyra undervisningsgrupper; diskussioner online med metakognitiv guidning (Online +meta), diskussioner Online utan metakognitiv guidning (Online), ansikte-mot-ansikte diskussioner med metakognitiv guidning (Ftf +meta) samt ansikte-mot-ansikte diskussioner utan metakognitiv guidning (Ftf).

Elevernas självreglering bedömdes via elevenkäter före och efter interventionen innehållande påståenden som behandlade strategianvändning före, under och efter problemlösning (exempel på påståenden: före jag löser ett problem försöker jag formulera det med mina egna ord, när jag löser problem så brukar jag organisera data i en tabell, och när jag har löst ett problem så kontrollerar jag att svaret är logiskt). De två undervisningsgrupper som fått undervisning i metoden Online fick ytterligare en enkät som behandlade den specifika självregleringen för miljön (exempel på påståenden: när jag löser problem online så ökar mitt intresse för matematik, och problemlösning online uppmuntrade mig att förklara min lösning). Matematisk kompetens mäts i studien för tre olika matematiska förmågor; förmåga att lösa standarduppgifter (beräkning, problemlösning och resonemang), förmåga att lösa verklighetsanknutna uppgifter (förstå uppgiften, använda strategier, processa information, resonemang) samt olika typer av argument eleverna använde för sina lösningar (enbart svar, svar med beräkning, användande av vardaglig matematisk terminologi, användande av

matematiska uttrycksformer, och användande av matematiskt-logiska argument). Denna kunskap bedömdes vid för- och eftertest, där uppgifterna vid eftertesten anpassades till det aktuella arbetsområdet (algebra). Elever som fått undervisning i metoden Online fick ytterligare matematiktest som löstes Online (verklighetsnära problemlösning).

Resultaten visar att metakognitiv guidning gav avsevärda effekter på elevernas matematikprestaioner, och att E-lärandemiljön också var fördelaktig. En jämförelse mellan de två grupperna som fått undervisning online visade att gruppen som fått metakognitiv guidning presterade statistiskt signifikant bättre ($p<0,01$) på verklighetsanknutna uppgifter lösta online (förståelse för uppgiften, användning av matematiska strategier, processa information, matematiskt resonemang). Gruppen Online+meta använde också formella (90 %) och logiska (75 %) argument i högre grad än Onlinegruppen (30 % respektive 35 %). Även gruppen som fått metakognitiv guidning men inte online (Ftf+meta), presterade statistiskt signifikant bättre än gruppen som bedrivit diskussioner online men inte fått metakognitiv guidning. Skillnaden var signifikant ($p<0,001$) när det gällde standarduppgifter, verklighetsanknutna uppgifter och ett flertal aspekter av självreglering (effektstorlek 0,94). I en jämförelse mellan grupperna som fått metakognitiv guidning presterade gruppen i E-lärande miljön bättre när det gällde standarduppgifter, verklighetsanknutna uppgifter och ett flertal aspekter av självreglering (effektstorlek för Online+meta jämfört med Ftf+meta var (0,70, $p<0,001$).

Kistner et al. (2010) utförde en korrelationsstudie över tid i syfte att få insikter i om och hur lärare stödjer självreglerat lärande (SRL) i vanliga tyska klassrum och vilka konsekvenser det får för elevers prestationer i matematik. Datainsamling utgjordes av videofilmade lektioner där tjugo matematiklärare genomförde tre lektioner om Pytagoras sats. Videofilmerna utgjorde den tyska delmängden av en tidigare tysk-schweizisk studie. De tjugo klasserna som filmades var jämnt fördelade mellan de två spåren ”gymnasium” och ”real-schule” (elever i ålder 15 år). Hur lärare valdes ut beskrivs inte. Ingen intervention föregick datainsamlingen. Analys av videoinspelningar av lektioner utgjorde data för variablerna lärares stöd för elevers självreglering samt lärmiljöers stöd för elevers självreglering.

Självreglering definieras i studien som en lärandes kompetens att autonomt planera, genomföra och utvärdera lärandeprocesser vilket innefattar kontinuerliga beslut om kognitiva, motiverande och beteendemässiga aspekter. Med observationsprotokollet ATES (assessing how teachers enhance self-regulated learning) mättes hur ofta och vilka kognitiva (organisering, bearbetning, problemlösande strategier), metakognitiva (strategier för att planera, övervaka och utvärdera lärandeprocessen) och motiverande strategier (resurshantering, attribuering och beteendekontroll) som undervisades, samt om dessa var implicita (läraren uppmanar elever att använda en strategi) eller explicita (läraren anger att eleverna ska använda en strategi). I observationsprotokollet mättes även i vilken grad (4-gradig skala) lärare skapar lärmiljöer som stödjer elevers självreglering. Lärmiljöer bedömdes ur fyra aspekter; stöd för social interaktion mellan elever; stöd för aktivt konstruerande av lärande; stöd för elevers själv-bestämmande (self-direction) samt stöd för lärande i autentiska situationer för att skapa transfer. Ingen observation av elevers SRL gjordes.

Matematisk kompetens definieras inte i studien men elevers matematiska prestation mäts för generell matematisk förmåga, kunskaper relaterade till Pytagoras sats samt förståelse för bevis. För elevers förmåga att prestera i matematik gjordes totalt sex test: tre förtest, (generell matematisk förmåga, förståelse för bevis och kunskaper relaterade till Pytagoras sats), ett eftertest direkt efter arbetsområdet avslutats (förståelse för Pytagoras stats), samt ett uppföljande test (förståelse för bevis) och slutligen ett avslutande test (generell matematisk förmåga).

Resultaten från de observerade videolektionerna visar att explicit undervisning av strategier ($r=0,52$, $p=0,1$) samt antalet strategier som behandlar organisation ($r=0,47$, $p=0,02$) korrelerar positivt med förmågan förståelse för bevis från för- till eftertest. För övriga kognitiva och metakognitiva strategier (3 st.) fanns inga signifikanta korrelationer. Inte heller de motiverande strategier som lärare använde visade resultat i form av elevers lärande på test. Författarna påpekar att lärare använder lite tid till att undervisa om strategier och när de gör det är det oftast implicit. Lärmiljöer med egenskaper som stödjer konstruktivistiskt lärande korrelerade signifikant för elevers förmåga att förstå Pytagoras sats ($r=0,71$, $p<0,01$) samt förståelse för bevis ($r=0,49$, $p=0,02$). Lärmiljöer med egenskaper som stödjer lärande i autentiska situationer (transfer) korrelerade signifikant med elevers resultat på förståelse för Pytagoras sats ($r= 0,56$, $p=0,01$). Inga skillnader kunde ses för lärmiljöer som stödde

kooperativt lärande eller elevers självbestämmande. Författarna påpekar dock att få lärare skapar en lärandemiljö som understödjer självreglering så underlaget för bedömning var troligen litet.

Cleary & Gregory (2013) har i en korrelationsstudie utförd i USA undersökt samband mellan elevers självreglering och prestation i matematik. Åttiosju elever i årskurs nio och en lärare i matematik (algebra) deltog i studien. I artikeln finns ingen information om hur elever valdes till studien. Ingen intervention gjordes.

Elevers självreglerade lärande (SRL) mättes dels genom en lärarenkät och dels genom elevenkäter. Lärarens bedömning av elevers SRL gjordes i en enkät, SRSI-TRS (self-regulation strategy inventory-teacher rating scale). Där bedömer läraren hur ofta eleverna engagerar sig i olika hjälpsökningsstrategier, deras användning av självmotiverande taktik och organisatoriska beteenden. Det andra måttet på SRL är elevers självskattningar i enkäter. I en av enkäterna (en del av enkäten SRSI-SR) skattar eleverna regleringsbeteenden som inte är anpassningsbara, MRB (maladaptive regulatory behaviors). Enkäten innehöll uppgifter av typen "Jag undviker att fråga om saker i klassen som jag inte förstår". Eleverna gör också självskattningar om strategianvändning när de ska göra prov (t ex strategier vid förberedelse inför prov, hjälpsökning under prov), intresse för matematik (t ex påståendet att lära matematik är mycket intressant) samt upplevelse av egen kapacitet (self-efficacy). För elevers prestation i matematik gjordes två mätningar. Först mättes elevers resultat på ett standardiserat matematiktest. Detta utgjorde mått på tidigare matematikkunskaper som en variabel som ingick i studien. Ett ordinarie lärarprov som gjordes två månader efter elevskattning utgjorde mått på elevers aktuella matematiska prestation.

Resultaten visar att elevernas självskattning av SRL som behandlade regleringsbeteenden som inte är anpassningsbara (MRB) korrelerade med deras matematikprestationer ($p<0,01$). Elevernas skattningar av tron att kunna klara en uppgift (self-efficacy), intresse och provstrategier var inte signifikant korrelerade med matematikprestationerna. Lärarnas skattningar av elevernas SRL korrelerade också signifikant med matematikprestationer ($p<0,01$). I studien beräknas semipartiella proportioner och dessa lärarbedömningar av SRL förklarar 9,4 % av variansen i matematikprestationer efter att ha kontrollerat för 28 % av variansen som kan attribueras till de övriga variablerna i studien. Motsvarande andel för elevbedömningarna var 2,4 %.

I en grekisk korrelationsstudie av Metallidou & Vlachou (2010) undersöks skillnader mellan grupper av elever med olika uppfattning av uppgifters värde (task-value beliefs). I studien redovisas även korrelation mellan elevers självreglerade lärande (SRL) och matematikprestation, ett delresultat som är relevant för denna kartläggning. Tretton lärare och 263 elever i årskurserna 5 (n=114) och 6 (n=149) ingick i studien. I artikeln redovisas inte uppgifter om hur deltagarna valdes.

I studien definieras självreglerat lärande som bestående av kognitiva, metakognitiva och motiverande processer, men man beskriver även SRL som en produkt. Matematisk kompetens definieras inte i studien. Både mått på matematiska prestationer och elevers SRL är bedömningar från lärarnas enkätsvar. Lärarna ombads bedöma elevernas prestation i matematik på en skala mellan 1–20 jämfört med sina klasskamrater. Lärarna bedömde hur ofta elever genom sitt beteende visade närvaro av SRL (genom motivation att genomföra uppgiften, sätta mål, planering samt självutvärderande förmågor). Elevers metakognitiva kunskaper (veta vad, hur, när eller varför strategier ska användas), bedömde lärarna på en 5-gradig skala (9 uppgifter).

Resultaten visar en statistiskt signifikant korrelation mellan elevers SRL och matematik för bägge variabler av SRL, d v s för kunskap om och användning av strategier (korr=0,78, $p \leq 0,001$) samt för självreglering av beteende (korr=0,83, $p \leq 0,001$).

Jones, Alexander & Estell (2010) har i en korrelationsstudie från USA undersökt om elevers (åk 9, n=111) upplevda självreglerade lärande (SRL) är på samma nivå som andra medlemmar av samma grupp. I studien undersöktes också om SRL är korrelerat till matematikprestationer och det är den delen av studien som är relevant för den här kartläggningen. I artikeln ges ingen information om hur urval av deltagare gått till. Studien fokuserar självreglerat lärande som förmåga att övervaka och själv hantera sin lärandeprocess. Matematisk kompetens definieras inte i studien. SRL mättes med elevers självskattningar i en enkät (Motivated strategies for learning questionnare, MSLQ), med sjugradig skala gällande metakognition, reglering av miljö, reglering av ansträngning, hjälpsökande och kamratlärande (använda kamrater för att lära) i förhållande till matematik (algebra). Elevers matematikprestationer mättes med elevernas senaste betyg. Betyget omvandlades till en 13-gradig skala.

Av de ingående variablerna så visade enbart reglering av ansträngning signifikant korrelation (stegvis regression) med matematikprestation ($\beta = 0{,}48$; $p < 0{,}001$). Aspekten kamratlärande har en negativ korrelation med matematikprestation vilket enligt artikelförfattarna kan spegla att duktiga elever inte behöver eller upplever vinsten av att fråga andra elever som inte kan lika mycket ($\beta = -0{,}29$; $p = 0{,}006$.)

Ness, B & Middleton, M. 2012, har i en fallstudie från USA beskrivit hur stöd för självreglerat lärande implementerats individuellt under matematiklektioner till en elev i behov av särskilt stöd (learning disabled, LD). Självreglerat lärande (SRL) definieras i studien som en cykel bestående av tre faser; planering, utförande och självvärdering. Elever som kan självreglera utvecklar kapacitet att göra anpassningar under denna cykel. Artikelns författare framhåller att elever med behov av särskilt stöd kan ha särskilda svårigheter att lära SRL (svårigheter med att planera, organisera, sätta rimliga mål, övervaka sin progression) vilket i sin tur kan leda till att de ägnar lång tid till uppgifter utan att ha förmåga att identifiera ett alltför stort kunskapsgap. Slutligen är risken att dessa elever i självvärderingsfasen attribuerar misslyckande eller framgång till den ansträngning som använts för att lösa uppgiften istället för att reflektera över vilka strategier som valts för uppgiften.

Deltagande i studien är en elev i årskurs 6, Chris, som identifierats som en elev med inlärningssvårigheter, med ett IQ på 119 men med matematik- och läsprestationer under förväntad nivå. Chris hade extra stöd i form av en specialpedagog under matematiklektioner. Den förberedande datainsamlingen gjordes dels för att få ett ingångsvärde att jämföra med utgångsvärden för elevens matematikkunskaper och SRL, men även för att designa interventionen. Elevens matematikkunskaper mättes med betyg i ämnet före och efter interventionen. Elevens självreglering mättes före interventionen genom lärarbedömning, observationsprotokoll (av forskare), elevintervju och elevenkät. Under interventionen mättes SRL genom observationer och efter interventionen bedömdes elevens SRL återigen. Det man mätte var elevens beteende bedömd av läraren (upplevd ansträngning) och av forskaren (hade penna, hade relevant bok framme, satt i bänken innan läraren började lektionen, ögonen fästa på läraren när lektionen började). Interventionen planerades med utgångpunkt från tidigare forskning (beskriven i artikeln) om stöd för SRL för elever i behov av särskilt stöd i de tre faserna i självregleringscykeln samt de ingående data som analyserats avseende elevens SRL. Speciallärarен satte tillsammans med eleven mål relaterade till de tre faserna i självregleringscykeln. En förkortning; MARS (material, förvänta lärandetrösklar, bli kvar i uppgiften, sitt i bänken och lyssna på läraren) användes för att hjälpa eleven att förbereda sig för lektionen och för att stödja självvärdering. Eleven instruerades i MARS (presenterat i ett arbetsblad) och fick stöd av specialpedagog i användandet av MARS före (förbereda) och i slutet (utvärdera) av matematiklektionerna.

Resultaten visar att elevens matematikbetyg ökade från C- till B-. Enligt speciallärarens bedömning förbättrades elevens beteende ett steg på den 8-gradiga skalan, från 7 till 6. För observerade beteenden under lektioner (observation av forskare), ökade medelvärdet för ”bli kvar i uppgiften” från 48 % till 67 % och ”förberedd för lärande” från 50 % till 87 %. För elevens engagemang för lärande (matematiklärarens bedömning) hade medelvärdet ökat från 25-50 % till 50-75 %. Föräldrar rapporterade att de upplevde en förbättring hos Chris vad gäller förståelsen för vikten av att planera att delta i klassen.

Ocak & Yamaç (2013) har i en turkisk studie undersökt förutsägande (predictor) och förklarande (explanatory) samband mellan elevers självreglerande strategier, motiverande föreställningar, attityd till matematik samt prestation i matematik. Deltagande i studien var 204 elever i årskurs 5 från olika skolor i en turkisk stad. Ingen information om hur eleverna valdes ges i artikeln.

Självreglering mäts i studien genom en elevenkät MSLQ (motivated strategies for learning questionnare) som mäter motivationskaraktäristika (tron på att klara av en uppgift, testoro, inre målorientering, yttre målorientering, kontroll av föreställningar om lärande, och uppgifters värde) och lärandestrategier (kognitiva och metakognitiva strategier samt strategier för att hantera resurser). Elevers attityd till matematik mättes genom en annan enkät (MTO). Elevers medelbetyg (grade point average, GPA) i matematik utgjorde variabeln matematisk prestation. Två stycken SEM (structured equation modeling) användes för att avgöra hur oberoende variabler förutsäger beroende variabler.

Resultaten av den första modellen visar att tro på att man ska klara av en uppgift (self-efficacy) påverkar matematisk prestation positivt och testoro påverkar negativt. Tillsammans förklarade dessa två variabler 41 % av matematisk prestation. Modellen var enligt beräkning av fit index ”excellent”. I studien redovisas också resultat om motiverande (oro vid test, inre målorientering, uppgifters värde och tron på att klara av en uppgift)

och självregleringsfaktorer (metakognitiv självreglering) som förutsäger (predict) elevers attityd till matematik. Den andra modellen visade hur motiverande variabler förutsäger (predict) kognitiva strategier och metakognitiva självreglerande strategier positivt. Även denna modell hade ”excellent fit”.

I en kroatisk korrelationsstudie (Buric & Soric, 2012) var syftet att undersöka relationen mellan elevers känslor av hopp och hopplöshet vid matematiska test (n=365, elevers ålder 16 år). Upprinnelse till dessa känslor anses vara kognitiv kontroll, värdeuppfattning samt strategier av vilja. Därför ville man i studien även undersöka effekten på elevers prestation i matematik av delar av dessa självregleringsområden. De fyra självregleringsfaktorer som ingick i studien var upplevd kognitiv kontroll över prestationer, upplevelse av värdet av lärande, förstärkning av self-efficacy och stressreducerande handlingar, samt negativt baserade incitament. Dessa självregleringsfaktorer mättes genom elevernas självskattning i enkäter. Elevers prestation i matematik mättes med tidigare och nuvarande betyg, rapporterade av eleverna i enkäten (5-gradig skala).

I analysen användes strukturell ekvationsmodellering (SEM). Resultaten visar signifikant korrelation mellan alla dessa fyra självregleringsvariabler och matematisk prestation (både tidigare och nuvarande betyg) i modellen där alla variabler inklusive elevers känslor av hopp och hopplöshet ingick. De standardiserade effekterna var 0,4 för de tre första faktorerna och -0,8 för negativt baserade incitament. Alla effekter var statistiskt signifikanta ($p<0,05$ för förstärkning av self-efficacy och $p<0,001$ för de andra tre faktorerna).

Marchis (2012) har i en korrelationsstudie från Rumänien undersökt elevers självreglerade lärande (SRL) och prestation i matematik. I studien ingick 335 slumpvis utvalda elever från nordvästra Rumänien i årskurserna 9-11 (åldrarna 14/15 – 17/18 år). De kategorier av självreglering som undersöks beskrivs som tillhörande någon av de tre faserna som utgör självreglerat lärande; företänksamhetsfas, genomförandefas och reflektionsfas. I företänksamhetsfasen ingår kategorierna intresse, sätta mål samt uppgiftsanalys. I genomförandefasen ingår kategorierna hjälpsökande och självkontroll. Reflektionsfasens kategori är självbedömning. Elevers självreglerade lärande (SRL) mättes i en enkät (21 uppgifter om SRL, 5-gradig skala) vid ett tillfälle. Matematisk prestation mättes med senaste betyg vilket eleverna rapporterade i enkäten.

Resultaten visar korrelation mellan prestation i matematik och flera av självregleringsförmågorna. Korrelationskoefficienterna för frågorna inom följande kategorier ligger i intervallet 0,23-0,35: elevers intresse för matematik (värdet av matematik i vardagslivet eller framtiden), uppgiftsanalys (förstå problemet, identifiera givna data och använda dem samt förståelse för de krav problemet ställer), självkontrollerande förmåga (övervaka och kontrollera problemlösningsprocessen), förståelse av uppgifters svårighetsgrad (veta om man kan lösa ett problem eller vad man saknar för att kunna göra det). Inga signifikansnivåer (p-värden) redovisas i studien. I studien presenteras även resultat om elevers grad av självreglering (i procent) inom olika kategorier, men dessa resultat är utanför denna kartläggnings intresse.

I en portugisisk studie (Rosario, Nunez, Valle, Gonzalez-Pienda & Lourenco, 2013) var syftet att öka förståelsen för hur självreglerat lärande medierar relationen mellan motiverande faktorer och prestation i matematik. I studien ingick 756 elever i åldrarna 12-19 år (medel 13,9 år) från 4 slumpvis utvalda skolor.

Datainsamlingen gjordes med elevenkäter, elevloggbok samt elevers betyg. Den del av elevenkäten där eleverna skattade SRL-strategier delades in i tre faktorer; planering, utförande och utvärdering. Ett exempel på en fråga ur faktorn SRL-strategier för utvärdering är ”Jag jämför det betyg jag får med det mål jag satt för ämnet”. En annan del av enkäten behandlade elevers tro på att de klarar sin egen självreglering med hjälp av SRL-strategier när de lär sig matematik (self-efficacy för SRL). Ett exempel på en fråga från faktorn är ”Hur väl kan du använda strategier för att memorera ett område på ett heltäckande sätt?”. Ytterligare en del i enkäten behandlade upplevd användbarhet av SRL-strategier. Ett exempel på en fråga från faktorn användbarhet för SRL-strategier är ”Hur användbart tycker du att det är att anteckna för att sedan utveckla och använda dem för att lära dig materialet mer detaljerat?”. I datainsamlingen ingick även en elevloggbok över hur mycket tid som använts för att studera matematik under en vecka. Matematisk prestation mättes utifrån betyg på ett skriftligt prov under terminen samt elevens betyg på hemläxors utförande. I artikeln finns ingen uppgift om hur betygsinformationen samlades in.För dataanalys användes metoden strukturerad ekvationsmodellering, SEM (structured equation modelling). Variablerna som ingick i SEM-modellen grupperades i tre kategorier: förutsägande (årskurs, studerandetid samt antal skolår eleven gått om årskurser), process (elevskattad användning av SRL-strategier, upplevd self-efficacy för SRL, upplevd användbarhet av SRL), produkt (prestation i matematik). I analysen användes matematisk prestation som latent variabel.

Resultaten från SEM indikerar att självreglerat lärande är positivt och signifikant relaterat till prestation i matematik (relationskoefficient 0,43, p= 0,000). Den rapporterade användningen av SRL-strategier förklarar 16 % av prestationen i matematik, inkluderande de indirekta effekterna av de övriga variablerna genom SRL. Användningen av SRL-strategier är kopplad till elevers tro på att de själv kan reglera sitt lärande i matematik genom att använda SRL-strategier, self-efficacy för SRL, (relationskoefficienten=0,44; p=0,000), och upplevd användbarhet av SRL-strategier (relationskoefficienten=0,42; p=0.000). Elevers rapporterade användning av SRL-strategier minskar dock under skolåren 7-9.

Montauge (2008) har i sin artikel sammanställt resultaten från interventionsstudier som undersöker effekten av undervisning om kognitiva strategier på elevers (elever i behov av särskilt stöd) lärande i matematik (problemlösning). Syftet var att identifiera viktiga principer för framgångsrik undervisning om kognitiva strategier för elever i behov av särskilt stöd. Vid urval av ingående artiklar så sattes krav på artiklarnas tydlighet i urval, metod, analys samt tydlig beskrivning av interventionen. Interventionen skulle också kunna definieras som evidensbaserad eller lovande. Enligt artikelförfattaren har alla artiklar som beskriver studier av enstaka elever redovisat positiva resultat. I Montagues sammanställning av artiklarna (både enskilda studier och reviews) framgår i många fall vad eleverna blivit bättre på men mer sällan i vilken grad.

Resultaten presenteras i form av åtta principer samt beskrivning av hur undervisning om kognitiva strategier bör organiseras. De åtta principerna är:

- Var sparsam med instruktioner
- Var medveten om elevers olikheter
- Undervisningsstrategier fungerar olika i olika sammanhang
- Jämförbara prestationer innebär inte att elever är i jämförbara processer eller använder jämförbara strategier
- Vid undervisning om strategier måste hänsyn tas till elevens kunskapsbas och kapacitet
- Användning av samma typ av strategi innebär inte nödvändigtvis att prestationen blir densamma (elever i behov av särskilt stöd kan behöva ytterligare stöd)
- Bra undervisningsapproach för elever i behov av särskilt stöd är inte nödvändigtvis bra undervisningsapproach för normalpresterande elever och vice versa.
- Strategier som elever behärskar kan de inte nödvändigtvis generalisera

Slutsatsen är också att kognitiv strategiundervisning bör ske i smågrupper, intensivt och tidsbegränsat.

Dignath, & Büttner (2008) har i sin metaanalys analyserat 74 interventionsstudier (utförda i årskurs 1-6 och 7-10) som fokuserar på att stödja elevers självreglerade lärande i matematik, läsning eller skrivning. Av dessa innehöll 28 studier ämnet matematik. De undersökte vilken effekt interventionerna haft på elevers prestation i ämnena, strategianvändning och motivation för studier. För att inkluderas i metaanalysen skulle studierna behandla självreglering, intervention med träning för elever (inte elever i behov av särskilt stöd) och studiernas resultat skulle behandla elevprestation, självregleringsstrategier, känslor eller motivation. Publikationerna som inkluderades skulle vara publicerade mellan 1992 och 2008. Den använda metod och de presenterade resultaten i studierna skulle vara av sådan kvalitet att det var möjligt att beräkna effektstorlekar. Sökning skedde i databaserna Eric, Psycinfo samt den tyska databasen Psyndex. Informationen från de olika studierna kodades för utfall(skolprestation, användning av kognitiva och metakognitiva strategier samt motivation), interventionens innehållsliga karaktär (metakognitiv reflektion, möjlighet till grupparbete samt typ av strategi; kognitiv, metakognitiv och motivation) samt karaktäristika för interventionens implementering (årskurs, implementering genomförd av forskare eller lärare samt interventionens längd). En standardprocedur för meta-analys användes för att möjliggöra jämförelse av effektstorlekar mellan de olika studierna. Effektstorlekar har beräknats som standardiserad medelskillnad mellan interventionsgrupp och kontrollgrupp. Vid beräkning av effektstorlek har hänsyn tagits till antal deltagare, antal effektstorlekar presenterade samt om en studie haft flera interventionsgrupper så att resultat från enskilda studier inte fick större påverkan på slutresultaten av fel orsak. Den potentiella effekten (relationen mellan variabler och effektstorlekar) undersöktes med hjälp av metaanalytisk regressionsanalys

Resultaten visar att interventioner i självreglerat lärande är effektfullt och når större effektstorlek om den behandlar skolämnet matematik än skriva/läsa. Interventioner i självreglerat lärande som behandlar skolämnet

matematik visade positiva effekter för elevernas prestationer med en medeleffektstorlek på 0,96 för årskurs 1-6, och en medeleffektstorlek på 0,23 för årskurs 7-10. Deras analys av studierna som innehöll ämnet matematik visade också att ju fler undervisnings- och träningssessioner som ingick i interventionen desto högre elevprestationer i matematik. De jämförde också effekterna av interventioner med olika teoretisk bakgrund och strategiinstruktion. För de tidigare skolåren fanns också ett positivt samband mellan interventioner som fokuserar på kognitiv strategiinstruktion jämfört med interventioner som fokuserar på metakognitiv reflektion. För senare grundskoleår var effekten av interventionerna på elevers matematikprestationer högre om interventionenernas teoretiska bakgrund fokuserade på motivationsteorier i jämförelse med metakognitiva och socialkognitiva teorier om självreglerat lärande. Metaanalysen inkluderade också andra ämnen och interventionernas effekt på andra konstrukt än prestationer. Exempelvis var interventionernas effekt på elevernas strategianvändning i deras självreglerade lärande större i de tidigare skolåren också om interventionerna baserades på en socialkognitiv teoretisk bakgrund, och om de innehöll kognitiv och metakognitiv strategiinstruktion samt metakognitiv reflektion.

2.4 FORMATIV BEDÖMNING: RESULTAT, NATIONELLA PUBLIKATIONER

Detta avsnitt behandlar publikationer i den svenska databasen SwePub och publikationer i Web of Science som innehåller den svenska kontexten (WoS innehöll dock inga sådana publikationer). Alla resultaten baserar sig på utvalda artiklar enligt sökmetod beskriven i Avsnitt 2.2. Inga artiklar som motsvarade sökkriterierna fokuserade på begreppen kamrat- och självbedömning. På grund av att endast nio publikationer i SwePub motsvarade kriterierna för inkludering i denna forskningsöversikt (och fyra av dessa publikationer utgörs av två studier som publicerats på flera ställen (konferensbidrag)) så besvaras forskningsfrågorna för dessa nationella publikationer gemensamt för de olika strategierna för formativ bedömning. Dessa svar baseras på artiklarna som inkluderats i forskningsöversikten och som sammanfattas nedan. Sex av publikationerna identifierades med termen formativ bedömning, en med termen feedback och två med termen självreglering/självreglerat lärande. Två av publikationerna är tidskriftsartiklar, två är avhandlingar varav en licentiatavhandling, en är en bok som baseras på en avhandling och fyra publikationer är konferensbidrag. Konferensbidragen innehåller oftast inte så mycket information, och kan därmed endast bidra med en fingervisning om den forskning som pågår.

Formativ bedömning, feedback och självreglerat lärande

Definition av formativ bedömning, feedback och självreglerat lärande

De definitioner som finns under internationell forskning ovan (avsnitten *Definition av formativ bedömning, Definition av feedback* och *Definition av självreglerat lärande*) gäller även här.

Forskningsfråga 1 och 2: Vilket samband finns mellan formativ bedömning och elevers prestationer i matematik, och vilka egenskaper hos formativ bedömning är centrala i detta samband?

I tre av de publicerade studierna studerades Forskningsfrågorna 1 och 2, och i alla tre studier fanns statistiskt signifikanta samband mellan formativ bedömning och elevers matematikprestationer. I två av studierna analyserades effekterna av lärares användning av en formativ bedömning som karaktäriseras av att hela klassrumspraktiken är formativ. Den tredje studien är en korrelationsstudie. Den formativa klassrumspraktiken i det två första studierna kan beskrivas som innehållande både ett fokus på lärares insamling av information om elevernas tänkande och kunnande med efterföljande modifiering av undervisning, samt stöttning av elevers aktiva engagemang i att själv bedöma och reglera sitt eget lärande och stödja sina klasskamrater i deras självreglerade lärande. I båda studierna användes Wiliam & Thompson's (2007) ramverk för formativ bedömning som inkluderar en stor idé och fem nyckelstrategier. Den stora idén innebär att information om elevers kunskaper återkommande framkallas och baserat på denna information anpassas sedan undervisningen för att passa elevernas identifierade lärandebehov. Strategierna inkluderar att tillsammans med eleverna uppnå en gemensam tolkning av lärandemålen, konstruera uppgifter och frågor som engagerar eleverna i lärande och identifierar deras lärandebehov. De inkluderar också feedback som hjälper eleverna närmare lärandemålen, och stöd för elevernas aktiva agerande som undervisningsresurser för varandra och sitt eget självreglerade lärande.

Den ena av dessa studier (Balan, 2012) är en interventionsstudie i en klass i årskurs ett på gymnasiet, med en annan klass som kontrollgrupp. Efter en termin presterade elever i interventionsgruppen signifikant bättre ($p<0,001$) resultat på ett problemlösningstest än elever i kontrollgruppen. Skillnaden var stor med en effekt-storlek på 1,4 (Cohen's d). Vid en jämförelse mellan grupperna med avseende på de problemlösningsuppgifter som ingick i det nationella provet för kurs A så presterade interventionsgruppen även här signifikant bättre än kontrollgruppen ($p<0,05$; $d=0,6$). Interventionsgruppen presterade även bättre på det nationella provet för Kurs A, men denna skillnad var inte statistiskt signifikant. Ett inslag i den formativa bedömningspraktiken i denna intervention var dock att skapa situationer som gör lärandet synligt genom använde av utmanande uppgifter i

form av problemlösningsuppgifter, och studien baseras endast på en lärare i interventionsgruppen. Den andra studien som använt Wiliam & Thompson's (2007) ramverk är däremot större och inkluderar 22 slumpmässigt utvalda åk 4 lärare i en kommun. Dessa lärare, som utgjorde interventionsgruppen, fick genomgå en omfattande fortbildning i formativ bedömning. Övriga åk 4 lärare i kommunen utgjorde kontrollgruppen (Anderssson, 2014; Andersson, Vingsle & Palm, 2013). Resultaten visar att klasserna till lärarna i interventionsgruppen presterade statistiskt signifikant bättre på ett test i slutet av ett läsår, jämfört med klasserna till konrollgruppen, när hänsyn tagits till resultaten på ett förtest i början av läsåret ($p<0,05$, $d=0,8$). De matematikkunskaper som testades var motsvarande kursinnehållet i matematik för årskursen.

Den tredje studien med relevans för Forskningsfråga 1 och 2 är en korrelationsstudie (Samuelsson, 2011). I denna studie undersöks 219 elevers utveckling av självregleringsfärdigheter under högstadiet, samt dessa färdigheters relation till elevernas prestationer i matematik. Elevernas självregleringsfärdigheter mättes via självskattning i en elevenkät och definieras av inre motivation (elevernas intresse och glädje av matematik), instrumentell motivation i matematik, självuppfattning (self-concept) i matematik och ångest (anxiety) i matematik. Elevernas matematiska prestation mättes med betyget i matematik i årskurs nio. Resultaten visar att då de fyra faktorerna analyserades individuellt så korrelerade alla till betyg. Inre motivation ($r=0,40$, $p<0,01$), instrumentell motivation ($r=-0,52$, $p<0,01$), självuppfattning ($r=0,64$, $p<0,01$) och ångest ($r=-0,54$, $p<0,01$). Den multipla regressionskoefficienten var signifikant ($r=0,69$, $p<0,001$) vilket innebär att det förekommer en relation mellan självregleringsfärdigheter och betyg. I regressionsmodellen var självuppfattning starkt relaterat till betyget i matematik ($p<0,01$) medan inre och instrumentell motivation samt ångest inte hade någon statistiskt signifikant relation till betyg.

Forskningsfråga 3: Vilka egenskaper hos fortbildningar är centrala för att lärare ska vilja och kunna använda formativ bedömning på ett framgångsrikt sätt i sin undervisning?

Endast tre av studierna innehöll någon form av fortbildning för lärarna. Alla dessa tre fortbildningar byggde på ramverket av Wiliam & Thompson (2007). Fortbildningen av åk 4 lärare (Anderssson, 2014; Andersson m fl., 2013) och åk 7 lärare (Boström, 2014) bygger på samma fortbildning som drevs parallellt för ett större antal lärare. I studien av en gymnasielärare (Balan, 2012) fick den enda läraren i studien ett eget stöd av forskaren. Det projekt som de två första studierna var en del av hade ett specifikt syfte att studera vilka egenskaper som är viktiga i stödet av lärares utveckling av en framgångsrik formativ klassrumspraktik. Egenskaper som framträdde som viktiga var bland annat tillgång till extern expertis, och möjligheterna att pröva nya undervisningstekniker och uppleva positiva utfall av genomförandet av dessa i klassrummet. En annan central faktor var att föreläsningar och litteraturläsning varvades med klassrumspraktik och gemensamma reflektioner som stöd för utvecklingen av den formativa bedömningen. Den relativt ansenliga mängd tid som lärarna fick till sitt förfogande för fortbildningen var också viktig för lärarnas möjöigheter att utveckla undervisningen. Även läraren i studien av Balan (2012) hade tillgång till extern expertis och reflektionsmöjligheter över tid.

Forskningsfråga 4: Vilken typ av forskning bedrivs för att besvara forskningsfrågorna?

Åldern på eleverna i de sju tillgängliga studierna varierar från årskurs 4 upp till gymnasiets första år. Typ av formativ bedömning varierar också men fyra av sju studier använder en innebörd av formativ bedömning som baseras på ett ramverk av Wiliam & Thompson (2007) som integrerar flera strategier för formativ bedömning. Att en majoritet av studierna fokuserar på samma ramverk beror till viss del på att det totala antalet studier är litet och flera publikationer kommer från samma större forskningsprojekt. Att en majoritet av studierna också har sin bas i fortbildningsinsatser har samma förklaring. Ett par studier är korrelationsstudier av elevers självregleringsfärdigheter och lika många är fallstudier där lärares undervisning analyserats. På grund av forskningsfrågornas karaktär i denna kunskapsöversikt används statistiska metoder i flera studier. Studierna som inriktades mot formativ bedömning som en integration av flera strategier av formativ bedömning var både situerade i normala klassrumssituationer och över en längre tid, vilket inte var fallet för studierna som fokuserade feedback och självreglering. Detta var också tendensen i de studier som inkluderats från sökningen i

Web of Science. I studierna om formativ bedömning som en intererad helhet låg också ett större fokus på lärarfortbildning och analys av lärares undervisning i relation till den.

Forskningsfråga 5: Hur ser svensk klassrumspraktik ut idag med avseende på formativ bedömning?

Ingen av studierna i denna översikt kan ge en bild över hur svensk matematikklassrumspraktik, med avseende på formativ bedömning, ser ut på nationell nivå. Inga av studierna har heller haft som huvudfokus att beskriva nuvarande undervisning med avseende på formativ bedömning. Några av studierna med andra huvudsyften kan dock ge vissa insikter från några klassrum och kommuner.

Fyra publikationer (Andersson 2014; Balan, 2012; Boström, 2014; Vingsle 2014) beskriver förändringar i klassrumspraktiken efter en fortbildning vilket då ger vissa indikationer på egenskaper hos undervisningen innan fortbildningen. Två studier, Andersson (2014) och Boström (2014), är delstudier i ett större projekt som undersöker effekten av en fortbildning i formativ bedömning på lärarnas praktik och elevernas prestationer i matematik. Dessa studier presenterar preliminära resultat vid konferenser vilket medför att publikationerna innehåller begränsat med information. De beskriver i sina studier de förändringar som 38 slumpvis valda matematiklärare gjort efter en omfattande fortbildning i formativ bedömning, där tidigare nämnda ramverk av Wiliam & Thompson (2007) använts för innebörden av formativ bedömning. Detta kan ge en indikation på att det finns en utvecklingspotential hos lärarna i denna kommun när det gäller denna typ av formativ bedömning. Även Balans (2012) studie stödjer detta, där en lärare i nära samarbete med forskaren utvecklade sin formativa klassrumspraktik. Resultaten i Anderssons (2014) och Boströms (2014) studie indikerar till exempel att lärarna efter fortbildningen ändrat sin klassrumspraktik mot en mer formativ praktik, och att många av dem nu oftare samlar information om elevernas kunskaper. Dessa resultat indikerar att lärarna före fortbildningen inte samlade information om elevernas kunskaper så ofta och följaktligen inte hade möjlighet att göra täta undervisningsanpassningar baserat på identifierade lärandebehov. Dessa studier beskriver också hur undervisningen, med avseende på formativ bedömning, bedrivs i klassrum efter sådana fortbildningar. Studien av Vingsle (2014) beskriver ett sådant klassrum i än större detalj. I sin licentiatavhandling studerade hon en av de lärare som deltagit i fortbildningen som ligger till grund för studien som beskrivs av Andersson (2014). I Vingsles fallstudie beskrivs vilka aktiviteter, kunskaper och förmågor denna lärare nyttjar när hon använder sig av formativ bedömning i sin klassrumspraktik. Denna klassrumspraktik innehåller många formativa inslag. Resultaten visar att läraren regelbundet använder tekniker för att samla information om elevernas matematiska tankar från stora delar av klassen och använder denna information som underlag för modifiering av innevarande lektion eller planering av nästa lektion. Läraren använder elevsvar som underlag för fördjupad informationssamling genom att låta eleverna diskutera vilka svar som är korrekta respektive inkorrekta. Denna och ytterligare aktiviteter är exempel på hur läraren hjälper eleverna att engagera sig i, och själva ta ansvar för, sitt lärande.

Studien av Björklund-Boistrup (2013) fokuserar på feedback. I denna studie analyseras fem lärare i årskurs 4. Björklund-Boistrup identifierar fyra förekommande bedömningsdiskurser i matematikundervisningen, 1) ”Gör det fort gör det rätt”, (2) ”Vad som helst duger”, (3) ”Öppenhet med matematik”, (4) ”Resonemang tar tid”. Dessa diskurser ger skilda aktiva erbjudanden om studerandes aktiva agentskap och lärande i matematik. Slutsatser om hur vanliga dessa diskurser för feedback är utanför dessa klassrum är svåra att dra från studien.

Både studien av Samuelsson (2011) och studien av Fägerstam & Samuelsson (2012) fokuserar på aspekter av självreglering. Resultat i båda dessa studier indikerar att elevernas självregleringsfärdigheter sjunker från årskurs 7 till årskurs 9. Elevernas självregleringsfärdigheter mättes med en elevenkät och definieras av inre motivation (elevernas intresse och glädje av matematik), instrumentell motivation i matematik, självuppfattning (self-concept) i matematik och ångest (anxiety) i matematik. Studien av Samuelsson inkluderar 219 elever och resultaten visar att alla fyra faktorerna minskar under högstadietiden (inre motivation ($p<0,001$), instrumentell motivation ($p<0,001$), självuppfattning ($p<0,001$) och ångest ($p<0,01$)). Fägerstam & Samuelssons studie är en interventionsstudie om hur utomhusundervisning (outdoor education) i matematik eventuellt påverkar elevernas prestationer i matematik och deras självregleringsfärdigheter. Undervisningen för kontrollgruppen med 28 elever genomfördes dock med ”traditionell” matematikundervisning, vilket gör att analysen av dessa elever ger viss information om självreglering hos elever med nuvarande undervisning även om resultaten inte så lätt kan

generaliseras till andra elever. Resultaten från denna studie visar att den enda statistiskt signifikanta skillnaden över tid var elevernas inre motivation (elevernas intresse och glädje av matematik) som minskade de första 10 veckorna av studien i början av årskurs 7 ($p < 0,01$).

Sammanfattning av publikationerna

Balan (2012) avser i sin avhandling ” Assessment for learning - a case study in mathematics education” att undersöka om en förändrad bedömningspraktik, mot en mer formativ bedömningspraktik, har positiv påverkan på elevernas matematiklärande. Hon avsåg även att undersöka vilka dessa förändringar var, och hur elever och lärare upplevde förändringarna. Studien var en kvasiexperimentell interventionsstudie som nyttjade en mixad metoddesign och genomfördes i årskurs 1 på gymnasiet. De 45 deltagande eleverna valdes slumpvis till att delta i interventionsgruppen eller kontrollgruppen, även två lärare valdes slumpvis till vardera gruppen (klassen). Interventionen bestod i att introducera en formativ bedömningspraktik genom att implementera de fem strategierna i ramverket för formativbedömning föreslaget av Wiliam & Thompson (2007). Implementeringen skedde i nära samarbete mellan forskaren och läraren som träffades varje vecka. Datainsamling skedde vid tre tillfällen; i början, i mitten och i slutet av ett läsår. I början av läsåret gjordes ett problemlösningstest, och vid nästa datainsamlingstillfälle gjordes ett nytt problemlösningstest och ett nationellt prov för Kurs A. Problemlösningstesten var baserade på uppgifter från gamla nationella prov. Baserat på dessa tre test kunde jämförelser mellan grupperna göras med avseende på prestationer i matematik. Även intervjuer med alla elever i interventionsgruppen genomfördes vid det andra datainsamlingstillfället. Datainsamlingen bestod också av självrapporteringsenkäter vid alla tre tillfällen för att mäta elevernas matematiska föreställningar (beliefs). Enkäterna var indelade i tre delar som testade epistemologiska föreställningar, matematisk självuppfattning (self-concept) och föreställningar om bedömning. Kvantitativa data behandlades statistiskt med t-test och ANOVA. Kvalitativa data analyserades med kvalitativ innehållsanalys.

Resultaten visar att efter en termin presterade eleverna i interventionsgruppen signifikant bättre ($p<0,001$) på problemlösningstestet än eleverna i kontrollgruppen (på det första testet förelåg ingen skillnad i resultat). Skillnaden var stor med en effektstorlek mätt i Cohen’s d på 1,43. Eleverna i interventionsgruppen presterade något bättre på det nationella provet i sin helhet, men inte signifikant bättre. En jämförelse mellan grupperna gjordes även med avseende på de problemlösningsuppgifter som ingick i det nationella provet. Analys med ANOVA visade då att eleverna i interventionsgruppen presterade signifikant bättre än eleverna i kontrollgruppen ($p<0,05$; Cohen’s $d= 0,6$).

Resultaten från enkäterna indikerar också att interventionen haft viss positiv påverkan på elevernas matematikrelaterade föreställningar. Resultaten från intervjuerna indikerade att eleverna i interventionsgruppen kände igen och uppskattade de olika formativa bedömningspraktikerna som resurser för deras lärande. Författaren menar att lärarens och elevernas uppfattningar av interventionen antyder att alla de implementerade strategierna bidrog till förbättrat lärande, men också att de olika komponenterna påverkade och förstärkte varandra vilket gjorde utvärderingen av varje enskild strategi omöjlig. Enligt författaren indikerar denna respons från både elever och läraren att de formativa bedömningspraktikerna var kopplade på ett komplext sätt.

I avhandlingen presenteras interventionen och den bidrar därmed också med en inblick i hur lärarens formativa klassrumspraktik kom till uttryck och några exempel presenteras här. Matriser användes för att göra mål och kriterier explicita och begripliga. För att skapa situationer som gör lärandet synligt så användes utmanande uppgifter i form av problemlösningsuppgifter. För att tillhandahålla eleverna med formativ feedback så samlades information om deras lärande in bl.a. då de arbetade med problemlösningsuppgifter och denna information användes sedan av läraren då hon gav feedback till grupperna. Grupplösningarna av uppgifterna diskuterades sedan, inte bara utifrån svagheter och styrkor utan även utifrån alternativa lösningsmetoder. För att kommunicera var eleverna stod i relation till mål och kriterier och för att visa vad som behövde göras för att nå längre så användes matriser. För att aktivera eleverna som resurser för varandra och som ägare av sitt eget lärande så introducerades och användes en kombination av kamrat- och medbedömning när interventionsgruppen arbetade med gruppuppgifter. Kamrat- och medbedömaraktiviteter kombinerades på detta sätt med erhållandet och formulerandet av feedback. För att åstadkomma detta så ombads eleverna att bedöma och ge feedback genom att använda bedömningsmatriser.

Andersson m fl. (2013) och Andersson (2014) är konferensbidrag som beskriver samma delstudier tillhörande ett större forskningsprojekt. Boströms (2014) konferensbidrag beskriver en annan delstudie i samma forskningsprojekt. I projektet undersöks effekten av en fortbildning i formativ bedömning på lärarnas praktik och elevernas prestationer i matematik. Det övergripande syftet med projektet är att bidra med kunskap om viktiga faktorer i stödjandet av lärares implementering av en formativ bedömningspraktik. Andersson (2014) studerar 22 slumpvis valda lärare som undervisar i årskurs 4 och Boström (2014) studerar 14 slumpvis valda lärare som undervisar i årskurs 7. Då dessa studier presenterats vid konferenser medför det att endast kortfattad information om dem finns att tillgå och de presenterar preliminära resultat.

Formativ bedömning är konceptualiserat som en klassrumspraktik baserad på användningen av fem nyckelstrategier och en grundläggande idé av att använda belägg om elevers lärande för att justera undervisningen till att bättre möta elevernas lärandebehov (Wiliam & Thompson, 2007). Fortbildningen innehöll teori om formativ bedömning, den grundläggande idén och fem nyckelstrategier. Den erbjöd också tekniker att använda och tid till diskussioner. Under träffarna så planerade lärarna tillsammans implementeringen, delade erfarenheter och diskuterade möjligheter att utveckla sin formativa praktik. Lärarna hade 120 timmar tillsammans med forskningsprojektledaren och de andra deltagande lärarna under våren 2011. De hade också liknande tidsmängd till att individuellt läsa kurslitteratur, reflektera och planera hur de kunde använda teknikerna i klassrumspraktiken.

Andersson (2014) beskriver resultaten för årskurs 4 och de indikerar att fortbildningen hade effekt på klassrumspraktiken. Intervjuer och klassrumsobservationer genomfördes både före och efter fortbildningen för att undersöka lärarnas eventuella förändring av klassrumspraktiken med avseende på formativ bedömning. Även två enkäter om fortbildningen besvarades både direkt efter fortbildningen och ett år senare. Analysen av lärarnas eventuella förändringar skedde med hjälp av ett analysverktyg som utvecklats utifrån de fem nyckelstrategierna och den grundläggande idén. Resultaten indikerar att lärarna gjorde stora förändringar i klassrumspraktiken, om än i varierande grad. De flesta lärarna lockade fram belägg för elevers lärande från alla elever och engagerade dem i lärandeaktiviteter baserat på informationen om deras lärandebehov. Andra tekniker, t.ex. tekniker för kamratbedömning användes mindre. Lärarna upplevde effektivitet i sin undervisning när bedömningsinformation användes till att justera undervisningen för att bättre möta elevernas lärandebehov. Faktorer i fortbildningen som stödde lärarnas förändring var att fortbildningen inkluderade tid med en expert på formativ bedömning och att fortbildningen var omfattande i tid, både vad gällde mängd och längd. Andra faktorer som stödde lärarnas förändring var möjligheten att testa tekniker och uppleva positiva utfall i klassrummen samt att lärarna tillsammans kunde dela erfarenheter och komma över svårigheter.

Den eventuella effekten på årskurs 4 elevernas prestationer i matematik mättes med ett förtest (i början av årskurs 4) och ett eftertest (i slutet av årskurs 4). Testerna avsåg att pröva ordinarie kursinnehåll i matematik för årskursen. Klasserna som undervisats av en lärare som deltagit i fortbildningen (interventionsgruppen) ökade sina resultat signifikant mer än klasserna som undervisats av lärare i kontrollgruppen ($p<0,05$). Effektstorleken mätt i Cohen's d var 0,8.

Syftet med Boströms (2014) studie var att undersöka årskurs 7 lärarnas eventuella förändringar av klassrumspraktiken, på grund av fortbildningen, och att identifiera orsaker till förändringarna. Fjorton slumpvis valda årskurs 7 lärare deltog i studien som genomfördes på samma sätt som Andersson (2014). Preliminära resultat visar att alla lärare var motiverade att förändra och förändrade klassrumspraktiken, men i varierande grad. Den vanligaste och mest frekventa förändringen var att lärarna i större utsträckning samlade in information om elevernas kunskaper för att kunna anpassa undervisningen efter deras lärandebehov. Faktorer som var viktiga för att förändringen skulle ske har också identifierats i studien.

Vingsle (2014b) är ett konferensbidrag och Vingsle (2014a) är en licentiatavhandling som beskriver samma studie. Vingsle (2014a) avser i sin licentiatavhandling undersöka vilka aktiviteter, kunskaper och förmågor en lärare använder när hon genomför formativ bedömning. I fallstudien följs en lärares formativa klassrumspraktik i matematik under 10 veckor i årskurs 5. Ingen intervention ingår i studien, men läraren har två år tidigare deltagit i en kompetensutveckling i formativ bedömning (Andersson et al., 2013). Läraren valdes till studien som en av flera lärare som förändrat sin undervisning i förhållande till formativ bedömning. Datainsamlingen består av ljud- och videoinspelade lektioner. Med formativ bedömning menas i studien sekvenser under elevaktiva genomgångar när lärare, elever eller kamrater framkallar information om elevers kunnande och

tänkande (t ex genom frågor eller uttalanden), tolkar den framkallade informationen och slutligen använder sig av den tolkade informationen för att anpassa undervisning och lärande till elevernas identifierade lärandebehov. De identifierade sekvenserna analyseras med avseende på aktiviteter som orkestreras av läraren (som till exempel när läraren inbjuder elever som aktiva deltagare, ställer frågor och skapar nya aktiviteter utifrån den insamlade informationen) och de kunskaper och förmågor (ämneskunskaper, pedagogiska ämneskunskaper och generella pedagogiska kunskaper) som läraren använder. De analyserade sekvenserna är valda för att inkludera olika tidsintervall mellan insamlad information och använd information.

Resultaten visar att läraren regelbundet använder ett antal aktiviteter för att samla information från så många elever som möjligt. Detta gör hon bland annat genom att ge elever skriftliga uppgifter i slutet av lektioner och använder denna information som underlag för planering av nästa lektion. När hon ställer frågor till klassen mitt under en lektion besvaras dessa ofta med hjälp av små elev-whiteboards där eleverna skriver sina svar. Läraren kan då använda elevsvaren på olika sätt för att anpassa sin undervisning efter elevernas lärandebehov eller som underlag för fördjupad informationssamling genom att låta eleverna diskutera vilka svar som är korrekta respektive inkorrekta. Denna och ytterligare aktiviteter är exempel på hur läraren också hjälper eleverna att engagera sig i och ta ansvar för sitt lärande. När läraren inbjuder elever att aktivt delta i den formativa klassrumspraktiken blir lektionsinnehållet oförutsägbart och läraren hanterar situationer där hon möter oväntad eller okänd matematik vilken hon snabbt behöver förhålla sig till samtidigt som gruppens och individers behov kan vara olika. De kunskaper och förmågor läraren använder under dessa sekvenser är komplexa (från flera kunskapsområden samtidigt) och krävande.

Den sjätte publikationen i en svensk kontext fokuserar på feedback. Björklund Boistrup (2013) baserar sin bok ”Bedömning i matematik pågår - återkoppling för elevers engagemang och lärande” (2013) på sin avhandling ”Assessment discourses in mathematics classrooms: a multimodal social semiotic study” (2010). I boken finns därmed information om den forskning som hon baserar sin avhandling på. Avhandlingens syfte var att ”analysera och förstå tydliga och mindre tydliga bedömningshandlingar i diskursiva praktiker i matematikklassrumskommunikation i termer av elevers möjligheter för elevers aktiva agentskap och lärande” (Björklund Boistrup, 2010, s. 5). I avhandlingen studeras matematikundervisningen och Björklund Boistrup har ett intresse för bedömning och ett särskilt fokus på feedback. Hon använder ett brett begrepp för bedömning som inbegriper de situationer i klassrumsarbetet där det sker en kommunikation med feedback mellan lärare och elev. I studien så presenterades fyra frågeställningar. I den första frågeställningen studerades bedömning så som den visade sig i den feedback (återkoppling) som hon kunde se i kommunikationen mellan lärare och elev. Den andra frågeställningen handlade om vilket fokus dessa bedömningshandlingar hade, d.v.s vad feedbacken handlade om. Den tredje frågeställningen fokuserade vilka roller som olika uttrycksformer kunde spela i de bedömningar som hon fångade i klassrummen, t.ex. att feedback visades genom blickar och gester. I den fjärde och sista frågeställningen så sammanstrålar resultaten från de tre första och det är dessa resultat som sammanfattas här.

Studien genomfördes i klassrumsmiljö där fem klassrum i årskurs 4 besöktes under en vecka. Urvalet av lärare baserades på att få en stor variation vad gällde deras bakgrund (alla var matematiklärare men hade arbetat olika länge och hade delvis olika utbildning), men även upptagningsområdet för de olika skolorna beaktades (stad, land, olika socioekonomisk bakgrund). Klassrummet filmades och utöver detta så filmades lärarna med en bärbar kamera. Två slumpvis valda elever i varje klass videofilmades med en fast kamera. Lärarna och de två eleverna hade även ljudinspelare på sig. Även skriftligt material om de utvalda elevernas matematikundervisning samlades in.

Björklund Boistrup använder två teoretiska inriktningar i sin analys. Dessa är socialsemiotik och ett diskursivt och institutionellt perspektiv. Materialet analyserades med hjälp av datorprogrammet Videograph för multimodal transkribering. Där skrevs vad som sades, kroppsrörelser, gester, ansiktsuttryck och blickar. I analysen så uttolkades sedan fyra diskurser. Uttolkningen av diskurserna skedde i ett samspel mellan syftet av studien, datamaterialet, tidigare forskning om diskurser samt resultat från de tre första analys- och resultatkapitlen om de tre första frågeställningarna. Varje diskurs uttolkades genom (a) vilka typer av bedömningshandlingar som var närvarande, (b) vad bedömningshandlingarna (t.ex. återkopplingar) hade för fokus kopplat till matematikklassrummet, (c) vilka roller de olika uttrycksformerna och andra resurser spelar i bedömningarna och (d) vilka möjligheter till agentskap och lärande diskurserna erbjuder.

Det sammanfattade resultatet av analysen var att Björklund Boistrup uttolkade fyra förekommande bedömningsdiskurser i matematikundervisningen. Dessa var (1) ”Gör det fort gör det rätt”, (2) ”Vad som helst duger”, (3) ”Öppenhet med matematik”, och (4) ”Resonemang tar tid”. Vidare poängterar hon att dessa fyra diskurser ger skilda erbjudanden om elevernas aktiva agentskap och lärande i matematik.

Samuelsson (2011) undersöker i sin korrelationsstudie över tid 219 elevers utveckling av självregleringsfärdigheter under högstadiet och relaterar detta till elevernas prestationer i matematik. I studien så undersöks elevernas självregleringsfärdigheter med en elevenkät (självuppskattning), ursprungligen designad och använd i PISA 2003. Inom PISAs ramverk så är självregleringsfärdigheter inre motivation (elevernas intresse och glädje av matematik), instrumentell motivation i matematik, självuppfattning (self-concept) i matematik och ångest (anxiety) i matematik. Inre motivation innebär här att elevens motivation att uppnå sitt mål kommer inifrån eleven själv. Instrumentell motivation innebär att motivation att uppnå målet kommer från en källa utanför eleven. Självuppfattning är definierat som självuppfattning om ens egen förmåga och kompetens som påverkar möjligheten till framgång i t.ex. matematik. Ångest har definierats på två olika sätt, dels som en känsla av spänning, oro eller rädsla som interfererar med matematisk prestation, och dels som ett tillstånd av obehag, vilket inträffar i situationer som involverar matematik. Den första definitionen fokuserar på effekten av ångest på kognitiva utfall, och den andra definitionen belyser ångests inverkan på självförtroendet (self-esteem). I enkäten får eleverna på en 10-gradig skala uttrycka hur väl de håller med olika påståenden. Elevernas matematiska prestation mäts med betyget i matematik i årskurs nio.

Deltagare i studien var 219 elever från 10 olika klasser och från 10 olika skolor. Studien var treårig och innefattade tre testtillfällen som genomfördes av läraren. För att analysera hur elevgruppernas självregleringsfärdigheter (inre och instrumentell motivation, självuppfattning, och ångest) förändrades över tid genomfördes fyra analyser av varians, ANOVA, (val av statistisk metod har att göra med att andra faktorer också analyserades). De fyra variabler som användes för att undersöka självregleringsfärdigheter användes här som beroende variabler. Analysen för att undersöka hur de fyra självregleringsfärdigheterna förutsäger betyg i matematik skedde i två steg. Först analyserades de fyra faktorerna individuellt med Pearson’s produktmoment korrelationstest och därefter genomfördes en multipel regressionsanalys. Då användes betyg i matematik i årskurs nio som beroende variabel och självregleringsfärdigheterna som oberoende variabel.

Resultaten visar att alla fyra faktorerna minskar under högstadietiden. Inre motivation ($p<0,001$), instrumentell motivation ($p<0,001$), självuppfattning ($p<0,001$) och ångest ($p<0,01$). Resultaten visar också att då de fyra faktorerna analyserades individuellt så korrelerade alla till betyg. Inre motivation ($r=0,40$, $p<0,01$), instrumentell motivation ($r=-0,52$, $p<0,01$), självuppfattning ($r=0,64$, $p<0,01$) och ångest ($r=-0,54$, $p<0,01$). Den multipla regressionskoefficienten var signifikant ($r=0,69$, $p<0,001$) vilket innebär att det förekommer en relation mellan självregleringsfärdigheter och betyg då eleverna slutar grundskolan. I regressionsmodellen var självuppfattning starkt relaterat till betyget i matematik ($p<0,01$) medan inre och instrumentell motivation samt ångest inte hade någon statistiskt signifikant relation till betyg.

Fägerstam & Samuelsson (2012) bidrar i sin studie med information kring Forskningsfråga 5, om hur den svenska klassrumspraktiken ser ut med avseende på formativ bedömning, även om fokus i studien inte var att undersöka detta. Fägerstam & Samuelsson undersöker i interventionsstudien hur utomhusundervisning (outdoor education) i matematik påverkar elevernas prestationer i matematik och deras självregleringsfärdigheter. Hur kontrollgruppens självregleringsfärdigheter förändras över tid kan då bidra med information kring Forskningsfråga 5. Studien genomfördes i årskurs sju. Två lärare och deras elever utgjorde interventionsgruppen och tre lärare och deras elever i utgjorde kontrollgruppen. I kontrollgruppen genomfördes ”traditionell” matematikundervisning. I det ”traditionella” klassrummet följde lektionerna ofta samma mönster; läraren hade en genomgång och sedan räknade eleverna individuellt i läroboken. Elevernas självregleringsfärdigheter mättes på samma sätt som i Samuelsson (2011), med en enkät designad för PISA 2003. Mätningar gjordes i början av åk 7, efter 10 veckor, och i slutet av årskurs 7. I kontrollgruppen var det 28 elever som besvarade enkäten vid de tre testtillfällena och det var dessa elevers resultat som sedan analyserades. Resultaten visade att den enda statistiskt signifikanta skillnaden över tid var kontrollgruppelevernas inre motivation (elevernas intresse och glädje av matematik) som minskade de första 10 veckorna av studien i början av åk 7 ($p <0,01$).

2.5 FORMATIV BEDÖMNING: SLUTSATSER OCH DISKUSSION

Sammanfattning forskningsfrågor

Formativ bedömning kan genomföras på olika sätt. Den kan t ex fokusera på lärarens bedömningar och modifieringar av undervisningen baserat på informationen från bedömningen. I de studier som ingått i denna rapport har detta ibland åstadkommits med hjälp av datorprogram och med fortbildning av lärare. Formativ bedömning kan också fokusera på själva feedbacken från läraren. Men den kan också fokusera på ett stöd av elevernas bedömning av varandras kunskaper eller stöd för deras självbedömningar och vidare modifieringar av sitt lärande i ett ansvarsfullt självreglerat lärande.

De två första forskningsfrågorna som ställs i denna forskningsöversikt är vilket samband som finns mellan formativ bedömning och elevers prestationer i matematik (Forskningsfråga 1), och vilka egenskaper hos formativ bedömning som är centrala för detta samband (Forskningsfråga 2). Resultaten från det stickprov av artiklar som analyserats visar att alla ovanstående strategier för formativ bedömning har ett positivt samband med elevers prestationer i matematik. Effektstorlekarna har oftast varit medelstora eller stora. Däremot är det svårare att utifrån de ingående studierna dra slutsatser om effekten av olika mer specifika egenskaper hos någon av de olika strategierna för genomförande av formativ bedömning. Till exempel indikerar studierna, i likhet med forskning från andra områden än ämnet matematik, att feedback har potential att påverka elevernas prestationer men att denna effekt är beroende på vilken typ av feedback eleverna får och på vilket sätt de får denna feedback. Sambandet mellan feedback och elevprestationer är dock komplext eftersom flera faktorer i en lärandesituation kan påverka detta samband. Det behövs därför många studier för att både undersöka olika faktorers påverkan och dessutom upprepande studier för säkrare slutsatser. De få antalet studier som identifierats i WoS, och skillnaden mellan studierna när det gäller de aspekter av feedback som studerats, gör att de inte på ett tydligt sätt kan verifiera varandras resultat. Säkra och detaljerade slutsatser om hur feedback påverkar elevernas matematikprestationer är därför svåra att dra grundat på dessa studier. Dessutom situeras endast två studier i normal klassrumsmiljö, inga studier inkluderar effekter av lärarfortbildning i feedback, och inte en enda svensk studie inriktas på effekten av feedback på elevers prestationer i matematik.

Men det är också möjligt att integrera alla dessa strategier, och underliggande teorier, för formativ bedömning till en sammansatt helhet. Wiliam & Thompson (2007) föreslog hur dessa olika strategier kunde integreras i ett enhetligt teoretiskt ramverk där alla dessa delar kunde ingå. Lärarna gör då frekventa bedömningar av elevernas kunskaper, ger formativ feedback och anpassar sin undervisning efter elevernas identifierade lärandebehov, och de stödjer också både elevernas självreglerade lärande och deras möjligheter att agera som undervisningsresurser för varandra. Alla dessa aktiviteter baseras på den grundläggande idén med formativ bedömning att aktörerna i klassrummet arbetar för en gemensam tolkning av målen, kontinuerligt bedömer elevernas kunnande i relation till dessa mål och anpassar undervisning och lärande efter de lärandebehov som identifieras genom bedömningarna. Genomförande av en sådan undervisning är betydligt svårare, än genomförandet av en enskild strategi, men skulle kunna tänkas vara ännu mer framgångsrik för att hjälpa eleverna att lära sig matematik. Det finns studier som undersökt effekterna av en sådan undervisning, men inte så många. Av alla studier i denna forskningsöversikt är det endast de svenska studierna av Balan (2012) och studierna från forskargruppen i Umeå (Andersson, 2014; Andersson et al., 2013; Boström, 2014; Vingsle, 2014a,b) som studerat denna integrerade användning av formativ bedömning. Resultaten från dessa studier indikerar också att det är en komplex och avancerad undervisningspraktik, men att stora lärandeökningar hos eleverna är möjliga. De få studier med denna inriktning i denna rapport gör dock att många fler studier är nödvändiga innan mer noggranna och säkrare generaliserbara slutsatser kan dras. Det finnas minst två möjliga anledningar till de få studierna. En är att ramverket är relativt nytt. Den andra är att det är svårare att genomföra den här typen av studier då datorprogram (som inkluderats i flera studier där fokus legat på en enda strategi) får svårt att användas för hela denna integration, och fortbildning för att lärare ska kunna genomföra detta blir då mer komplex och kostsam då det blir mycket svårare att lära sig och implementera en

sådan integrerad formativ bedömning som använder sig av hela detta förhållningssätt och de olika tillgängliga teknikerna. Detta ställer därför höga krav på fortbildningar med denna inriktning i form av stöd och resurser.

Resultaten som indikerar positiva samband mellan formativ bedömning och elevers matematikprestationer sätter ett fokus på hur lärare skulle kunna stödjas för att kunna utveckla sin undervisning mot en mer formativ klassrumspraktik av hög kvalitet. Den tredje forskningsfrågan i denna kunskapsöversikt handlar om vilka egenskaper hos fortbildningar som är centrala för att lärare ska vilja och kunna använda formativ bedömning på ett framgångsrikt sätt i sin undervisning. Det låga antalet studier med fokus på denna forskningsfråga som hittades i sökningen av publikationer indikerar en brist på studier om effekter av fortbildning i formativ bedömning i matematik, och vilka egenskaper hos sådana fortbildningar som är avgörande för att de ska bli framgångsrika. I denna forskningsöversikt fanns endast tre studier (Phelan et al., 2011, Andersson et al., 2013 och Boström, 2014, där de två senare ingick i samma svenska forskningsprojekt) som fokuserar detta och både beskriver egenskaperna hos fortbildningen och dess effekter på elevernas prestationer (Lee et al., 2012 studerar hinder för implementering av fortbildningsinnehåll, men inte effekter på elever). Dessa studier ger insikter i viktiga egenskaper hos fortbildningar men är alldeles för få för att tillförlitliga slutsatser som kan generaliseras ska kunna dras. Olika typer av innehåll kan kräva olika egenskaper hos fortbildningar så även om forskning om fortbildningar på annat innehåll ger värdefulla insikter behövs studier för fortbildningar med det specifika innehållet formativ bedömning i matematik. Att så få studier om fortbildningar i just formativ bedömning dykt upp i artikelsökningen ligger i linje med de bedömningar som framförts av andra forskare att det saknas en stabil forskningsbas för hur framgångsrik fortbildning i formativ bedömning, oavsett ämne, ska utformas (Schneider & Randel 2010; Wiliam, 2010). Även Dignath & Büttner (2008) för i sin metaanalys över interventionsstudier om självreglerat lärande (som presenterats i denna forskningsöversikt) fram bristen på studier om effekten av lärarfortbildning på lärares implementering av en undervisning som stöder elevernas självreglerade lärande och elevernas prestationer.

Den fjärde forskningsfrågan som ställs i denna kunskapsöversikt är vilken typ av forskning som bedrivits för att besvara de övriga forskningsfrågorna. Det kan då först konstateras att studierna inte fördelas jämnt över de olika strategierna för formativ bedömning. Ungefär hälften av studierna har analyserat självreglerat lärande, och färre studier har fokuserat på de övriga strategierna. Forskningen i de olika studierna har varit av olika karaktär, men på grund av forskningsfrågornas karaktär så har många av studierna haft en kvantitativ approach i sin dataanalys. Många av studierna har också varit korrelationsstudier där korrelationen mellan formativ bedömning och elevers prestationer mätts vid ett visst tillfälle. Andra studier har innefattat korta interventioner. Slående är bristen på studier över längre tidsperioder i normal klassrumsmiljö, och studier av fortbildningsinsatser i formativ bedömning. Detta gör det svårt att dra säkra slutsatser av hur en ökning av elevprestationer, via en satsning på formativ bedömning, ska utformas och kan realiseras med framgång i den faktiska praktiken. En annan slående slutsats från denna kunskapsöversikt är den i det närmaste avsaknaden av svenska studier. Inte en enda studie hittades i WoS och när SwePub undersöktes så hittades endast något enstaka befintligt större projekt och studier som ännu inte hunnit publiceras i vetenskapliga tidskrifter. Resultaten av sökningen i SwePub visar att det finns svensk forskning på området, men att den delvis är under uppbyggnad och endast finns på ett fåtal ställen i landet.

En annan fråga som blir relevant till följd av de positiva sambanden mellan formativ bedömning och elevers matematikprestationer är hur svensk matematikundervisning ser ut med avseende på formativ bedömning. Några enstaka svenska publikationer har studerat effekter av fortbildningsprogram i formativ bedömning och kan ge en viss fingervisning om denna aspekt i några skolor och kommuner. Men när det gäller denna fjärde forskningsfråga hittades inte en enda studie i SwePub med fokus på att analysera nuvarande klassrumsundervisning i matematik med avseende på formativ bedömning, och som kan ge en tydligare bild om undervisning ur denna aspekt.

Identifierade behov av forskning med relevans för skolpraktiken

Baserat på denna forskningsöversikt, kort sammanfattad ovan, identifieras följande behov av forskning som kan vara värdefull för både det svenska utbildningsvetenskapliga forskningsfältet och det svenska skolväsendet:

1. *En fullständig forskningsöversikt inom området formativ bedömning i matematik*
   Denna befintliga kunskapsöversikt visar på ett samband mellan formativ bedömning och elevers matematikprestationer. Begränsningarna för genomförandet gör dock att endast ett urval av befintliga artiklar om formativ bedömning och dess effekter har kunnat ingå i översikten. Det betyder att det är svårt att dra mer säkra slutsatser om mer noggranna samband mellan olika aspekter av formativ bedömning och matematikprestationer, och i vilken utsträckning olika typer av forskning inom området existerar. Då denna rapport indikerar starka samband mellan formativ bedömning och matematikprestationer (vilket också ligger i linje med andra mer generella översikter över formativ bedömning (t ex Black & Wiliam, 1998a; Hattie, 2009)) så vore det värdefullt om denna rapport kompletterades med en mer fullständig översikt över den befintliga forskningen på området. Med den nuvarande översikten så finns incitament för skolor att prioritera utvecklingssatsningar inom detta område. Inför större sådana lokala satsningar skulle det vara värdefullt om starkare slutsatser kring sambandet mellan olika aspekter av formativ bedömning och matematikprestationer kan dras, så att skolsatsningar verkligen kan bygga på vetenskaplig grund. En sådan översikt skulle inkludera publikationer från alla relevanta databaser.

2. *Förstärkning av den svenska forskningen inom formativ bedömning i matematik*
   a. *Studier i den svenska kontexten generellt*
      En slutsats som kan dras direkt från denna forskningsöversikt är att den svenska forskningen på området (forskning inom den svenska kontexten) är kvantitativt marginell. Inte en enda studie fanns tillgänglig i Web of Science. Sökningen i SwePub identifierade ett fåtal publikationer som indikerade att det finns forskningsprojekt i ett sådant skede att kommande internationella publikationer kan vara på väg. Med det är endast ett fåtal och forskningsmiljöer på området där det publicerats någon form av rapport, avhandling, bok eller tidskriftsartikel verkar endast existera på några enstaka ställen i landet. Med tanke på de positiva samband som forskningsöversikten visar mellan formativ bedömning och elevers prestationer i matematik behöver denna forskning förstärkas. Den svenska skolan ska vila på en vetenskaplig grund, men även om det går att dra lärdomar av utländsk forskning så fungerar inte undervisning och lärande nödvändigtvis på samma sätt i alla kulturer och kontext. Det är därför av vikt att det även bedrivs forskning om formativ bedömning i matematik inom den svenska kontexten. Dessutom kan forskning i svenska klassrum fungera som ett stöd för skolutveckling och lärares undervisningsinnovationer genom samarbeten mellan forskare, lärare och skolledare i gemensamma forsknings- och skolutvecklingsprojekt med inriktning på formativ bedömning.

   b. *Studier av formativ bedömning som en integrerad helhet*
      Den teoretiska utvecklingen av formativ bedömning har gått mot att kunna integrera de olika strategier för formativ bedömning som kan användas i klassrummet. De individuella strategierna har delvis olika teoretisk underbyggnad och konkret genomförande, men de har alla en gemensam huvudidé om att göra bedömningar av elevers tankar, färdigheter och kunskaper och baserat på informationen från dessa bedömningar anpassa undervisning och lärande till elevernas lärandebehov. Att använda alla befintliga aktörer i klassrummet (lärare, elev, och klasskamrater) i en sådan gemensam process bör ha förutsättningar att vara än mer fördelaktigt för elevernas lärande än om endast en av aktörerna driver denna process. Denna forskningsöversikt indikerar dock att studier som fokuserar på denna helhetssyn på formativ bedömning än så länge är få. Det kommer framöver att fortfarande vara av vikt att studera enskilda strategier för genomförande av formativ bedömning. Men eftersom potentialen för en undervisning som bygger på en helhetssyn på formativ bedömning teoretiskt (och med stöd från de få tillgängliga studierna) kan bedömas

vara stor när det gäller en ökning av elevers lärande, så är det viktigt att täcka bristen på studier inom denna inriktning.

c. *Studier med hög ekologisk validitet*
Behovet är stort av forskning med hög ekologisk validitet, vilket inkluderar studier i normala klassrumsmiljöer över längre tidsperioder. En majoritet av studierna i denna forskningsöversikt är antingen korrelationsstudier där olika variabler har mätts vid enstaka tillfällen, eller interventionsstudier där en specificerad och väl kontrollerad intervention har genomförts under relativt kort tid. Dessa studier är värdefulla men behöver kompletteras med fler studier där insikter genereras om hur formativ bedömning kan fungera för vanliga lärare under normala klassrumsbetingelser under en längre period. Denna typ av studier är centrala för att lärare ska kunna tillämpa en undervisning som bygger på vetenskaplig grund.

d. *Studier av fortbildning i formativ bedömning och dess effekter*
Om undervisningen i svenska skolor ska kunna utvecklas mot en praktik som i större utsträckning innehåller aspekter av formativ bedömning så behöver lärare stöd för en sådan utveckling. Behovet av olika typer av fortbildning inom området är stort, men endast några enstaka artiklar i denna forskningsöversikt undersöker fortbildning. Detta resultat, i kombination med andra forskares identifikation av forskningsbehov, indikerar att det saknas en stabil forskningsbas för hur sådana fortbildningar bör utformas för att ge lärarna tillräckligt stöd för att vilja och kunna utveckla sin undervisning till en sådan nivå av formativ bedömning att de lärandeökningar som funnits i forskningslitteraturen kan uppnås. Det finns många exempel där fortbildningar inte har utformats på ett sådant sätt att dessa mål har uppnåtts (se t ex Schneider & Randel, 2010).

Implikationer för skola och undervisning

För skolor och kommuner som står inför beslut om innehåll i kommande fortbildningar och skolutvecklingssatsningar kan denna forskningsöversikt ge viss vägledning. Baserat på analysen av de 40-tal publikationer översikten bygger på har olika strategier för formativ bedömning visat sig ha goda effekter på elevers prestationer i matematik. Men osäkerhetsfaktorerna är flera. För det första gör det begränsade antalet publikationer som ingår i översikten, tillsammans med att dessa publikationer studerar så många olika sätt att genomföra formativ bedömning, att slutsatser om varje specifik strategi för formativ bedömning blir osäkra. För det andra är antalet studier med hög ekologisk validitet få. Få studier har inkluderat slumpvis utvalda lärare som fått möjlighet att utveckla sin undervisning med avseende på formativ bedömning, och analyserat denna utvecklade formativa klassrumspraktik samt dess effekter i klassrummet under normala förhållanden. För det tredje är de flesta studierna genomförda i andra länder än i Sverige. Endast ett fåtal studier genomförda i den svenska skolkontexten existerar, och egenskaper i genomförandekontexten kan spela roll för vilka egenskaper som är viktiga i genomförandet och effekterna av genomförandet. Det finns dock också faktorer som ökar tillförlitligheten i resultaten. En faktor som ökar sannolikheten att resultaten från de befintliga studierna även ska gälla på många ställen i Sverige är att effekterna på elevernas prestationer identifierats i många länder och kontexter. En annan faktor är att studierna sammantaget genomförts i olika åldrar och för olika strategier för formativ bedömning. Dessa strategier har den gemensamma egenskapen att undervisning och/eller lärande baseras på information om elevernas kunnande eller tänkande och anpassas efter elevernas lärandebehov. Resultaten ligger också i linje med tidigare forskning (t ex Black & Wiliam, 1998; Hattie, 2009; National Mathematics Advisory Panel, 2008).

Inför större sådana lokala satsningar skulle det vara värdefullt om starkare slutsatser kring sambandet mellan olika aspekter av formativ bedömning och matematikprestationer kan dras, så att skolsatsningar verkligen kan bygga på vetenskaplig grund. För att kunna öka säkerheten i slutsatserna skulle en utökad forskningsöversikt vara värdefull. En sådan översikt skulle inkludera publikationer från alla relevanta forskningsdatabaser. Det kan också poängteras att det vid planering av fortbildnings- och skolutvecklingssatsningar är viktigt att beakta befintlig kunskap om hur sådana lämpligen kan utformas när det gäller innehållet formativ bedömning. För att

skolutvecklingsinsatser ska bli framgångsrika behöver även utformningen av dessa att bygga på vetenskaplig grund. Denna forskningsöversikt inkluderar inte publikationer om fortbildning i allmänhet utan bara de som handlar om formativ bedömning och som dessutom är registrerade i Web of Science eller SwePub. Dessa är väldigt få vilket ligger i linje med bedömningar framförda i publikationer utanför denna kunskapsöversikt som både konstaterar att detta gäller rent generellt och att det är ett av de viktigaste framtida forskningsområdena när det gäller formativ bedömning (Schneider & Randel, 2010; Wiliam, 2010).

Metoddiskussion och forskningsöversiktens begränsningar

Tidsramarna för skrivandet av denna kunskapsöversikt möjliggjorde inte en fullständig sökning av publikationer, vilket betyder att en begränsning i sökningen behövde göras. Det betyder i sin tur att de publikationer som inkluderats i översikten är ett urval av befintliga publikationer och val av begränsningar i sökmetoden påverkar detta urval. Oavsett vilken begränsning som görs kommer en del publikationer att inkluderas medan andra exkluderas. I detta fall valdes att begränsa sökningen till publikationer tillgängliga i Web of Science och SwePub. En fördel med Web of Science är att sökningen inte begränsas till vissa subjektivt utvalda tidskrifter eller publikationsställen. Web of Science innehåller också publikationsställen för områden inom såväl utbildningsvetenskap som inom andra vetenskapsdiscipliner. En nackdel med att fokusera på WoS är att vissa tidskrifter som kan vara speciellt lämpliga för sökning inom det aktuella forskningsområdet inte inkluderas. Att inkludera sökning i SwePub har den uppenbara fördelen att översikten kan ge en bild av forskning i den svenska kontexten, som är en del av uppdraget för kunskapsöversikten, även om forskningen inte finns representerad i WoS. SwePub är tänkt att innehålla all tillgänglig svensk forskning så till skillnad från sökningen för den internationella forskningen i Web of Science utgör sökningen i SwePub ingen begränsning av urvalet när det gäller publikationsställen. En sådan utökad sökning när det gäller den svenska forskningen är genomförbar eftersom den svenska forskningen inom området inte är så omfattande.

Ett annat val var att i sökningen i WoS bara inkludera publikationer i tidskrifter. Denna begränsning behöver dock inte nödvändigtvis avsevärt påverka inkluderingen av viktiga studier. De flesta viktiga forskningsresultat som baseras på forskning av hög kvalitet torde, även om de publicerats i en bok, även ha publicerats i en tidskriftsartikel. När det gäller sökningen av svenska publikationer i SwePub utökades sökningen till att inkludera alla typer av publikationer. Detta var möjligt eftersom antalet svenska publikationer var så modest. Det betyder med stor sannolikhet att ytterst få publikationer som studerat den svenska kontexten inte identifierats i sökningen.

Även val av söktermer i sökningen påverkar de publikationer som inkluderas i översikten. I denna sökning användes olika termer som i forskning används för att beteckna formativ bedömning. Både termer som formativ bedömning och bedömning för lärande har inkluderats. Men dessutom har termer som feedback, kamratbedömning och självreglerat lärande använts. Det betyder att flera olika strategier för formativ bedömning finns representerade. En översikt som inte använt dessa olika termer skulle ge en snävare bild av forskningen kring formativ bedömning med den innebörd som kan uttolkas ur den definition på formativ bedömning som ligger som grund för denna översikt. Detta är också en innebörd som är vanlig i litteraturen och som använts i den kunskapsöversikt (Black & Wiliam, 1998a) som verkar vara den mest citerade inom området. Detta val har förstås också nackdelar. Begrepp som t ex feedback är allmänna begrepp så sökningar där titel, sammanfattning eller nyckelord innehåller ett sådant sökord kommer att få träff på väldigt många publikationer som inte är direkt relevanta för de forskningsfrågor och kontext som ingår i denna kunskapsöversikt. För att det ändå skulle vara möjligt att inkludera de olika strategierna för genomförande av formativ bedömning i denna kunskapsöversikt så begränsades sökningen till att söktermerna som t ex formativ bedömning och feedback skulle ingå i publikationens titel. Det betyder att vissa publikationer där dessa termer endast finns i sammanfattningen inte kommer med. En bedömning är dock att i de flesta publikationer där formativ bedömning är ett fokus finns dessa termer med i titeln. En värdering gjordes också att det var viktigare att få med publikationer om olika strategier av formativ bedömning än att få med några få extra publikationer där termerna inte finns med i publikationens titel.

I uppdraget att producera denna kunskapsöversikt ingick inte att göra en bedömning av kvaliteten i de ingående studierna. Alla publikationer som identifierats i sökningen i Web of Science ingår dock i

refereegranskade tidskrifter, vilket borde innebära att studierna har en grundläggande kvalitet. Baserat på den diskussion som förts i forskarsamhället kring kvaliteten i publicerade studier (t ex Dunn & Mulvenon, 2009; Kingston & Nash, 2011) kan det dock inte uteslutas att några av studierna har brister i kvaliteten.

Resultaten från kunskapsöversikten behöver därför tolkas i ljuset av de begränsningar och möjligheter som projektets ramar möjliggör, samt de val som gjorts inom dessa ramar och som diskuterats ovan. Vår bedömning är att ramarna och nödvändiga val har varit tillräckliga för att generera värdefulla resultat om tillgänglig forskning med viktiga implikationer för skolpraktiken. Det har också varit möjligt att identifiera vidare behov av forskning med relevans för skolpraktiken. Det är dock även vår bedömning att en mer fullskalig kunskapsöversikt är både nödvändig och skulle vara mycket värdefull för att få en säkrare och mer detaljerad bild över kunskapsläget inom området formativ bedömning som i denna befintliga kunskapsöversikt visar på en stark koppling till elevers prestationer i matematik.

3.1 CLASSROOM TEACHING: INTRODUCTION

As Hemmi and Ryve (2014) outline, effective teaching as described in the literature during the sixties and seventies stressed the classroom management of learning and aspects of the classroom that were possible to quantify. There was a heavy focus on teachers' presentation of content, and variables such as clarity and flexibility were measured and connected to students' outcomes (Wilson, Cooney, & Stinson, 2005). In these studies, classroom practices and teaching were conceptualized in terms of teachers' activities and their impact on students' learning. However, later studies conceptualize teaching as "classroom interaction among teachers and students around content directed towards facilitating students' achievement of learning goals" (Hiebert & Grouws, 2007). We can see that students become part of the teaching, which harmonizes with the reform movement during the nineties in which students' thinking, activities and contribution to the classroom discourse were emphasized. In addition, as noted by Hemmi and Ryve (2014), stressed aspects also include "the promotion of students' understanding of mathematics through the use of meaningful verbal, situational and visual representations (e.g. Yackel, 2000); the promotion of classroom interactions around problem-solving and students' multiple solution methods (Franke, Kazemi, & Battey, 2007); and high levels of questioning and explaining among teacher and students" (p. 3). The influential Standards of the National Council of Teachers of Mathematics (NCTM, 2000) stress the idea that the teacher should take the students' cognition and knowing as the departure for planning and enacting teaching. To be able to take their departure in students' thinking, the teachers' knowledge is stressed. That is, a research area focusing on the mathematical knowledge for teaching mathematics has developed, with suggestions that teachers need common and specialized content knowledge, as well as knowledge of students and mathematics, of teaching and mathematics and of curricula (Ball, Thames, & Phelps, 2008).

In 2007, Hiebert and Grouws presented a review of research on effective teaching in mathematics. While stressing the need for much more research and the methodological difficulties involved in examining effective teaching, the authors conclude that there is evidence of effective teaching of mathematical skills and conceptual understanding. Firstly, mathematical skills and procedural knowledge are developed by students if the teacher engages in rapidly paced teaching that includes many product-type questions, followed by a massive amount of error-free practice (see Hemmi & Ryve, 2014; Hiebert & Grouws, 2007). Secondly, the development of students' conceptual understanding requires an explicit focus on conceptual aspects as well as the processes of students' struggle with mathematical ideas and concepts.

Another review of effective teaching in mathematics, reported in Anthony and Walshaw (2007) and cited by Brodie (2011), states that while evidence linking specific teaching styles with student knowledge development is weak, important aspects of classroom practice include: an emphasis on mathematical ideas and concepts and their connections through the use of appropriate tasks; an explicit and deep engagement with student mathematical thinking; having high expectations of all students; and both challenging and supporting all students in reaching these expectations.

However, classroom practices are not just teachers' presentation of content and students' engagement in mathematical thinking; they also involve aspects such as mathematical problems and tasks (Watson & Mason, 2005), classroom norms (Yackel & Cobb, 1996), curriculum materials that support both students and teachers (Davis & Krajcik, 2005), new technology, and teachers' questioning and moves (Smith & Stein, 2011). Further, cross-cultural studies reveal that all the above mentioned aspects and processes are culturally specific and that we therefore have to seriously consider the adaptation potential (Clarke, 2013) when it comes to research findings on effective classroom teaching in one culture and transmitting it to another. For instance, Hemmi and Ryve (2014) detected clearly distinguished ways of reasoning about effective mathematics teaching among teacher educators and mathematics teachers in Finland and Sweden. Further, and of particular relevance for the current review, although research suggests key aspects of effective mathematical classroom practices connected to the reform movement aiming to help students develop conceptual understanding and problem-solving competences, reason mathematically, and make connections between mathematical areas, we do not yet know much about how teachers might achieve these high-level practices through questions, interactional moves and the orchestration of classroom discussions.

To sum up, within research there are different definitions of teaching, a variety of ways of conceptualizing effective teaching, and different "lens zooms", in that some research focuses on micro-processes within classroom interactions while other research tries to understand how the broader cultural context affects classroom teaching. Within the current project and time constraints, we place particular focus on research on classroom teaching and instructional strategies, methods and moves, with special interest in the work and actions of teachers.

Our review questions are:

- What characterizes research on classroom teaching practices, teaching approaches and teaching methods in mathematics?
- What characterizes research on teachers' instructional strategies used to establish classroom practices in mathematics?
- What does research tell us about teaching for the learning of mathematical competencies?

3.2 CLASSROOM TEACHING: METHODOLOGY

We base our review process on the logic and stages from Gough et al. (2013). Below, we describe the steps and rationales for our scope, and the processes of searching, screening, and coding.

Scope

Here we describe our inclusion criteria for the review. For an article to be included in the review, it must meet the following three criteria:

- It is about mathematics teaching/learning/education
- It is related to compulsory school (Grades 1-9)
- It addresses the teacher's role

Search methodology

Here we describe our search process.

Search restrictions

We searched Web of Science Core Collection. The indexes we included are:

- Science Citation Index Expanded (SCI-EXPANDED) Science Citation Index Expanded (SCI-EXPANDEDScience Citation Index Expanded (SCI-Expanded)
- Social Sciences Citation Index (SSCI)
- Arts & Humanities Citation Index (A&HCI)

We chose to use Web of Science Core Collection as the database for running our searches. Of course, there are certain limitations in this choice. For example, the journal *ZDM* is not included among the journals cited in Web of Science. Further, the journal *Mathematical Thinking and Learning* was not included in Web of Science until 2009, which excluded, for example, a seminal article by Stein, Engle, Smith, and Hughes (2008) published in this journal (which would otherwise have been captured by our search strings).

Of course, it would have been possible to supplement our search~~es~~ by also running our search strings in the databases MathEduc and ERIC, which are relevant for the field of mathematics education. It also would have been possible to run our searches in the database SwePub for Swedish research. However, within the time constraints of this project, we limited our searches to Web of Science in order to focus on high-quality journal articles.

We limited our search to the English language, the year span 2008-2014, and the document types "article" and "review". The year span was chosen in order to focus on recent research. Regarding document types, we chose to exclude conference papers because the review process for conference papers is not always of high scientific standard, and because many of the important findings from conference papers in mathematics education are usually also published in journal articles. We chose to exclude doctoral theses and books due to time limits, but also because many important results published in theses and books are also published in journal articles. Publishing journal articles in high-impact journals is a strong driving force in academia, as having published articles is seen as crucial in the competition over employment and research funding. Besides journal articles, we chose to include review articles because they offer an overview of the field, which is in line with the aim of this project.

The year 2008 was chosen because we took our outset from a review, published in September 2008, that has a similar focus as the present subproject. The review, by researchers from New Zealand (Walshaw & Anthony, 2008), is called "The teacher's role in classroom discourse: a review of recent research into mathematics

classrooms" and is published in the journal *Review of Educational Research*. In the abstract (p. 516), the authors state that:

> "Current curriculum initiatives in mathematics call for the development of classroom communities that take communication about mathematics as a central focus. In these proposals, mathematical discourse involving explanation, argumentation, and defense of mathematical ideas becomes a defining feature of a quality classroom experience. In this article, the authors provide a comprehensive and critical review of what it is that mathematics teachers actually do to deal with classroom discourse. Synthesizing the literature around a number of key themes, the authors critically assess the kinds of human infrastructure that promote mathematical discourse in the classroom and that allow students to achieve desirable outcomes. From the findings, they conclude with implications for teachers."

We had no restrictions due to, for example, research areas or domains.

Search strategy

Inclusion criteria, keywords and search combinations were chosen to frame and provide direction for the goal of mapping the research, which strives to study teaching in mathematics and give an account of certain qualities in mathematics teaching. The keywords were therefore particularly keyed to basic elements of classroom practice in mathematics and to teachers' role in classroom practice in mathematics.

As we strove for a broad search strategy, we chose not to include search words like "effective", "ambitious", "productive", and "rich", since it was clear at an early stage that these words limited our searches in ways that were not desirable; these terms are used in different ways by different researchers, or are not used at all in studies that are highly relevant for the aims of this review.

We searched in TOPIC in Web of Science; i.e. in the title, keywords and abstract. To be included, an article had to contain both the word math* and a two-word combination consisting of one of the words teach*, instruction*, or pedagogic*, directly followed by one of the words practice*, strateg*, method*, approach*, model*, or move*. Hence, the title, keywords or abstract had to contain both math* as well as one of the following 18 (3·6) two-word combinations:

- teach* practice*
- instruction* practice*
- pedagogic* practice*
- teach* strateg*
- instruction* strateg*
- pedagogic* strateg*
- teach* method*
- instruction* method*
- pedagogic* method*
- teach* approach*
- instruction* approach*
- pedagogic* approach*
- teach* model*
- instruction* model*
- pedagogic* model*
- teach* move*
- instruction* move*
- pedagogic* move*

Search strings were iteratively developed through reading a number of abstracts. At first we had only teach* or instruction* directly preceding practice*, strategy*, method*, or approach*. When we started to read abstracts, however, we noted that the word pedagogic* was sometimes used in combination with practice*, strategy*, method*, or approach*. Therefore, we added pedagogic*. Additionally, we noted that the words model* and

move* were used in some abstracts, so we added these words as well. These additions were made in line with our strive for a broad search strategy. We found a total of 622 articles, which we screened for relevance according to our inclusion criteria. Next, we describe our screening process based on these 622 journal articles.

See Appendix A for the complete search strings we used in Web of Science Core Collection.

Screen methodology

The hits based on our broad literature search (622 journal articles) were screened for relevance based on the stated inclusion criteria (see *Scope* section). The screening was done in two steps: first based on the title, keywords, and journal name, and second based on the abstracts.

Screening 1: Screening based on titles, keywords, and journal names

All 622 journal articles were screened based on their title, keywords, and journal name. We exported all the article information to Endnote X7, which was used as a tool to organize the articles into categories. The following categories were used:

- Potentially relevant on the basis of title, keywords, and journal name
- Discarded on the basis of title, keywords, and journal name – not about mathematics education (107)
- Discarded on the basis of title, keywords, and journal name – not related to compulsory school (Grades 1-9) (60)
- Discarded on the basis of title, keywords, and journal name – not about teacher's role (29)
- Discarded on the basis of title, keywords, and journal name – about very specific objects of study1 (12)

The number of articles discarded on the various grounds is shown in parentheses. In total, 208 articles were discarded in Screening 1. Uncertain cases were discussed among the three researchers within the subproject and decided upon collectively.

Screening 2: Screening based on abstracts

The remaining journal articles – after the screening on the basis of title, keywords, and journal name – were screened on the basis of abstracts. Simultaneous with the screening of abstracts, a coding was done of the articles that remained relevant (see *Code methodology* section). The categories we used were the following:

- Potentially relevant on the basis of abstract
- Discarded on the basis of abstract – not about mathematics education (28)
- Discarded on the basis of abstract – not related to compulsory school (Grades 1-9) (22)
- Discarded on the basis of abstract – not about teacher's role (21)
- Discarded on the basis of abstract – about very specific objects of study (4)

The number of articles discarded on the different grounds is shown in parentheses. In total, 75 articles were discarded in Screening 2. Uncertain cases were discussed among the three researchers within the subproject and decided upon collectively. Additionally, articles focusing on teacher education and professional development were excluded. After Screenings 1 and 2, the category "About professional development and teacher education" contained 97 articles altogether. There was an intention to review the articles on teacher education and

[1] Examples of very specific objects of study include mathematics teaching for deaf or blind children.

professional development separately, but due to time constraints this was not done within the scope of this project.

After the two screenings, 242 (i.e. 622 - 208 - 75 - 97) articles remained potentially relevant for scope of the review. These articles were coded on a number of aspects based on the abstract. Our coding process is described in the next section.

Code methodology

Based on the two screening processes, 242 articles were found potentially relevant for the review. In order to create a basis for the further analysis and classification of our survey, we reviewed the abstracts of these articles according to the following coding focus:

- Object of study
- Method
- Number of participants
- Context
- Results
- Implications for practice

We collected the coding of the articles in a table with these six aspects, basing our coding solely on the information available in the abstract. Every aspect could not be found in every abstract; therefore, we coded aspects that were missing in the abstract as "No information". Some articles did not have enough information in the abstract to allow us to determine whether they matched our inclusion criteria. Then, we read parts of the full-text article, such as the description of context in the methods section. For example, if the abstract referred to high school or secondary school, we had to scan the methods section to see which grades were dealt with. Another example is that it was not always obvious from the abstract if the focus was on mathematics. After the coding and scanning process, 201 articles remained relevant for the review based on the inclusion criteria.

Structuring the mapping

The next step of the mapping was to structure and characterize trends and interests within the discourse of research in mathematics education that focus on teaching methods, classroom practice and teachers' role in classroom practice in mathematics[2]. To do this, we took a closer look at the abstracts of our 201 articles, trying to order them in relation to the nature of the research projects and objects of study. Appendix B includes a matrix with the coding of the 201 articles sorted into categories with the following headings: Promoting the learning of mathematical processes, Promoting the learning of mathematical products, Student attributes, and Practices. The reasons behind this sorting are explained below.

In order to describe and illustrate the various categories, we selected a number of articles for full-text reading. These articles were not selected randomly, but were also not chosen according to certain scientific qualities or with a wish to convey a certain message about how teachers should act or what we consider should be regarded as good teaching. In our screening, we looked for articles that clearly illustrated the category in which they had been classified.

To answer, specifically, our third research question, we dig deeper into the articles dealing with students' learning of mathematical competencies, i.e. those under the heading Promoting the learning of mathematical processes.

[2] Note that this mapping is only of a subset of the research field defined by our inclusion criteria.

3.3 CLASSROOM TEACHING: RESULTS OF MAPPING

Cobb (2007) argues for mathematics education as a design discipline. Considering mathematics education from a design perspective implies that we structure the object of study as a kind of connection between two or more analytical and didactical variables (Ruthven, Laborde, Leach, & Tiberghien, 2009). On the most general level, we can say that teaching is design and learning is what teaching as design is related to and aims to promote. Searching for such relationships was the main task in structuring our mapping.

	Student knowledge		Student attributes (SA)	Practice (P) (Teaching approaches)
	Processes (Proc)	Products (Prod)		
Interactional strategies/patterns (IS)	3	10	4	11
Teaching approaches (TA)	9	42	8	
Learning material (LM)	5	14	2	7
Background variables (BV)	1	16	2	34
Characterization (Char) Total: 33				

Figur 1. Categories and analytical relationships of the mapping.

From our coding of the abstracts we distinguish a set of local and fine-grained analytical relationships related to the teaching and learning in mathematics (Figure 1). The matrix in Figure 1 provides an overview of the nature and research directions, gained from our coding of the 201 articles (see Appendix B). It gives us an overview of the trends, directions and focuses highlighted and prominent in mathematics education research, which is close to classroom practice. Of the 201 articles, 168 were structured according to an analytical relationship between two didactical variables; in the remaining 33, the object of study did not follow such a structure. In this group of studies, the object of study was to uncover and characterize a phenomenon such as characterizing students' understanding of or strategies for specific subject matter content (Wagner & Davis, 2010). Below, we will return to discussing the trends and focus that can be inferred from the overview pictured in the matrix, and what implications for further research can be drawn from the information the overview provides. We will also dig deeper into and describe the research that particularly focuses on students' learning in relation to *mathematical processes*. The basic reason we do this is that we consider viewing mathematical knowledge from a process perspective richer than viewing it from a product perspective (NCTM, 2000). In this way, we find it reasonable

to believe that we will be able to capture the most fair and representative picture of how research conceptualizes mathematical classroom practice. But, before doing this, we will elaborate on the analytical categories represented in the matrix.

Analytical categories

We first describe the analytical categories in our framework, and then illustrate the findings with full-text articles from the different categories. Of course, it may be too simplistic to say that one variable is the design variable (independent variable) and another is the target variable (dependent variable). Two didactical variables are reflexively dependent. However, from our coding we find that research is relatively clear about the analytical relationship that is the object of study, and that the objects of study often follow the principle that there is some design aspect (independent variable) that should result in some effect (dependent variable). What we distinguish as independent variables in research focusing on classroom practice is gathered in the left column in the matrix (Figure 1), while the target variable (the dependent variable) is presented in the top row of the matrix. The number of articles in each category is shown in each square of the matrix. For example, 42 articles have been categorized as TA+Prod. It is rather common for an article to belong to two categories (squares). In such cases, the article's most prominent category is counted in Figure 1.We will use the set of target variables as an overall structure for our presentation. But, to understand the relationship between the different didactical variables there is a need to clarify our classification of the design variables; that is, the variables that are connected to aspects of teaching.

Our mapping discloses how research connected to classroom practice focuses on different levels of teaching and on different aspects of teaching in mathematics, and we found it appropriate to group these differences according to *teaching approaches*, *interactional strategies*, *learning material* and *background variables*. Note that the references can be found in Appendix B.

Interactional strategies

Studies belonging to this category explicitly refer to teachers' moves and actions. The teacher, and how he/she acts in interaction with the students, is central to the investigation. The focus is on how teachers communicate and engage with their students, and what role the communication and engagement play in students' learning of mathematics. It could be about, for instance, how a teacher uses gestures (Shein, 2012) and questions (Panayiotou, Kyriakides, Creemers, McMahon, Vanlaar, Pfeifer, Rekalidou, & Bren, 2014) to promote students' learning. It could also concern how teachers struggle with following up on students' ideas in order to develop the mathematical classroom practice (Akkus, 2013).

Teaching approaches

Teaching approaches refer to studies taking a broad perspective on classroom teaching in mathematics. The teachers' actions and interactional behavior are not the main object of investigation. The teacher may be important, but it is the more general and overall structures of teaching that are the object of study. In our review we found different examples of teaching approaches, such as technology-based teaching (Allsopp, McHatton & Farmer, 2010), mathematical games (Bragg, 2012), *problem-based teaching* (Xin, Zhang, Park, Tom, Whipple & Si, 2011), and contrasting ways of using textbooks in mathematics teaching (Even & Kvatinsky, 2010).

Learning material (task design)

Some studies focus on how a specific artefact or design principle can support or challenge students' learning in mathematics (Gravemeijer & van Eerde, 2009). In contrast to focusing on overall teaching approaches, this group of studies looks into some specific aspect of a teaching approach. For instance, instead of investigating computer-assisted learning in general terms, studies belonging to this category may investigate the role of visualization or simulation in the learning of mathematics (David & Tomaz, 2012). It could also be the case that the study investigates a particular computer program, designed for particular purposes in the learning of

mathematics. In this group we also include task design. When we talk about task design, we refer to types of mathematical tasks and sequences of tasks (Hattikudur & Alibali, 2010). Included in this category are also interventions evaluating specifically guiding instructions for solving mathematics tasks (Krawec, Huang, Montague, Kressler, & de Alba, 2013; Orosco, 2014), and task design in relation to cognitive load (Ngu, Yeung & Tobias, 2014).

Background variables

Background variables are not design variables in the same way as teaching approaches, interactional strategies and learning material. However, in our mapping we found that several studies did connect classroom practice and students' learning to something different from these groups of didactical design variables. Particularly, we found studies giving accounts of personal attributes such as teachers' beliefs (Bray, 2011), attitudes (Palardy & Rumberger, 2008) and knowledge (Tchoshanov, 2011) in order to explain their classroom practice and their students' learning and performance. There are also studies referring to student attributes in order to explain why they do or do not learn mathematics in a certain classroom practice. Other examples of this category are studies investigating aspects of mathematics teaching and learning in relation to school culture (Imbo & LeFevre, 2009), parents (Lubienski, Lubienski, & Crane, 2008), and general issues of implementation of programs for improvement (Brown, Pitvorec, Ditto & Kelso, 2009; Meaney, Trinick & Fairhall, 2013).

Characterization

In one group of studies (33 articles), the object of study does not follow the structure of an analytical relationship between two didactical variables. In this group of studies, the object of study is to uncover and describe analytical categories of a phenomenon. It may concern the characterization of curriculum material (Bryant, Bryant, Kethley, Kim, Pool, & Seo, 2008; Sherin & Drake, 2009), the role of interactive whiteboards (Beauchamp, Kennewell, Tanner & Jones, 2010), mapping the mathematics in classroom discourse (Herbel-Eisenmann & Otten, 2011), or profiling students' understanding or strategies of specific subject matter content (Wagner & Davis, 2010). Describing a certain practice or teaching approach is the focus in itself, rather than how the practice or approach impacts students' learning.

Grant, Stronge and Xu (2013) analyze and characterize key variables of effective classroom practice in mathematics by observing classrooms of China and US teachers who have received national awards for their teaching. In both cultural contexts, the teachers a) used a variety of instructional activities that spanned across different cognitive levels, b) were opportunistic planners in order to maximize meaningful student learning, c) had high student engagement, d) presented effective classroom management skills, and e) maintained a learning environment that was conducive to optimal learning.

The purpose of Walkowiak, Berry, Meyer, Rimm-Kaufman, and Ottmar (2014) is to introduce a tool to measure standards-based mathematics teaching practices, the Mathematics Scan (M-Scan), and to examine its validity and score reliability. The object of study is the M-Scan tool as a quantitative tool for measuring teaching practices that are widely acknowledged as effective in mathematics teaching. Sixty video-taped lesson observations were made in 60 third- and fourth-grade classrooms, and were coded using the M-Scan as well as two other measures: RTOP and CLASS. The validity of the M-Scan was analyzed in three ways: by content review by experts, by response processes of coders, and by relating the M-Scan to the other two measures. Also, the reliability of the M-Scan was analyzed. Regarding validity, results show that valid inferences can be made from the M-Scan scores about the extent to which standards-based mathematics teaching practices are present in classrooms. Regarding reliability, results indicate that the M-Scan shows promise for broader use in large-scale research focused on mathematics teaching practices.

Illustration and validation of analytical relationships

To illustrate and validate the analytical relationships classified in Figure 1, we have read articles in full text from a number of different categories. We have compared the classification made based on the object of study with the classification made based on reading the whole article. Below, the analytical relationships are

elaborated on and illustrated by presenting the results of our full-text reading. In Appendix C, these articles are presented according to the headings we used for coding full-text articles.

Student knowledge

What should count as mathematical knowledge or knowing mathematics is a continuing debate in mathematics education. In recent decades we have seen a change from conceptualizing mathematical knowledge exclusively related to mathematical products to doing so in terms of mathematical processes (Kilpatrick, Swafford & Findell, 2001; Niss, 2007). Mathematical products relate to conventions, symbol systems, concepts and methods of mathematics. Mathematical processes refer to students' ability to investigate, discuss and see connections and patterns in mathematical situations. In relation to the dichotomy between product and process, we can also formulate a change from mathematical knowledge to *knowing* mathematics. Adding to the notion of mathematical knowledge, the idea of knowing mathematics points to processes of how conventions, concepts and methods are put into use in a mathematical discourse (Sfard, 2008). A shift towards a more process-oriented perspective on mathematical knowledge and knowing is also evident in international policy documents, of which one of the more elaborate examples is the NCTM Principles and Standards (NCTM, 2000). Niss (2004) describes the ability to handle mathematical processes within a system of eight competences, which is also the framework that OECD/PISA use as a basis for their international comparisons of student achievement (Niss, 2007). The Swedish school curriculum (Swedish National Agency for Education, 2012) has also broadened the view of knowledge and knowing in mathematics. Process knowledge is linked to central content (products) in order to offer students opportunities to develop the ability to formulate and solve mathematical problems, evaluate strategies and methods, use and analyze mathematical concepts, conduct and follow logical mathematical reasoning, and use mathematical language to talk about and explain calculations and solutions.

In the coding and analysis of the abstracts, we distinguished between studies that emphasized a product view of mathematics and those looking at the classroom practice or students' outcomes from a process-oriented perspective on knowledge and knowing of mathematics.

Promoting the learning of mathematical processes

Studies coded as belonging to the process group referred to learning outcomes in terms of students' ability to communicate, reflect on and argue for a solution. Studies focusing on problem-solving ability and conceptual understanding were also coded to this group.

Interactional strategies and the learning of mathematical processes

Matsumara, Cadman Slater and Crosson (2008) investigate how classroom climate and rigorous teaching practices relate to students' participation in whole-class discussions. Thirteen 6^{th}- and 7^{th}-grade mathematics teachers working at schools with mainly minority students from low-income families participated in the study. Two lessons with whole-class discussions were observed for each teacher. Rubrics were used to measure the quality of classroom climate, rigor of instruction, and students' interactions. Matsumara et al. (2008) found that the presence of explicit rules in the classroom for respectful, prosocial behavior significantly predicted the number of students who participated in the class discussions. They also found that "the quality of students' participation in class discussions – that is, the degree to which they built on other students' contributions and explained and supported their responses – was predicted by teachers pressing students to explain their thinking in discussions and by the rigor of the questions posed to students in the discussion" (p. 293).

Teaching approaches and the learning of mathematical processes

Cotic and Zuljan (2009) compare students' problem-solving abilities and attitudes towards mathematics for two kinds of instruction: problem-based versus conventional. The problem-based instruction used mathematical problems close to students' lives and interests. Also, emphasis was put on problems with multiple answers, problems with insufficient or superfluous information, and problems with contradictory data or no solution. That problems can be solved in a variety of ways was an important part of the instruction. The conventional

instruction consisted of the teacher demonstrating procedures to the students, whereupon the students practiced with tasks similar to the example given by the teacher. The research design followed the structure of a pre-test–intervention–post-test design. The pre- and post-tests, conducted at the beginning and end of the school year, respectively, focused on a) cognitive measures (procedural and problem-solving performance) and b) affective measures. Third-grade students (n = 179) were assigned to one of the two groups, and received classroom instruction accordingly. The results show that students who received problem-based instruction were more successful in solving more difficult problems than those who received conventional instruction, but there were no statistically significant differences regarding procedural computation between the two groups. Regarding affective variables, there were statistically significant differences in some variables but not others.

Learning material and the learning of mathematical processes

The goal of Star and Rittle-Johnson (2009) was to evaluate whether comparing solution strategies is more effective than the sequential study of strategies in learning about computational estimation. Students from two schools participated in the study. At School A, a private urban school, 69 fifth-grade students participated (32 girls and 37 boys); and at School B, a small rural school, 45 fifth graders and 46 sixth graders participated. The research design followed the structure of a pre-test–intervention–post-test design, including a retention test. The results of the study contribute to a growing body of research demonstrating that comparing multiple strategies in problem-solving activities facilitates learning. The focus here is on estimation, which is both a critically important real-world skill and a mathematical domain that is significantly more complex than equation solving, which has been the target of prior work. Comparison helped students develop a larger repertoire of estimation strategies, improved their ability to select the most appropriate strategies for computing an easy estimate, and increased some students' retention of conceptual knowledge. The current results provide experimental evidence that, when learning how to estimate, it pays to compare.

Background variables and the learning of mathematical processes

In Thorvaldsen, Vavik and Salomon (2012) it is both background variables and teaching approach that influence students' learning. The authors compare "best practice" classes (KappAbel winners) with average Norwegian classes with regard to the use of ICT tools, pedagogical practices, and teachers' education. Problem-solving is central in the KappAbel contest. The teachers of the KappAbel winners and a control group report on their teacher education and teaching practices via a Likert-scale questionnaire. The results of the study are that: 1) KappAbel teachers emphasized reasoning-oriented more than instrumental- or rule-oriented teaching, 2) KappAbel teachers had more often studied at universities than control teachers, and 3) KappAbel teachers used ICT tools that are specific to mathematics (like spreadsheets) for purposes of exploration more than control teachers, although KappAbel teachers saw weaker general effects of ICT than control teachers. The authors draw the conclusion that it is the teacher-guided student activity, i.e. the inquiry-guided learning environment, in which ICT may be used, that makes the difference for students' learning. What is special in KappAbel classes is the combination of the three interdependent variables a) teachers' higher mathematics mastery and b) their choice of specific ICT tools, through which c) a more problem-solving and inquiry-based teaching approach is possible. It is hence not ICT per se that makes the difference. Thorvaldsen et al. (2012) stress that classroom observations would be needed to substantiate these conclusions.

Promoting the learning of mathematical products

Several of the studies belonging to the product group measured student outcomes using standardized tests. Other articles did not measure student learning through tests, but referred to learning outcomes basically in the language of domain-specific concepts and methods.

Interactional strategies and the learning of mathematical products

The purpose of Stronge, Ward and Grant (2011) was to examine the impact teachers had on student learning gains and, then, to examine the instructional practices and behaviors of effective teachers. The effectiveness of participating teachers was classified according to their students' academic growth in a single year. Classroom

differences between effective and less effective teachers were examined in terms of both their teaching behaviors and their students' classroom behaviors. The teachers who were more effective in terms of student achievement were more organized, used routines and procedures with greater efficiency, and held higher expectations for their students' behavior. The top-quartile teachers had fewer classroom disruptions, better classroom management skills, and better relationships with their students than did bottom-quartile teachers. The specific interactive parameters examined concerned instructional focus, feedback, clarity, teacher and student questions, and disruptive behavior by the teachers. Although there proved to be large differences between the two teacher groups regarding these parameters, the difference was not significant at the level of $p < .05$. Stronge et al. (2011) conclude by claiming that "to make a difference in the quality of education, we must be able to provide ready and well-founded answers to the question, What do good teachers do that enhances student learning?" (p. 351). Even though the authors are aware of the limitations of their study and that their results should be applied with caution, they suggest that their study moves research closer to answering this fundamental question.

Shein (2012) reports on a revelatory case study of a 5th-grade teacher's ways of orchestrating the discourse and interaction in her classroom to create opportunities for English language learners to participate in the repair of mathematical errors during a unit on finding the area of geometric shapes. The analysis of discourse takes on a binocular perspective of considering gesture and speech as a unity. The teacher's pointing, representational, and writing gestures were studied in relation to her questioning and revoicing. This research was guided by a social learning theory that characterizes learning as active and interactive participation in communities of practice. The findings highlight three qualitatively different ways in which gestures were used by the teacher to promote student participation in mathematical classroom practice: (a) in grounding mathematical discussions, (b) in revoicing student strategies, and (c) in narrating the meaning of geometric features.

Teaching approaches and the learning of mathematical products

As an argument for carrying out the study presented in Ayres (2013), it is noted that in order to handle high cognitive load caused by problem complexity, specific instructional methods are required. Isolating elements is one such strategy, as it reduces element interactivity by initially presenting partial tasks before progressing to whole tasks (Ayres, 2006a; Pollock, Chandler, & Sweller, 2002). The current study investigates whether this strategy could be improved. To investigate the idea that spending more time practicing the most cognitively demanding components error rates would decrease due to this extended practice, the study set up three groups of a total of 54 Year 8 students from Sydney. The students were divided into low/high prior knowledge and then randomly assigned to one of three groups engaging in full-worked strategy ($N = 20$), the equal-isolated strategy ($N = 18$), or the targeted-isolated strategy ($N = 16$). The results of the study show that the targeted-isolated strategy was most effective when compared with the full-worked example strategy, and a significant advantage was found in completing the acquisition problem-solving tasks. Further, in comparing the newly developed targeted-isolated strategy with the equal-isolated strategy, no overall advantage was found for the targeted approach on test measures, the cognitive load measure, or the efficiency measures. However, there was significant interaction found on the Transfer test. Students with low prior knowledge were disadvantaged by the targeted strategy, while those with higher levels of prior knowledge were neither disadvantaged nor advantaged.

Learning material and the learning of mathematical products

David and Tomaz (2012) investigate how visual representations can structure geometry activity in the classroom. They present one illustrative episode from a public elementary school in Brazil to show how drawings of geometrical figures play a powerful role in structuring and modifying geometry activity in the classroom. The framework of Activity Theory was used to characterize the episode as a system of interconnected activities. The analysis focuses on the changes and transformations perceived in these activities, and explores the idea of miniature cycles of learning actions to focus on the mathematical learning that is taking place. The analysis describes how the drawings structured the mathematical activity in the classroom. At some moments they played the role of an artifact while at others, even if not previously planned by the teacher,

they became the focus and the very object of the activity. In particular, visualization created potential learning related to the following mathematical products: new procedures for doing the calculation of areas; characterization of geometrical figures; the norms and rules for the use and interpretation of a drawing in mathematics; the representation of a right-angled triangle by a drawing, and the notion that not all triangles are right-angled; and the notation for a right angle and the conservation of areas.

Background variables and the learning of mathematical products

Boonen, Van Damme and Onghena (2014) investigate the effects of teachers (background qualifications, attitudes and beliefs, and instructional practices) on student achievement in mathematics, reading, and spelling in 1st grade. The study's theoretical framework and methodology are based on recent work by Palardy and Rumberger (2008). Data from the SiBO Project, a longitudinal study in Flemish primary education, were analyzed for the present study using two-level regression analysis. The results showed that teacher background had the largest effect on mathematics achievement. In mathematics, students with more experienced teachers tended to perform better, whereas those with teachers doing more in-service training tended to perform worse. Of the attitudes and beliefs variables examined, job satisfaction was the only one that was significantly associated with math achievement. No variables measuring instructional time or modality were associated with math achievement in Grade 1. Findings revealed that instructional practices had the greatest effect on student achievement in reading and spelling, but not on student achievement in mathematics. Overall, the results suggest that teachers had a modest to strong effect on student achievement in first grade.

Student attributes

Student attributes are comparable to the notion of background variables. In our review, we note that some studies do not investigate teaching in relation to students' knowledge but are instead interested in targeting variables such as students' beliefs, attitudes, behavior and motivation. We refer to such variables as different kinds of attributes of students.

Interactional strategies and influences on student attributes

In a study from the US (Berry & Kim, 2010), four teachers are followed through four lessons with a focus on characterizing their interactions with students. Data from Year 1 classrooms were collected, and the researchers were particularly interested in understanding the interactional patterns established in classrooms characterized as inclusive, in that students with disabilities form a substantial part of the class. The study found six categories of teacher utterances: (a) question/elicit, (b) respond to students' contributions, (c) organize/give instructions, (d) present/explain, (e) evaluate, and (f) asssociate. Content-related teacher talk comprised elicitations of recall-based responses, preceding acknowledgement of these responses. This pattern was similar across the four teachers, who rarely asked students to provide explanations, share ideas, or assist peers. The authors speculate about implications from practice, suggesting that teachers should use a variety of approaches in teaching. For instance, they should use direct instruction to teach basic mathematics knowledge, such as mathematics facts; use self-instruction (i.e., verbal prompts to mediate cognitive and metacognitive operations) to teach processes (Kroesbergen et al., 2004); and use interactive instruction involving "discussion of strategies (i.e., general metacognitive and domain-specific heuristics), multiple solutions to problems, and a 'debriefing' component" (Woodward & Montague, 2002, p. 97) to teach problem-solving.

Teaching approaches and influences on student attributes

A common argument for introducing new technologies in mathematics teaching is that it is expected to increase students' motivation to learn mathematics. Following this tradition, Baki and Cakiroglu (2010) investigate students' views and attitudes regarding learning mathematics in, and from engaging with, the computer-based learning environment of *Learning Objects* (LOs). LOs, web-based learning labs, can be designed to support the exploration and investigation of mathematical ideas and relations, and enhance conceptual understanding. Students with different learning needs may also benefit from LOs, for instance in studying new topics or reviewing previous ones, doing homework, or working on projects. LOs can also help teachers introduce new

topics or concepts that are more difficult to explain using traditional teaching methods, by providing students with more extended activities. In the present study, the LOs were built on drill and practice, problem-solving, decision making and concept teaching activities. In the study, 30 9th-grade students at a high school in Trabzon, Turkey studied topics from the 9th-grade mathematics curriculum in an LO-based learning environmen. Both quantitative and qualitative data were gathered. Quantitative data included a web-based evaluation form and Internet records of working time period and the frequency of solving particular tasks. The qualitative data were gathered from interviews conducted with six students and the teacher. The interviews were used to gain further information about the students' opinions about and attitudes towards working with LOs. These views and attitudes were then analyzed and categorized as themes. Also, the data from these interviews were compared with the data received from observations made by the researchers. In order to understand the reasons why the teacher preferred LOs and the aims for using them, the teacher's views were obtained through semi-structured interviews at certain time intervals. The qualitative data obtained from the teacher were categorized and presented according to frequency of usage. In the interviews, the students were asked to present their experiences with LOs throughout the study as liked and disliked cases. They were asked to refer to LOs as either being or not being engaging, enjoyable or interesting. They compared their lessons with LOs to their previous lessons without LOs, and also specified their preferences for using LOs. The students' views and attitudes about the computer-based learning environment of LOs indicate that the innovations in courses by using LOs have influenced the students in a positive way. They made positive comments about the motivational and learning themes; for instance, they expressed that LOs were not boring, that they enjoyed working with them, and that they found them interesting and motivating. They also added that they were pleased with their overall performance; the LOs helped them study and learn more, and encouraged them to take more responsibility in their learning process. The students liked the LOs, as they were considered to offer enhanced interactivity and visual potential. Being in a learning environment under their own control through the use of LOs increased the students' interest in the course. Games or competitions, two important factors in LOs, created an interesting learning environment for the students. Overall, the students evaluated the Learning Object Repository (LOR) as *highly* sufficient in terms of Learning Value, Added Value , and Usability of the Design. The Technology Function was considered sufficient. Negative views expressed generally involved similar topics, such as getting tired and worrying about exams. The authors offer no reflection on practical implications.

Learning material and influences on student attributes

Schukajlow et al. (2011) developed a self-report instrument to measure how different problem formats may affect students' enjoyment, interest, value and self-efficacy expectations. Using this instrument, 224 German ninth graders were asked about their enjoyment, interest, value and self-efficacy expectations concerning three formats of mathematical problems: intra-mathematical problems, word problems and modeling problems. Enjoyment, interest, value and self-efficacy were assessed before and after a ten-lesson teaching unit promoting modeling competency related to the topics "Pythagoras' theorem" and "linear functions". The study aimed to answer the following research questions: (1) Do students' enjoyment, value, interest and self-efficacy expectations differ depending on the type of task? (2) Does the treatment of modeling problems in classroom instruction influence these variables? (3) Are there any differential effects for different ways of teaching modeling problems, including a "directive", teacher-centered instruction and an "operative-strategic", more student-centered instruction emphasizing group work, and strategic scaffolding by the teacher? Using the instruments developed in the study, it was found that students' enjoyment, value, interest and self-efficacy expectations are essentially identical for modeling problems, "dressed up" word problems and intra-mathematical problems. However, teaching oriented towards modeling problems had positive effects on some of the student variables, with the student-centered teaching method producing the most beneficial effects. So student-centered teaching methods seem to be better suited for improving both students' achievements and their affect when dealing with modeling problems.

Background variables and influences on student attributes

Birgin, Baloglu, Cathlioglu, and Gurbuz (2010) investigate mathematics anxiety on a sample of Turkish students in relation to mathematics achievement levels, perceived enjoyment of the mathematics teaching method (PET), perceived enjoyment of mathematics (PEM), and perceived help with mathematics from parents (PPM). The study group, consisting of 220 sixth- through eighth-grade Turkish students,, represents the culture and socioeconomic status of a suburban community. It is considered that this geographic region of Turkey will reflect the overall picture of the country in terms of mathematics anxiety; because, according to the PISA results, the mean mathematics anxiety score of the region is the same as that of the entire country.
A Personal Information Form and the Mathematics Anxiety Scale for Elementary School Students (MASESS) were used to collect the data in the study. The form included demographic items such as gender and grade level and questions about mathematics achievement level (assessed based on the mathematics grade from the previous year), PET, PEM, and PPM. Three standard multiple regressions were computed between mathematics anxiety as the dependent variable and mathematics achievement, PET, PEM, and PPM as independent variables for three grade levels.

For all three grade levels, the highest zero-order correlations were between mathematics anxiety and mathematics achievement. Whereas the lowest anxiety was measured among the sixth graders, the highest was among the eighth graders, which shows that mathematics anxiety levels increase as grades increase. Gender difference in mathematics anxiety is the single most studied environmental antecedent. In the present study, no significant difference between boys and girls was found. However, the inconsistency of the results is still present. This might be because mathematics anxiety is not a well-defined and measured construct. The Turkish study found stronger and more prevalent grade level differences than gender differences in mathematics anxiety. For all grade levels, PEM had a significant negative effect on mathematics anxiety. In addition, PEM was the second highest contributor to mathematics anxiety, following mathematics achievement, for the eighth graders. The study also found that PET had a significant negative effect on mathematics anxiety on sixth and seventh graders, but not on eighth graders. After mathematics achievement, this was the second highest contributor to mathematics anxiety for the sixth graders.

It is concluded that students' own perceptions of mathematics and the teaching method are more prevalent than their parents' help with mathematics. Results indicate that the impact of PET on mathematics anxiety seems to diminish with age, whereas the impact of PEM seems to be increasing. Therefore, it can be recommended that mathematics anxiety remediation programs focus more on attitudes towards teaching for younger children and towards mathematics as a whole subject for older children. Results from this study could be used by researchers to plan further studies of mathematics anxiety and to compare their results with the results of the present study. The results could also be used by educators or counselors to plan remedial programs or to revise current course requirements and teaching strategies. Future experimental studies are needed to determine the causes of mathematics anxiety, as most studies on the topic are correlational. In addition, further investigation of mathematics anxiety with younger populations is always encouraged.

Practice

Several articles are not very specific about learning and learning outcomes, instead targeting creating conditions for learning. These studies are based on the assumption that learning necessitates that there are already opportunities for learning established (Hiebert & Grouws, 2007). Hence, this group of studies investigates how to establish practices that are assumed to provide opportunities for learning. The Practice category corresponds to the design variable Teaching approach. However, in contrast to Teaching approach, which should be regarded as a means to promote learning, the variable Practice refers to studies that have the actual practice as the goal in an analytical relationship. In this way, Teaching approach can never be in an analytical relationship to Practice, because they consist of a similar structure. What differs is the role of this structure in analytical relationships.

Interactional strategies promoting practice

Akkus (2013) examines the teacher's role in creating argument-based class discussions. A teacher participating in a year-long professional development focusing on argument-based whole-class discussions based on students' ideas was selected for analysis. A criteria matrix capturing three areas of pedagogical practice was used for the analysis. The three areas are: creating dialogical interaction; controlling knowledge and the problem-solving process; and unit preparation and making connections. The results showed that the students were given limited opportunities to express and negotiate their ideas. The students were not given the responsibility for the problem-solving process, and several opportunities to address their misconceptions were lost. The authors explain these difficulties with the teacher's "comfort zone" in the traditional way of teaching.

A study from the US (Bozack, Vega, McCaslin, & Good, 2008) investigates how teachers can interact with students in the classroom to support the students' autonomy. Field notes from 106 classroom observations from the authors' larger Comprehensive School Reform (CSR) sample were analyzed. The analysis used Autonomy Supportive Behavior Instrument (ASBI) for classroom observations, an instrument based on self-determination theory (SDT) on how teachers can foster autonomy in the classroom. The findings show that all eight characteristics suggested by SDT were present in the field notes, but their prevalence and the nature in which they were evident differed. Students had little opportunity to choose how they wanted to work with objects; they had many opportunities to talk, but teachers rarely helped them relate ideas and concepts between topics or lessons. Explicit instances of encouragement or teachers engaging the experiences, expertise, or perspective of students were rarely identified. The findings suggest that teachers in these settings could extend their autonomy-supportive behavior by scaffolding relationships among concepts and ideas across lessons and subjects, i.e. create opportunities for students to make connections. Also, teachers could further extend their autonomy-supportive behaviors by not only responding to, but also elaborating on, the things their students say, and by engaging students' own experiences, expertise, and perspectives in the learning process.

Brodie (2011) sets out to examine the hybrid nature of classroom interactions in the reform mathematical classroom by developing a coding scheme focusing on teachers' interactional moves. The data were collected among four South African secondary teachers who were enrolled in a professional development program, and focused on the teachers' classroom practice for two weeks each, one before and one after the professional program. The study derives five categories of the teachers' interactional moves: affirm, direct, initiate, inform and follow up. Particular emphasis is put on the last category, follow up: insert, elicit, press, maintain, and confirm. The study shows, except for the coding scheme as such, that the teachers' shifts among types of moves were similar among those who used the same or similar tasks, and does not show clear differences across classrooms of different SES. As such, the study adds to evidence that suggests that a teacher with the interest and resources to change teaching in lower SES classrooms can do so. The author suggests that the coding scheme, while developed primarily for researchers, might also be used by teachers to talk and reflect upon their practice.

Learning material promoting practice

Discussions about school mathematics often address the importance of reasoning and proving for building students' understanding of mathematics. Bieda (2010) investigates processes and outcomes of implementing proof-related tasks in classroom practice. The study addresses the research question: What is the nature of the students' and teachers' actions and discourse in the enactment of tasks written to engage middle school students in justifying and proving? Data are collected through field notes on observations of seven middle school classrooms in the connected mathematics project (CMP). The results show that students regularly responded to proof-related tasks by offering conjectures that created opportunities to prove, and that students provided justifications for nearly half of the conjectures, resulting in proving events. However, the findings indicate that students' experience with proof-related tasks are insufficient for developing their proof-ability and understanding of what constitutes valid mathematical justifications. Moreover, although the students were regularly involved in responding to proof-related tasks, the teachers in the classroom observed did not provide sufficient feedback to sustain discussions based on students' conjectures and justifications. Findings also indicate that when a teacher offered feedback on students' justification, it was not sufficient to establish

standards for proofs in a mathematics classroom. For instance, teachers were as likely to sanction a justification with a positive appraisal if it was based on a non-proof argument as if it was based on a general argument.

Background variables and influences on practice

Bray (2011) examines how teachers' beliefs and knowledge influence their ways of handling students' errors during class discussion related to mathematics tasks. The study has a collective case study design, with four third-grade teachers at an urban school going through a one-year reform-oriented professional development program, Everyday Mathematics. To characterize each teacher's typical response patterns to student errors, analyses were made of four lessons with class discussions per teacher. To analyze each teacher's response patterns in relation to her beliefs and knowledge, data from pre- and post-interviews and measures of beliefs and knowledge (at the beginning and end of the year) were used. Also, a cross-case analysis was made of the case stories to identify dimensions of the teachers' error-handling practices that captured their response patterns. Results on the teachers' beliefs and knowledge show that three of the four teachers made significant shifts toward reform-oriented beliefs during the year. Results from the cross-case analysis reveal three dimensions of error-handling practices across the four teachers. These dimensions are: 1) Intentional focus on flawed solutions in whole-class discussions; 2) Promotion of conceptual understanding through discussion of errors; and 3) Mobilization of a community of learners to address errors. Teacher beliefs appear to be highly related to how they structure class discussions with respect to errors, while teacher knowledge seems to be highly related to the quality of their response patterns to errors during class discussions.

In addition to the mapping described above, we deepened our mapping of full-text articles that relate teaching to students' learning of mathematical competencies in order to answer review question 3. To do this, we selected for in-depth reading the articles of particular relevance for students' learning of mathematical competencies. This selection was made from the 201 articles (see Appendix B) as described in the next section.

Mapping of articles that relate teaching to students' learning of mathematical competencies

The focus of our review is to particularly map research results that articulate something about the effects of teaching on students' learning of mathematical competencies (see review question 3). Due to our focus on the development of mathematical competencies, we focus on students' learning of processes rather than products or student attributes. Therefore, we dig deeper in the mapping of the research results of full-text articles that are coded as Proc. As can be seen below, most articles in this category have the research design of pre-test–intervention–post-test in order to look at the effects of teaching on students' actual learning in terms of mathematical competencies.

The studies coded as P and that focus on how teaching can create opportunities for students' learning of mathematical competencies are also of interest in relation to mathematics teaching practices. However, here we choose to focus specifically on the studies that go furthest in the sense that they can say something about the effects of teaching on students' learning of mathematical competencies. In other words, we choose to focus on the studies that do not primarily focus on "opportunities to learn", although these are also very important (see section *Opportunities to learn*, below).

In total, 18 articles were categorized as Proc (see Appendix B). Two articles were excluded from reading in full-text: Kotsopoulos and Lee (2012) and Carbonneau, Marley and Selig (2013). Kotsopoulos and Lee (2012) was excluded because it builds on students' self-reported learning and understanding, while Carbonneau et al. (2013) was excluded because it is itself a meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. In the reading of the 16 remaining articles in full text that belong to the category Proc, three areas emerged. These are studies that deal with: 1) Problem-solving teaching, 2) Comparison as an effective teaching strategy, and 3) Pushing students to explain their thinking in detail. Next, we will present these areas.

Problem-solving teaching

Cotic and Zuljan (2009) compare students' problem-solving abilities and attitudes towards mathematics for two kinds of instruction: problem-based versus conventional. The problem-based instruction used mathematical problems close to students' lives and interests. Also, emphasis was put on problems with multiple answers, problems with insufficient or superfluous information, and problems with contradictory data or no solution. That problems can be solved in a variety of ways was an important part of the instruction. The conventional instruction consisted of the teacher demonstrating procedures to the students, whereupon the students practiced with tasks similar to the example given by the teacher. The research design followed the structure of a pre-test–intervention–post-test design. The pre- and post-tests, conducted at the beginning and end of the school year, respectively, focused on a) cognitive measures (procedural and problem-solving performance) and b) affective measures. Third-grade students (n = 179) were assigned to one of the two groups, and received classroom instruction accordingly. The results show that students who received problem-based instruction were more successful in solving more difficult problems than those who received conventional instruction, but there were no statistically significant differences regarding procedural computation between the two groups. Regarding affective variables, there were statistically significant differences in some variables but not others.

Le, Lockwood, Stecher, Hamilton and Martinez (2009) report on a three-year longitudinal investigation of the relationship between teachers' self-reports of reform-oriented teaching and mathematics and science achievement. Le et al. (2009) state: "Results suggested greater exposure to reform-oriented instruction was generally not significantly associated with higher student achievement but the effects became stronger with prolonged exposure to reform-oriented practices. Reform-oriented instruction showed stronger, positive relationships with open-ended measures than with multiple-choice tests in both mathematics and science and with problem-solving skills than with procedural skills in mathematics" (p. 200). The authors stress that measurement matters. Achievement measures that were aligned with the goals of reforms – problem-solving skills and open-ended assessments – were more related to reform-oriented instruction than procedural measures, which are common on US state and district tests. Further, the authors stress that new teaching approaches always have to be coordinated with other aspects of the teaching context, such as the teacher's skills and the curriculum. Note that Le et al.'s (2009) study builds on teachers' self-reported teaching practices, and includes no data from actual classroom instruction. The teachers filled in surveys for a period of a year, and kept logs for two to five days to indicate how much time they spent on different classroom activities. They also reflected on vignettes of teaching problems, and stated what their own actions would be in the situations depicted in the vignettes.

Gresalfi, Barnes, and Cross (2012) investigate how classroom practices shape affordances for student engagement, and how these affordances were realized. The aim is to examine the nuances of classroom interaction with a goal of interrogating the mechanisms that impact observed student behavior; the aim is not to make strict causal claims about the impact of particular teacher moves on student learning. Case studies were performed on two teachers' instructions and whole-class discussions in Grade 7. The students in the two classes worked on the same project-based statistics unit, consisting of three "real-world", inquiry-based tasks (also called "rich tasks" by the authors). Each task was worked on for one lesson. The tasks offered the students opportunities to engage procedurally, conceptually, consequentially, and critically. Consequential engagement involves connecting a rule or a decision with its implications. Critical engagement requires that students be intentional and reflect on the rationale behind their decision.

Most opportunities offered by both teachers involved engaging procedurally or conceptually. Far fewer opportunities were offered to engage consequentially and critically. To seek the answer as to why this was, the authors looked at each teacher and found great differences: for one teacher, the opportunities to engage consequentially or critically much more often resulted in the same form of engagement for students. This teacher spent more time during launching the task on clarifying what the students' solutions should look like, i.e. what kinds of arguments are convincing. This move reinforced the norm that students needed to consider the mathematics of a problem, even when it involved a storyline. It is clear that teachers' framing of an activity significantly impacts the ways students are likely to engage with a task. Offering clear expectations is especially important in supporting students' engagement with content.

To summarize the reviewed studies on problem-solving teaching so far, the results of Cotic and Zuljan (2009) and Le et al. (2009) point to the possibility that problem-based instruction and reform-oriented instruction can have positive effects on students' problem-solving ability. Also, the results of Gresalfi et al. (2012) point at the importance of establishing productive norms and expectations for the success of problem-solving teaching.

Some studies focus on technology use in problem-solving teaching. Thorvaldsen, Vavik and Salomon (2012) compare "best practice" classes (KappAbel winners) with average Norwegian classes regarding the use of ICT tools, pedagogical practices, and teachers' education. Problem-solving is central in the KappAbel contest. The teachers of the KappAbel winners and a control group report on their teacher education and teaching practices via a Likert-scale questionnaire. The results of the study are that: 1) KappAbel teachers emphasized reasoning-oriented more than instrumental- or rule-oriented teaching, 2) KappAbel teachers had more often studied at universities than control teachers, and 3) KappAbel teachers used ICT tools that are specific to mathematics (like spreadsheets) for purposes of exploration more than control teachers, although KappAbel teachers saw weaker general effects of ICT than control teachers. The authors draw the conclusion that it is the teacher-guided student activity, i.e. the inquiry-guided learning environment, in which ICT may be used, that makes the difference for students' learning. What is special in KappAbel classes is the combination of the three interdependent variables a) teachers' higher mathematics mastery and b) their choice of specific ICT tools, through which c) a more problem-solving and inquiry-based teaching approach is possible. It is hence not ICT per se that makes the difference. Thorvaldsen et al. (2012) stress that classroom observations would be needed to substantiate these conclusions.

Song and Looi (2012) aim at investigating how teacher beliefs influence classroom practice and how practices influence student inquiry learning in a computer-supported collaborative learning (CSCL) environment. Student learning is examined using a theoretical framework that looks at how the group moves from divergent to convergent thinking: a) Idea generating, b) Idea linking, and c) Intellectual convergence. Additionally, Song and Looi (2012) add d) social support. The authors make a cross-case comparative analysis of two Singapore teachers with very different beliefs, both of whom had undergone more than a year of a professional development focusing on inquiry-based learning and innovative practices in mathematics. The inquiry principles are: 1) Working on authentic problems, 2) Encouraging different ideas, 3) Making progressive inquiry, 4) Providing collaborative opportunities, and 5) Doing embedded assessment. A fine-grained analysis is made of how the two teachers enact the same lesson on division and fractions with 5th grade students in a CSCL environment based on inquiry principles. The CSCL tool allows groups of students to post their ideas for public view in different representations (idea generating), comment on each other's ideas to improve the solution proposals (idea linking), support each other (social support), and reach a common understanding (intellectual convergence). The results showed that the teacher with "innovation-oriented" beliefs enacted the lesson in an inquiry-oriented way and effectively used the technology affordances. This contributed to inquiry, encouraging collaborative student learning. The other teacher believed that the students did not have enough prior knowledge of the concepts, and that they easily developed misconceptions if allowed to work without her guidance. She believed she had to use content-based instruction, which is strongly teacher-centered. She asked a great deal of right or wrong questions, judged students' answers as right or wrong, and did not encourage diverse ideas. Hence, a collaborative inquiry culture was not established. There was very little idea linking or social support, which meant that there was also little intellectual convergence in the class. The findings suggest that a teacher holding innovative-oriented beliefs and enacting inquiry-based practices is likely to lead to effective student inquiry learning.

Scheiter, Gerjets and Schuh (2010) analyzed whether animations can be used effectively for problem-solving teaching in mathematics. Since realistic animations had proven unsuitable for the teaching of problem-solving, hybrid animations were used. According to Scheiter et al. (2010), hybrid animations "explicitly show the transition from a concrete representation of the problem statement to the abstract representation necessary to construct a problem model" (p. 498). The research design followed the structure of a pre-test–intervention–post-test design. Students in the 9th grade (n = 32) were divided into two groups, both of which were to read three textbook chapters (on biology, chemistry, and politics). In each chapter, three solved examples were included that differed in their underlying solution procedure and algebraic formula. Additionally, one group

could watch hybrid animations of the examples as many times as they wanted. The total time was two hours. The pre-test consisted of items on prerequisite concepts for solving algebra word problems. To measure students' performance, the post-test distinguished between four different problems depending on how surface features (context) and structural problem features differed between the problems taught and the post-test problems: 1) Equivalent test problems with both surface and structural features similar to the taught problems, 2) Isomorphic test problems with only structural features similar to the taught problems, 3) Similar test problems with only similar surface features, and 4) Unrelated problems sharing neither surface nor structural features. The results showed a significant overall difference between the two groups in favor of the hybrid animations group. This applied to equivalent, isomorphic, similar, and unrelated problems, and the effects grew stronger with increasing transfer distance.

To summarize the reviewed studies on technology use in problem-solving teaching, Thorvaldsen et al. (2012) and Song and Looi (2012) draw the conclusion that it is the teacher-guided student activity rather than the tool itself that makes the difference in students' learning, and that teacher beliefs can influence classroom practice.

Several studies focus on approaches to problem-solving teaching for students with learning problems or learning disabilities. Xin, Zhang, Park, Tom, Whipple, and Si (2011) compare two different strategies for problem-solving: a conceptual model-based problem-solving (COMPS) approach, and a general heuristic instructional (GHI) approach. The study focuses on learning outcomes in terms of both multiplicative-division word-problem-solving and prealgebra for students with learning problems. The mathematical problems in the study are limited to two types of multiplicative-division word problems: equal groups and multiplicative comparision. The research design followed the structure of a pre-test–intervention–post-test design, including a maintenance test. Students with mathematical learning problems in Grades 3-5 (n = 27) were assigned to one of the two groups (COMPS or GHI). The COMPS group received systematic instruction on conceptual models for each of the two problem types, and were trained in detecting the problem type and mapping mathematical relations onto a conceptual model diagram. The GHI group was guided by a five-step general heuristic problem-solving checklist. The results of the study were that the COMPS group improved significantly more than the GHI group on the multiplicative-division word-problem-solving test and on the prealgebra model expression test (but there was no significant improvement on the prealgebra equation solving test). Xin et al. (2011) suggest that elementary students with learning problems "can be expected to move beyond concrete operations and to algebraically represent mathematical relations in conceptual models that drive the solution plan for accurate problem-solving" (p. 381). A limitation stated by the authors is that the problems in the study consisted of only two kinds of multiplicative-division problems. They state further that more complex problems with multiple steps and irrelevant information need to be included in future studies.

Jitendra and Star (2011) discuss schema-based instruction (SBI) as an alternative to traditional instruction (in textbooks in which all problems on a page can be solved by the same procedure) as a means for raising the problem-solving performance of students with learning disabilities. The authors state that the flexible use of strategies is based on comparison of multiple solution strategies as an effective teaching method, and refer to Star and Rittle-Johnson (2009) (see next section). SBI encourages students to go beyond surface features of a word problem to grasp its underlying mathematical structure. This is done by identifying the problem schema (semantic structure) of the problem type. The schemas in the domain of arithmetic word problems – Change, Group, Compare, Restate, and Vary – are hierarchically organized and stored in long-term memory, which is important for students with working memory deficits. In Jitendra and Star's (2011) article, a schema for ratio and proportion problems for middle school students is presented. The teacher explicitly and systematically uses a four-step strategy: 1) Find the problem type, 2) Organize the information in the problem using the diagram, 3) Plan to solve the problem, and 4) Solve the problem. The teacher leads the students through Steps 1-3, and the students solve the problem using the solution strategy identified in Step 3. The authors refer to a study by Jitendra et al. (2009) in which one student group in 7th grade received SBI and the control group received textbook procedural instruction on ratios and proportions. The results showed that the SBI group outperformed the control group on both a problem-solving post-test and a delayed post-test.

Orosco (2014) aims at investigating the effectiveness of a mathematics comprehension word problem strategy for Latino English language learners at risk for mathematics disabilities. The strategy, called Dynamic

Strategic Math, builds on a dynamic assessment framework and provided scaffolding based on reading and language comprehension levels. The tester adapted the word-problem-solving to each student's word-problem-solving cognition level through a vocabulary modification procedure with four levels of mathematical terms: 1) Everyday, common words, 2) General math words, 3) Specialized math vocabulary, and 4) Technical vocabulary. The tester assessed students' word-problem-solving ability with probes. Six 3rd grade students participated in the study. The research design followed the structure of a pre-test–intervention–post-test design. At the pre-test, all six students were at Level 1 (basic level with math terms used in everyday conversation). They received 17 sessions of 20-25 minutes each for a period of five weeks, in addition to their regular mathematics lessons. The intervention consisted of three parts: a) Explicit teaching of concepts and vocabulary, b) Teaching of strategies, and c) Co-operative practice of strategies (one student with one teacher). The results of the study were that all six students raised their level of performance from 1 to 3 (five students) or 4 (one student), and these levels were maintained in follow-up sessions. The authors state that the results suggest a causal relationship between the intervention and word-problem-solving accuracy. The results also suggest that comprehension strategies may help English language learners at risk for math disabilities to develop their word-problem-solving ability.

Krawec, Huang, Montague, Kressle, and Melia de Alba (2013) examine the effectiveness of SolveIt! (Montague, 2003) cognitive strategy instruction on the knowledge of problem-solving strategies. SolveIt! instruction explicitly teaches the cognitive processes and metacognitive strategies used by proficient problem solvers. The following seven cognitive processes are included in the SolveIt! routine: 1) Read the problem, 2) Paraphrase the problem, 3) Visualize the problem, 4) Hypothesize about problem solutions, 5) Estimate the answer, 6) Compute the answer, and 7) Check the accuracy. The research design followed the structure of a pre-test–intervention–post-test design. Middle school students in 7th and 8th grade with learning disabilities (n = 78) as well as average-achieving students (n = 83) participated in the study, either in the treatment or the comparison group. During the school year, students in the treatment group received a three-day initial instruction and then 30 minutes of weekly instruction according to the SolveIt! routine. Teachers in the comparison group were asked to proceed as they usually did, but to focus on word-problem-solving at least one lesson per week. The pre-test and post-test consisted of interview questions focusing on knowledge and reported use of problem-solving strategies. The students who had received SolveIt! instruction reported using significantly more strategies than the comparison group. This was the case for both average-achieving students and those with learning disabilities. The medium effect size was 0.52. The SolveIt! group improved significantly from pre- to post-test, while the comparison group did not. Average achievers reported using significantly more strategies than did students with learning disabilities (not surprisingly), whereas students with learning disabilities in the SolveIt! group actually increased their reported strategy use to a slightly higher level than the average-achieving students in the comparison group. The results showed that SolveIt! was equally effective regardless of ability level. Among the limitations stated by the authors is that it is not actually known whether the students used the reported strategies when solving problems – we only know what they said they did. According to the authors, however, previous studies have consistently shown the effectiveness regarding problem-solving performance of SolveIt! for students with learning disabilities. Another limitation is that the time spent on problem-solving in the comparison group is not known; only that they had problem-solving once a week and that the teachers mainly posed a worked example whereupon students solved similarly structured problems. Therefore, Krawec et al. (2013) state that there is a possibility that it is the time spent on problem-solving itself that makes the difference, rather than the SolveIt! intervention.

Montague (2008) reviews self-regulation strategies for improving problem-solving for students with learning disabilities. Seven studies are evaluated, and the self-regulation component is described for each study. Three of the studies are Montague's own, and the self-regulation components in Montague's case are the seven cognitive processes included in the SolveIt! routine (see Krawec et al., 2013) above. Other examples of self-regulation components among the articles reviewed are: 1) Read the problem out loud, 2) Select the important information, 3) Draw a representation of the problem, 4) Write down the steps for doing the computation, 5) Check the answer; or similarly: 1) Read the problem out loud, 2) Look for important information and circle it; 3) Draw pictures to help tell what is happening; 4) Write down the math sentence; 5) Write down the answer. The seven studies reviewed by Montague (2008) show positive results for students' performance. Montague's

(2008) reviewed self-regulation strategies are similar to Polya's (1945/57) four problem-solving phases: 1) Understand the problem, 2) Make a plan, 3) Carry out the plan, 4) Look back at the solution.

To summarize the reviewed studies focusing on problem-solving teaching for students with learning problems or learning disabilities, they claim that strategy instruction can be effective (Krawec et al., 2013; Montague, 2008). They also claim that schema-based or model-based instruction (Jitendra & Star, 2011; Xin et al, 2011) can be effective. However, as Xin et al. (2011) point out, more complex problems with multiple steps need to be included in future studies, as do more types of problems.

Comparison as an effective teaching strategy

Star and Rittle-Johnson (2009) evaluate whether comparing solution strategies is more effective than sequentially studying strategies for learning about computational estimation. Students in the 5th and 6th grades at two schools participated in the study (n = 157). The research design followed the structure of a pre-test–intervention–post-test design, including a retention test. The results of the study contribute to a growing body of research demonstrating that comparing multiple strategies for solving the same problem facilitates learning. The focus here is on estimation, which is both a critically important real-world skill and a mathematical domain that is significantly more complex than equation solving, which has been the target of prior work. Comparison helped students develop a larger repertoire of estimation strategies, improved their ability to select the most appropriate strategies for computing an easy estimate, and increased some students' retention of conceptual knowledge. The current results provide experimental evidence that, when learning how to estimate, it pays to compare.

Rittle-Johnson, Star, and Durkin (2012) build on their earlier findings for experienced learners (e.g. Star & Rittle-Johnson, 2009 – see above) to explore whether it pays to compare also in the case of novices. Eighth-grade students (n = 198) who were novices in multi-step equation solving were randomly assigned to one of three groups participating in different interventions in the classroom. In the first group, comparison of multiple procedures was done at once (first lesson of two). In the second group, the first lesson offered two examples with the same procedure while comparison of multiple procedures was delayed until the second lesson. In the third group, procedures were introduced one at a time in the first lesson while alternative procedures were not introduced until the second lesson. Explicit comparison was not done at all in the third group. The research design followed the structure of a pre-test–intervention–post-test design, including a retention test on procedural flexibility, procedural knowledge, and conceptual knowledge. Procedural flexibility is a key in problem-solving, with students learning and using multiple procedures appropriately to solve problems. Procedural flexibility consists of two parts: flexibility knowledge and the flexible use of procedures. The results show that comparing algebraic procedures immediately (first group) was more effective for procedural flexibility for novice learners than was delayed comparison of procedures (second group). In turn, greater procedural flexibility was related to greater conceptual and procedural knowledge. Delayed exposure without comparison (third group) was as effective as comparing procedures immediately, except for flexible use of procedures, for which immediate comparison was superior. In sum, it can be effective for novice learners to compare multiple procedures at an early point.

Hattikudur and Alibali (2010) investigate whether instruction comparing the equal sign with other relational symbols (<, >, and ≠) is more effective than instruction on the equal sign alone. Students in 3rd and 4th grades (n = 112) were divided into three groups: Comparing symbols, Equal sign, and Control. The research design followed the structure of a pre-test–intervention–post-test design. The pre- and post-test assessed 1) conceptual understanding of the equal sign, 2) equation encoding, and 3) problem-solving. The intervention consisted of each group having one lesson of one-to-one instruction. The students were first given descriptions and examples of each symbol (only = for the equal signs group and =, <, >, and ≠ for the comparing symbols group). They then solved three example tasks with the help of feedback from the instructor and, lastly, practiced on their own. The results of the study are that students in the comparing symbols group increased more in their conceptual understanding of the equal sign and also scored higher on a post-test on inequality symbols and inequality problem-solving than the other two groups. However, equation encoding and equivalence problem-solving did not show these gains. Still, students improved their conceptual understanding

of both equal and inequality symbols through a lesson on comparing symbols. The authors state that their findings show that an instructional method based on comparison can facilitate students' understanding of mathematical concepts.

To summarize the reviewed studies on comparison as an effective teaching strategy, all three – by Hattikudur and Alibali (2010), Rittle-Johnson et al. (2012), and Star and Rittle-Johnson (2009) – state that comparison can be effective for students' learning.

Pushing students to explain their thinking in detail

Webb, Franke, Ing, Chan, De, Freund and Battey (2008), investigating the role of teachers' instructional practices in collaboration among students, found correlations between teachers' questioning practices, students' explanations and student achievement. Three teachers, who received instruction in eliciting the details of student thinking, and their students in 2nd and 3rd grades, participated in the study. The results showed that although all three teachers asked the students to explain their thinking, they probed student thinking differently. The authors state: "The differences in the extent to which teachers asked students to elaborate on their explanations showed a strong correspondence with the nature and extent of student explaining to each other during collaborative conversations" (p. 377). The teacher with the highest amount of correct and complete student explaining pushed students to explain their thinking in detail by asking questions intended to make explicit their thought steps. This was the case for both correct and incorrect initial student explanations. Webb et al. (2008) are careful about inferring causal relationships from their study.

Matsumara, Cadman Slater and Crosson (2008) investigate how classroom climate and rigorous teaching practices relate to students' participation in whole-class discussions. Thirteen 6th- and 7th-grade mathematics teachers working at schools with mainly minority students from low-income families participated in the study. Two lessons with whole-class discussions were observed for each teacher. Rubrics were used to measure the quality of classroom climate, rigor of instruction, and students' interactions. Matsumara et al. (2008) found that the presence of explicit rules in the classroom for respectful, prosocial behavior significantly predicted the number of students who participated in the class discussions. They also found that "the quality of students' participation in class discussions – that is, the degree to which they built on other students' contributions and explained and supported their responses – was predicted by teachers pressing students to explain their thinking in discussions and by the rigor of the questions posed to students in the discussion" (p. 293).

To summarize the reviewed studies on pushing students to explain their thinking in detail, both Webb et al. (2008) and Matsumara et al. (2008) state the importance of making details explicit in student explanations for students' learning.

3.4 CLASSROOM TEACHING: CONCLUSIONS AND DISCUSSION

We conclude with a discussion of the methodology, of our reflections of the review in relation to school practice, and of aspects to consider for future work.

Methodological discussion

Here we discuss some aspects of our methodology for the review in relation to the database search and the categorization process.

Database search

Of course, the choice of search strings in Web of Science is crucial for the results of the review. We aimed for a broad mapping of the research landscape, but a broad search naturally has its limitations in precision. Some studies that would have been relevant for the mapping are always excluded by any choice of search strings. However, we believe that the aim of mapping the research in broad terms has been sufficiently achieved through the search process.

Categorization process

The categorization of the articles has not always been a straightforward and unambiguous process; to another viewer, for instance, some articles could be regarded as belonging to a different category. To raise the reliability of the coding process and strengthen the validity of the coding scheme, at times two people have coded an article independently. We have then compared our coding and resolved the differences. We have also utilized our reading of full-text articles to validate the categorization.

Using abstracts for coding has been a limitation in some cases when the abstracts do not contain enough information for categorization. As a step in our categorization process, in many cases we have read selected parts of the full-text article to be able to determine which category an article belongs to. This is particularly the case for deciding whether the results stay at opportunities for learning or can say something about the relationship between teaching and actual student learning.

Reflections of the review in relation to school practice

From the analysis of the articles that look at taching effects in relation to students' learning of mathematical competencies we distinguished three areas (see the section *Mapping of articles that relate teaching to students' learning of mathematical competencies*, above): 1) Problem-solving teaching, 2) Comparison as an effective teaching strategy, and 3) Pushing students to explain their thinking in detail.

Problem-solving teaching

Cotic and Zuljan's (2009) results show that students who received problem-based instruction were more successful in solving more difficult problems than those who received conventional instruction, with no differences found regarding procedural computation between the approaches. These results are in line with several other studies, such as Boaler (2002), who compared traditional and reform approaches to teaching and their impact on student learning, as well as Hiebert and Wearne (1992, 1996) and Carpenter, Fennema, Peterson, Chiang, and Loef (1989). This means that problem-based instruction can improve students' problem-solving abilities without the expense of procedural ability, which is a highly relevant research finding for school practice.

Attending to nuances in problem-solving teaching can make a big difference in students' learning. We suggest that it is important that teachers be aware and attentive of these nuances and incorporate them into their teaching. Examples from our review are Gresalfi et al.'s (2012) finding that a teacher who, when launching a problem, spent more time clarifying what the students' solutions should look like – i.e. what kinds of arguments are convincing – had students who more often engaged with the problem at higher levels. Of course, there are many nuances in inquiry-oriented problem-solving teaching that are important for teachers to take into account.

An example of the importance of carefully prepared pedagogy in technology use comes from Thorvaldsen et al. (2012), who draw the conclusion that what makes the difference in students' learning in ICT usage is not the ICT per se but the teacher-guided student activity, i.e. the inquiry-guided, problem-solving learning environment, in which ICT may be used. Similarly, Song and Looi (2012) conclude that "technology is only a tool that cannot automatically support learning without a mind for pedagogical innovations" (p. 153). The collaborative inquiry affordances offered by the tool are only used to establish an inquiry-oriented collaborative learning environment if the teacher holds inquiry-oriented beliefs. Related to technology use, Scheiter et al.'s (2010) study shows that letting students use animations can have positive effects on their problem-solving performance for problems of different transfer distances.

Among the studies focusing on problem-solving performance of students with learning problems are those by Jitendra and Star (2011) and Xin et al. (2011). The results of Xin et al. (2011) indicate that that a conceptual model-based approach is more effective than a general heuristic approach. Similar to model-based instruction is schema-based instruction, discussed by Jitendra and Star (2011) as an alternative to traditional instruction. A limitation stated by Xin et al. (2011) themselves is that the mathematical problems in their study were of only two kinds (equal groups and multiplicative compare), and that more complex problems with multiple steps need to be included in future studies. Similarly, Jitendra and Star's (2011) example was limited to only ratio and proportion problems. In our view, a reasonable question to be posed is what the results would be if students received these two kinds of instruction for a large number of different problem types over time? In that case there would be many different models or schemas for the students to handle in the conceptual model-based or schema-based approach. In their handbook chapter, Lesh and Zawojewskij (2007) write:

> The dilemma we see is that, on the one hand, short lists of descriptive processes (i.e. the conventional problem-solving strategies) appear to be too general to be meaningful for instructional purposes; on the other hand, long lists of prescriptive processes (i.e., more detailed and specific problem-solving strategies for classes of problems) tend to become so numerous that knowing when to use them becomes the heart of understanding them (p. 769).

The key question is whether the results from Xin et al. (2011) and Jitendra and Star (2011) would hold if the problem types, or the number of classes of problems in Lesh and Zawojewskij's (2007) wording, were more than just a few.

Krawec et al. (2013) focus on the problem-solving strategy knowledge and reported use by students with learning disabilities as well as average students. They find their own SolveIt! instructional approach effective compared to another solving approach of word-problems in which the teachers mainly posed a worked example whereupon students solved similarly structured problems. However, demonstrating a worked example and letting students solve similar tasks is not mathematical problem-solving as commonly defined in literature. Schoenfeld (1985) states: "If one has ready access to a solution schema for a mathematical task, that task is an exercise and not a problem." (p. 74). This is related to our reasoning above, that the studies of Xin et al. (2011) and Jitendra and Star (2011) are limited to only a few types of problems. It is important to always define what is meant by problem-solving in studying or discussing problem-solving, as emphasized by Schoenfeld (1992).

In relation to school practice in mathematics, it is of great importance for teachers to be aware of what they view as problem-solving. It is further crucial that they be aware of what constitutes a mathematical problem, i.e. that it is in the relation between the task and the individual student. A task that a teacher sets up in a class may hence be a challenging problem for some students and a routine exercise for others. How to handle this in problem-solving teaching is important for teachers to reflect on.

Comparison as an effective teaching strategy

The act of comparing several solutions seems to have the potential of being an effective teaching strategy. All three articles on comparison that were captured in our review (Hattikudur & Alibali, 2010; Rittle-Johnson et al., 2012; Star & Rittle-Johnson, 2009) followed the structure of a pre-test–intervention–post-test design. In these studies, several different kinds of process knowledge are studied – problem-solving ability, conceptual understanding and procedural knowledge. All three have relatively many participants, but are limited in the duration time of the intervention.

The results of these studies indicate that comparing several solution methods may be an effective teaching approach. This can be of relevance for school practice, regarding not only problem-solving lessons and lessons focusing on concepts, but also those focusing on solving procedural tasks. A regular instructional routine for comparing several solutions can lead to greater student achievement and the insight that in mathematics there are always multiple solution paths that lead to the correct answer.

Pushing students to explain their thinking in detail

Note that the results of both Webb et al. (2008) and Matsumara et al. (2008) stress the importance for the quality of student participation that the teacher pushes, or presses, them to explain their thinking in detail. In their review "Mathematics teaching and classroom practice", Franke, Kazemi, and Battey (2007) emphasize the act of making details explicit in classroom discourse. They write: "So one of the most powerful pedagogical moves a teacher can make is one that supports making detail explicit in mathematical talk, in both explanations given and questions asked" (p. 232). Interestingly, Webb et al. (2008) focus on student collaboration in small groups while Matsumara et al. (2008) focus on class discussions. Hence, it seems crucial for the quality of student participation in both small-group and whole-class discussions that the teacher asks productive questions to make details explicit in students' thinking.

It is of great importance for mathematics teachers to be aware of the importance of pushing students to explain their thinking in detail. This is also related to the issue of equity, since making details explicit in explanations makes the explanations more accessible for all students in a class.

Aspects for future work

Here we discuss some aspects to consider for the future work of Swedish research. We begin by discussing the relevance of the mapped research for the Swedish context.

Relevance for the Swedish context

We are aware that many studies captured by our review are from the US context. Specifically, among the 16 studies whose results we draw upon in our review articles that relate teaching to students' learning of mathematical competencies, 12 were conducted in an American context (these are Gresalfi et al.; Hattikudur & Alibali, 2010; Jitendra & Star, 2011; Krawec et al., 2013; Le et al., 2009; Matsumura et al., 2008; Montague, 2008; Orosco, 2013; Rittle-Johnson et al., 2012; Star & Rittle-Johnson, 2009; Webb et al., 2008, and Xin et al., 2011). Still, four studies are conducted in other countries: Germany (Scheiter et al., 2010), Norway (Thorvaldsen et al., 2012), Slovenia (Cotic & Zuljan, 2009), and Singapore (Song & Looi, 2012). We have not appraised the relevance of the reviewed articles for Sweden in relation to the context in which the studies were conducted. It would be interesting to see more studies in the Swedish context focusing on the effects of certain teaching approaches and interactional strategies on students' learning, particularly of mathematical competencies such as reasoning, communication, problem-solving, and conceptual understanding.

Research on effects of teaching on students' learning of mathematical competencies

In their review, Hiebert and Grouws (2007) emphasize that the research field is still seeking reliable recommendations of teaching methods. They state that theories connecting classroom teaching and students'

learning are little developed, and that we need theories to accumulate research findings in order to reveal general principles. It is also methodologically difficult to document the effects of teaching on learning.

In our review, we can see that several studies (e.g. Gresalfi et al., 2012; Matsumara et al., 2008; Webb et al., 2008) are careful about making causal claims regarding relationships between teaching and learning. Some studies (e.g. Orosco, 2014; Star & Rittle-Johnson, 2009) do, however.

Further, in our review, which goes from the year 2008, we can see that among the studies looking at students' learning, fewer focus on students' knowledge of mathematical processes (Proc) compared to products (Prod); see Appendix B. This means that fewer studies focus on students' development of mathematical competencies (cf. Skolverket, 2011) than on mathematical topics like algebra, geometry, etc. Many studies do not look at students' learning but rather focus on teaching practices (P); see Appendix B. These studies hence focus on the opportunities to learn, which is discussed below.

Opportunities to learn

The National Research Council (2001) states that the "opportunity to learn is widely considered the single most important predictor of student achievement" (p. 334). Hiebert & Grouws (2007) add that "this claim says that students learn best what they have the most opportunity to learn" (p. 378), and point out that students' previous knowledge has to be considered in their opportunity to learn.

The teacher's role is very important in creating opportunities for students to learn, although the curriculum of course also affects their learning opportunities. "The emphasis teachers place on different learning goals and different topics, the expectations for learning that they set, the time they allocate for particular topics, the kinds of tasks they pose, the kinds of questions they ask and responses they accept, the nature of the discussions they lead – all are part of teaching and all influence the opportunities students have to learn" (p. 379). Across many empirical studies, Hiebert and Grouws (2007) identify two key features that promote conceptual understanding:

- Teachers and students attend explicitly to concepts
- Students struggle with important mathematics

The first feature concerns making connections among mathematical facts, procedures, and ideas, while the second concerns students trying to make sense of challenging mathematics within reach, as opposed to only practicing what the teacher has demonstrated.

Hiebert and Grouws (2007) believe that these two features are justified by both empirical data and theoretical argument. Further, they state that the features also promote procedural skills and are effective across many contexts and teaching systems. Despite a lack of theories on teaching and methodological difficulties in documenting the effects of teaching on learning, these two key features have emerged.

Research methodologies

In our review it appears that surveys, correlational studies, and case studies constitute a large body of studies investigating interactional strategies, learning materials, and teaching approaches in mathematics. Typically, these methodologies have descriptive aims, with low inference by the researcher. When it comes to intervention studies, we note a trend towards pre- and post-test methodology. Intervention studies have typically prescriptive aims and, thus, may have more to offer when it comes to giving advice for practice. However, here the researcher is often distanced from the classroom. It is the researcher who designs the learning material to be tested, and it is also the researcher who is most active in constructing the test instruments. However, between the test situations, the researcher is not supposed to be very active. Intervention studies, which are based on the method of pre- and post-test, also risk not placing enough focus on the teaching itself; on what actually happens in the classroom and the mechanisms producing and constraining learning outcomes. The teaching of an intervention is explained and evaluated based primarily on tests. In order to develop practice-based and practice-relevant research in mathematics education, we see a need for the research community to broaden the application of research methodologies. We consider the methodology of educational design research (EDR) a promising alternative.

It is claimed that EDR has the potential to bridge the gap between educational practice and theory, because it aims at developing theories about both domain-specific learning and the means that are designed to support this learning (Gravemeijer & Cobb, 2013; Cobb, Confrey, diSessa, Lehrer & Schauble, 2003). EDR thus aims at producing both useful products (e.g., educational materials) and accompanying scientific insights into how these products can be used in education. It is also said to be suitable for addressing complex educational problems that should be dealt with in a holistic way (Bakker and van Eerde, 2015). Educational design resarch can be characterized as research in which the design of educational materials (e.g., computer tools, learning activities, or a professional development program) plays a crucial part in the research. That is, the design of learning environments is interwoven with the developing and testing of theory. Such an integrated perspective on design and theory building distinguishes EDR from intervention studies aiming solely at designing educational materials by testing and improving prototypes.

4.1 CURRICULUM PROGRAMS: INTRODUCTION AND AIM OF THE STUDY

In mathematics education, the term *curriculum* has been used in different meanings (Stein, Remillard, & Smith, 2007). Some have used it to refer to the content of teaching. A related meaning, often adopted by educational decision makers, relates to expectations for instruction as described in policy documents. Many mathematics education researchers and practitioners, however, have used the term to refer to "material resources designed to be used by teachers in the classroom" (Stein et al., 2007, p. 321). In the current review study, we use *curriculum programs* to refer to the student textbooks, the corresponding teacher guides, and any related material that supplements the student textbooks or teacher guides[3].

Our decision to zoom in on mathematics curriculum programs is informed by research stressing their wide use in classrooms (e.g., Grouws, Smith, & Sztajn, 2004; Mullis, Martin, & Foy, 2008; Pepin & Haggarty, 2001), and the fact that they are regarded as a crucial link between national curricula and teaching practice (Pepin, Gueudet, & Trouche, 2013; Valverde, Bianchi, Wolfe, Schmidt, & Houng, 2002). As such, they are powerful tools for impacting teaching practice and student learning. It should thus not be a surprise that many reformers have emphasized curriculum programs as an attempt to promote improvement in teaching mathematics (Remillard, 2005). Very often, the idea has been that experts have tried to improve instruction through innovative curriculum programs, which the teachers would use as intended by the developers. This corresponds with the commonly held assumption that teachers are merely conduits of curriculum, simply delivering the curriculum to students (Clandinin & Connelly, 1992). Meanwhile, it has been shown that a main cause of failure of many of these reform efforts was a disregarding of the central role of teachers in teaching mathematics (Remillard, 2005). Moreover, there is empirical evidence that the introduction of human interaction with curriculum programs influences the way lessons are taught and, consequently, influences students' learning opportunities (e.g., Freeman & Porter, 1989; Remillard & Bryans, 2004; Stein, Grover, & Henningsen, 1996). These findings stress the need to characterize the relationship between teachers and curriculum programs as interactional (Brown, 2009; Lloyd, Remillard, & Herbel-Eisenman, 2009; Taylor, 2013).

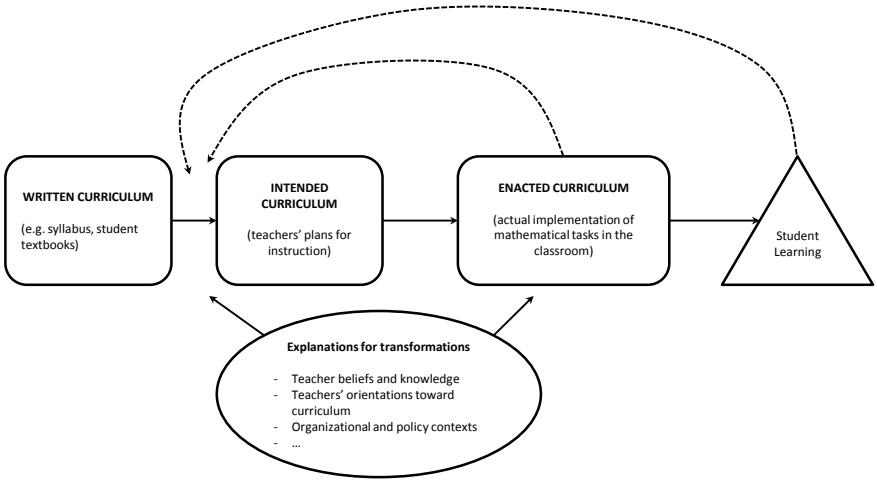
A comprehensive review study by Stein et al. (2007), on how curriculum[4] influences student learning shares a similar focus, and also stresses that no curriculum is self-enacting. The review study is published in the *Second Handbook of Research on Mathematics Teaching and Learning*, highly influential within mathematics education research. To our knowledge, the review study is unique in its kind, given its comprehensiveness and focus on mathematics education. It should thus not be a surprise that many other related studies have referred to it. Given the similarity in focus, the comprehensiveness, and the fact that the review study is frequently cited, we assume that important issues for the current review study are to be found in it. Therefore, we continue from this review both in searching more recent publications and by focusing on important areas that were put forward for future research.

Stein et al. (2007) have conceptualized the "unfolding" of the curriculum in a series of temporal phases, distinguishing between a written (e.g., syllabus, textbooks), intended (teachers' plans for instruction), and enacted curriculum (actual implementation of mathematical tasks) (See Figure 2). The conceptual framework as depicted in Figure 2 illustrates that student learning opportunities are influenced by teachers' interpretations and use of curriculum programs to plan instruction (i.e., intended curriculum) and by how these plans are

[3] At times we use curriculum, curriculum materials, and curriculum programs interchangeably. This is done when studies use these terms in a way that corresponds with our description of curriculum programs.

[4] Stein et al. (2007) use the term curriculum in a broad way to refer to curriculum programs, the curriculum goals intended by the teacher, and the curriculum as enacted in the classroom.

enacted in the classroom (i.e., enacted curriculum). The transformations in the curriculum are influenced by characteristics of the teachers (e.g., teacher knowledge), context, and curriculum.



Figur 2. Temporal phases of curriculum use (based on Stein at al., 2007, p. 322)

The conceptual framework depicted in Figure 2 illustrates that the influence of curriculum programs on teaching practice and on student learning is not straightforward. Along these lines, Stein et al. (2007) made recommendations for future research, which we argue are important for teaching practice given that one cannot merely impose a curriculum program to improve mathematics classroom practices. *First*, as teachers are increasingly expected to design mathematical classroom practices rather than faithfully following the guidelines in the curriculum programs, the research field has initiated attempts to come to a better understanding of educative curriculum programs – curriculum programs that support teacher learning in addition to student learning. More related research is needed, including how teachers respond to features of educative curriculum programs. *Second*, we need to better understand how teachers prepare for teaching, how they use the curriculum programs in their preparation, and how the curriculum programs can support them in planning for the lesson. *Third*, further research is needed to relate influencing factors that explain transformations between and within the phases of curriculum use. *Fourth*, there is a need for more research describing the effectiveness of curriculum programs. Table 5 lists these four areas for future research.

Tabell 5. Selected areas for future research, based on Stein et al. (2007)

Studies that …
1. Describe design principles of educative curriculum programs, and how teachers respond to or use these features.
2. Describe how teachers prepare for teaching.
3. Relate influencing factors that explain transformations between and within the phases of curriculum use. Particularly, studies that:
3.1. Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.
3.2. Combine the inclusion of schools and districts as settings for curriculum implementation.
4. Describe the effectiveness of curriculum programs. Particularly, studies that:
4.1. Describe the effectiveness of curriculum programs and provide data on the degree of implementation.
4.2. Describe what and how students learn from interacting with particular curriculum programs, rather than describing how much they learn.

In the current review, we aim to map the research terrain related to mathematics curriculum programs since the Stein et al. (2007) review. More particularly, we aim to provide an overview of research articles, published at

the earliest in 2008, focusing on curriculum programs in Grades 1-9, or in preservice teacher education. Second, once we have provided this overview – or map – we aim to zoom in on a subset of the mapped articles explicitly addressing the four selected areas for future research, as listed in Table 5, to exemplify the research field's current understanding concerning these areas. In sum, we aim to:

1) Map the research since 2008 on curriculum programs in mathematics education for Grades 1-9 or in preservice teacher education.
2) Exemplify the research field since 2008 with regard to selected areas for future research retrieved from Stein et al., 2007 (See Table 5).

In the following sections, we will describe the search procedure and the screening, coding and mapping of the research field.

4.2 CURRICULUM PROGRAMS: SEARCH STRATEGY, SCREENING AND CODING

Table 6 lists the steps undertaken in the search strategy. As a first step we searched for research articles indexed in the Social Sciences Citation Index of the Web of Science Core Collection (WoS). A search carried out on October 3, 2014, required that a) research articles include "mathematics" in their title, abstract or keywords, or in the journal's name, and b) the title, abstract or keywords mentioned "curriculum" or "curricula", "classroom material(s)", "textbook(s)", "teacher('s) guide(s)", "artifact(s)" or artefact(s)". As an additional requirement, research articles had to be listed by the WoS as articles or review articles, and had to be written in English. This resulted in a total of 1,100 research articles.

Tabell 6. Overview of search strategy

Step	Number of retained research articles
1. WoS search	1100
2. Grouping of journals according to WoS journal categories	1100
3. Selection of research articles published in most influential journals within each category	675
4. A. Screening of abstracts of research articles	245
B. Coding of abstracts of selected research articles	62
5. Selection of research articles, based on the criteria mentioned in Table 5	15

In the second step, as mentioned in Table 6, we grouped the research articles according to categories that WoS applies for grouping the journals. Given our aim to provide a comprehensive map of the related research terrain, we considered it important not to delete any category by definition. See Table 7 for an overview of the journal categories that resulted from this step.

In the third step, for each journal category we decided to focus on research articles published in journals with a high impact factor. Therefore, we retained those that were published in journals with an impact factor higher than or equal to 1 (JCR Social Sciences Edition 2013). In addition, following our aim to provide a broad overview, we ensured we selected articles from at least three different journals of each category. For instance, for the category *Mathematics education* we also included research articles from *Educational Studies in Mathematics* (IF: 0.639) and *Mathematical Thinking and Learning* (IF: 0.385), in addition to research articles from *Journal for Research in Mathematics Education* (IF: 1.000). Following this procedure, we ended up with 675 research articles[5] to be considered for screening and coding.

The fourth step was related to the selection, or screening, of research articles to meet the first research aim. Of the 675 research articles retrieved in Step 3, we decided to select research articles that:

a. Focus on mathematics education
b. Focus on curriculum programs
c. Deal with issues related to elementary school, lower secondary education, or teacher education
d. Do not deal primarily with professional development. However, if professional development accompanied the implementation of a curriculum program, the research article might still be included.

[5] Ultimately, we did not screen and code research articles published in *Curriculum Matters* (n = 6) and *The Elementary School Journal* (n = 14) as these journals were not at our disposal when this review was carried out.

For some research articles we based our decision on a reading of the title; however, for most of the research articles we read the abstract and keywords in addition to the title in order to make a decision. As Cobb and Jackson (2008) reported limitations in developing implications for policy and practice in mathematics education by relying exclusively on research conducted using a single methodology (for instance, relying solely on experimental research studies), we did not include particular selection criteria concerning the method of study.

Tabell 7. Overview of journal categories with corresponding number of research articles [a,d]

Journal category	**Total number of research articles considered for screening and coding**[b]
Science and science education*	106 (151)
Mathematics education*	106 (107)
Psychology	84 (99)
Psychology educational	80 (109)
General education*	55 (149)
Mathematics & science education*	44 (44)
Education special	35 (76)
Curriculum research*	29 (29)
Teacher education*	24 (58)
Computer science interdisciplinary applications	18 (18)
Evaluation, economics, policy, management, effectiveness*	15 (33)
History philosophy of science	14 (27)
Economics	13 (23)
Globalization, development, comparison, social justice, international education*	12 (22)
Review*	11 (11)
Technological resources*	10 (25)
Public environmental occupational health	10 (12)
Sociology	8 (15)
Information science library science	8 (14)
Multidisciplinary sciences	8 (11)
Mathematics interdisciplinary applications	7 (13)
Nursing	6 (10)
Linguistics	6 (23)
Social sciences interdisciplinary	5 (10)
Other WoS categories[c]	*56*

[a] See Appendix D for a description of the procedures applied to construct these categories.
[b] Based on Steps 1-3 of the selection strategy (See Table 6). Numbers in parentheses refer to the initial number of research articles per category prior to Step 3 of the search strategy. The numbers within and outside the parentheses add up to more than 1,100 and 675, respectively, due to some overlap between categories. Categories within the WoS category *Education and educational research* are marked with an asterisk (*).
[c] This category groups the remainder of the WoS categories other than the above-listed categories.
[d] Categories shaded in grey have been screened, coded and mapped.

Due to time constraints, we decided to screen and code abstracts from research articles in the highly relevant categories *Curriculum research*, *Mathematics education*, *Teacher education*, and *General education*. In addition, in line with our aim to provide a comprehensive overview, we screened and coded abstracts from research articles in other categories as well. The category *Economics* was chosen as we assumed that research articles in this category might provide information about the effectiveness of curriculum programs. The category *Sociology* was chosen as we assumed that it might include information about factors explaining transformations between and within phases of curriculum use in a broader setting such as a school or district. We included the category *Technological resources* because we expected it to be potentially informative concerning the evolution from paper-written curriculum programs to digital curriculum programs. As a result, in Step 4, 245 abstracts of research articles were screened (Step 4A in Table 6), of which 62 were selected for coding based on selection criteria a-d. The coding of these 62 abstracts (Step 4B; see Appendix E for the coded abstracts) related to the a) object of study, b) method, c) number of participants, d) context of the study, e) reported results and conclusions, and d) reported relevance. This served as the input for meeting the first research aim.

In the fifth step, based on the criteria listed in Table 5 we selected a subsample of 15 research articles from the 62 articles retrieved in Step 4. These 15 articles were then completely read and summarized (See Appendix F). This served as input for meeting the second research aim.

4.3 CURRICULUM PROGRAMS: MAP AND RESULTS

This section consists of two parts. In the *first* – aiming to address the first research aim – we describe the research terrain related to mathematics curriculum programs in Grades 1-9, or in teacher education since 2008. We do this based on our coding of the 62 abstracts of retrieved research articles (recall that these were selected based on a screening of abstracts of 245 research articles). An overview of the 62 coded research articles can be found in Appendix E.

In the *second* part – aiming to address the second research aim – we read a sample of these 62 research articles completely to exemplify the research field's state concerning 1) the design and use of educative curriculum materials, 2) teachers' use of curriculum programs in planning for teaching, 3) influential factors explaining transformations between and within the phases of curriculum use, and 4) the effectiveness of curriculum programs. Appendix F contains a summary of each of the selected research articles that were read to address the second research aim.

Research aim 1: Map the research on curriculum programs in mathematics education for Grades 1-9 or in preservice teacher education since 2008

In this section, based on our reading of abstracts of research articles for each selected journal category (Mathematics education, Curriculum research, Teacher education, General education, Economics, Sociology, Technological resources), we describe what research has focused on with regard to mathematics curriculum programs in Grades 1-9 or in teacher education since 2008. Table 8 provides a distribution of research articles per journal category according to the specific research focus as evident in the abstract.

Some general observations can be made based on the numbers in Table 8. *First*, and not surprisingly, most relevant research articles were found in the journal categories *Mathematics education*, *Curriculum research*, and *General education*. Rather surprisingly is the small number of research articles found in the *Teacher education journals.* This could be indicative of a lack of association between (the use of) mathematics curriculum programs and teacher education research. *Second*, one can see that three research foci reoccur frequently: 23 (37%) research articles focus on aspects contributing to the unfolding of the curriculum; 17 (27%) focus on the effectiveness of curriculum programs; and 13 (21%) focus on textbook analyses. Also, three research foci are addressed to a lesser extent: five (8%) articles focus on the inclusion of broader settings such as schools and districts when studying curriculum programs; five (8%) relate to educative curriculum programs, and three (5%) describe teachers' preparations for teaching. This suggests that more related research could be especially beneficial in relation to the latter three research foci. Further, two (3%) articles have a research focus other than the abovementioned foci. A *third* observation relates to the two research articles in the *Economics journals*. In our sample, both research articles are unique in their combined focus on the effectiveness of curriculum programs and the inclusion of broader settings such as schools, districts and municipalities. We observed only one other research article that did likewise. A *fourth* remark relates to the context of study: 44 of the 62 articles (66%) relate exclusively to a US context, whereas only seven (11%) relate exclusively to a context other than the US. Also, seven articles (11%) relate to multiple contexts, and we found no related information in the abstract for seven other articles (11%).

In the following sections, we describe each of the 62 research articles in more detail per journal category, based on our reading of their abstracts. In the final section (*Overview of research on curriculum programs in mathematics education for Grades 1- 9 or in preservice teacher education, since 2008 based on reading of abstracts*), we summarize the research for all the journal categories.

Tabell 8. Distribution of research articles per category and research focus, based on reading of abstracts

	Research focus								
Category	Textbook analysis	Effective-ness of curriculum programs	Unfolding of curriculum, including mediating variables	Preparing for teaching	Inclusion of schools and districts	Educative curriculum programs	Other	Total	Context of study
Mathematics education journals	12	11	9	1			1	34	US: 20 Other: 3 Multiple: 6 No info: 5
Curriculum research journals	1		8	1				10	US: 9 Other: 1 Multiple: / No info: /
Teacher education journals			4			1		4	US: 1 Other: 2 Multiple: / No info: 1
General education journals		3	1	1	2	4	1	9	US: 9 Other: / Multiple: / No info: /
Economics journals		2			2			2	US: 1 Other: 1 Multiple: / No info: /
Sociology journals					1			1	US: 1 Other: / Multiple: / No info: /
Technological resources journals		1	1					2	US: / Other: / Multiple: 1 No info: 1
Total	**13**	**17**	**23**	**3**	**5**	**5**	**2**	**62***	**US: 41** **Other: 7** **Multiple: 7** **No info: 7**

Note. Sum of total of columns adds up to more than 62, since some research articles stress multiple research foci in their abstract.

Category: Mathematics education journals

The screening of research articles in the category *Mathematics education journals* resulted in a selection of 16 research articles published in *Journal for Research in Mathematics Education,* 13 in *Educational Studies in Mathematics* and five in *Mathematical Thinking and Learning,* ultimately totaling 34 selected articles for which we coded the abstracts (see Appendix E).

Of these 34 articles, 12 are textbook analyses, of which five are comparative studies comparing different aspects of American textbooks to textbooks from eastern Asian countries, such as Japan (Alajmi, 2012), Korea (Hong & Choi, 2014; Son & Senk, 2010) and China (Sun, 2011). One study compares textbooks in Cyprus, Ireland and Taiwan (Charalambous, Delaney, Hsu, & Mesa, 2010).

Objects of study in these 12 research articles vary between particular content areas of mathematics such as fractions (Alajmi, 2012; Charalambous et al., 2010; Son & Senk, 2010) or quadratic equations (Hong & Choi, 2014), and more general aspects like reasoning and proof (Otten, Gilbertson, Males, & Clark, 2014; Stacey & Vincent, 2009; Stylianides, 2009; Thompson, Senk, & Johnson, 2012), constructing deep knowledge (Shield & Dole, 2013), or representations that facilitate the transition from concrete to abstract thinking (Ding & Li, 2014). One study explicitly reviews the content covered in 141 elementary textbooks published between 1900 and 2000 in the US (Baker et al., 2010).

Eleven research articles describe the effectiveness of curriculum programs. One focuses on the impact of challenging geometry and measurement units on students' achievement (Gavin, Casa, Adelson, & Firmender,

2013). Another focuses on the joint implementation of an ambitious curriculum program and a block schedule (Kramer & Keller, 2008). Yet another, large-scale, study analyzes the association between two reform curriculum programs and student performance (Post et al., 2008). As an inclusion criterion, students needed to have followed the same curriculum for at least three years. It was found that students learned traditional topics, but that achievement in problem solving and open-ended subtests was greater than on a procedures subtest. In an article on the use of calculators by students with and without disabilities, it was found that type of curriculum (traditional versus reform) did not play a significant role in the frequency and efficiency of their use of the calculator (Bouck, Joshi, & Johnson, 2013).

We also retrieved a number of studies that explicitly included implementation of the curriculum programs in their scope. One such article analyzes the processes and outcomes of implementing proof-related tasks (Bieda, 2010). Two large-scale articles focus explicitly on the impacts of curriculum and curriculum implementation on student achievement (Grouws et al., 2013; Tarr et al., 2008). Finally, one article focuses explicitly on curriculum implementation and describes the various ways curriculum implementation has been measured (Huntley, 2009).

Two large-scale articles report on associations between high school curriculum programs and achievement in postsecondary education (Harwell, Post, Medhanie, Dupuis, & LeBeau, 2013; Post et al., 2010). These were retrospective studies, and did not include in-depth implementation information. One article reports on associations between middle school curriculum programs and the problem posing skills of high school students (Cai et al., 2013).

Nine articles focus on changes in the curriculum as it unfolds throughout the temporal phases, and describe mediating factors that explain these transformations. Four articles focus explicitly on the interaction between teacher and curriculum program: In the context of whole-number lessons, Brown et al. (2009) explore how curriculum programs support and hinder teachers, and how teachers impact the enactment of the curriculum program. Lloyd (2008) zooms in on a student teacher's interactions with a curriculum program, and finds that features of the student-teaching placement, the student teacher's personal background, and characteristics of the curriculum program contributed to how this particular student teacher made use of the curriculum program. An article by Choppin (2011) describes how teachers' attention to student thinking influences the adaptations the teacher makes to the written curriculum. Gresalfi et al. (2012) studied two teachers using identical curriculum materials but differing significantly in their personal histories of teaching and instructional practices.

One article focuses on how affective stances toward the curriculum program might influence enactment: Charalambous and Philippou (2010) studied teachers' concerns and efficacy beliefs as affective responses to mathematics curriculum reforms.

One article (Ellerton, 2013) centers on engaging student teachers in problem posing activities, and argues that this is a useful way to redesign curriculum programs in order to do more than pay lip service to problem posing – a frequently mentioned goal in national curricula.

Two articles (Jansen, Herbel-Eisenmann, & Smith, 2012; Star, Smith, & Jansen, 2008) describe how students experienced reform curriculum programs, thereby possibly adding to our knowledge about another variable that might influence a curriculum program's impact on classroom practices: students' views of the curriculum program. Yet another article points at the importance of considering the needs of individual schools when implementing new curricula (Meaney, Trinick, & Fairhall, 2013).

In a study of how exemplary Chinese and US teachers view instructional coherence and how they see possible ways to achieve instructional coherence it was found that, whereas this was not mentioned by the US teachers, Chinese teachers stressed the importance of studying the textbooks to understand the curriculum authors' intentions and how they connected to prior and later pieces of knowledge (Cai, Ding, & Wang, 2014). As such, this study sheds light on how the two groups of teachers might differ in their planning for teaching.

One article reviews curricular research in mathematics education and aims to highlight advances related to curricular research, and to point at issues and areas that merit further exploration (Confrey et al., 2008).

Category: Curriculum research journals

The screening of the research articles in the category *Curriculum research journals* resulted in ten selected articles: nine published in *Journal of Curriculum Studies* and one in *Curriculum Inquiry*. Six of the articles focus on the individual and joint contribution of teachers' mathematical knowledge for teaching and curriculum programs to the mathematical quality of instruction (Charalambous & Hill, 2012; Charalambous, Hill, & Mitchell, 2012; Hill & Charalambous, 2012a, 2012b; Lewis & Blunk, 2012; Sleep & Eskelson, 2012)[6]. In doing so, this group of articles attends to an integrated analysis of curriculum resources and teacher resources, and how they interact. As such, these studies describe influencing factors that might explain transformations between the phases of curriculum use and reveal information about the enactment of curricula.

One study (Van Steenbrugge, Valcke, & Desoete, 2013) focuses on teachers' view of curriculum programs as a potentially mediating variable in the process of curriculum enactment, and thus also deals with an influencing factor to explain transformations in the curriculum. Another study (Choppin, 2009) focuses on how teachers learn from successive use of their curriculum program, which reveals information about mediating factors in the process of curriculum implementation and provides information on the enactment of curricula.

Sherin and Drake (2009) develop a curriculum strategy framework to describe teachers' use of the curriculum by focusing on how and when teachers read, evaluate and adapt the curriculum program, and in doing so, provide a description of how teachers prepare for teaching.

A final study relates to an analysis of the language about children throughout lesson conversations, in mathematics textbooks and standards documents (Parks, 2010).

Category: Teacher education journals

The screening of the research articles in the category Teacher education resulted in four selected articles: three published in *Teaching and Teacher education* and one in *European Journal of Teacher Education*. The four selected articles all focus on changes in the curriculum as it unfolds throughout the temporal phases and aim to describe factors that explain these transformations. A study by Davis et al. (2011) describes the resources on which the teachers draw when adapting the curriculum, and how this relates to particular curriculum adaptations. In doing so, the study has implications for teacher education, professional development and educative curriculum materials. In Stylianides and Stylianides (2008), an analytical framework is presented to describe the implementation of highly demanding mathematical tasks. Two studies include the school setting in their analysis of curriculum implementation. A study by Marz and Kelchtermans (2013) focuses on capturing the complexity of implementing educational innovations by describing how individual and collective processes of sense-making, as well as structural factors and processes in the school, mediate implementation practices. Haser (2010) finds that the Turkish national curriculum context interfered with the school and classroom contexts, often by presenting many additional difficulties.

Category: General education journals

The screening of research articles in the category *General education journals* resulted in nine selected articles: three in *Educational Researcher*, four in *American Educational Research Journal*, one in *American Journal of Education*, and one in *Harvard Educational Review*.

Three articles deal with the effectiveness of curriculum programs. A study by Roschelle and colleagues (2010) evaluates the impact of replacement units on student learning of advanced mathematics in middle school. The replacement units integrate an interactive representational technology and paper curriculum joined by teacher professional development. The study, replicated in three settings with varied teacher and student characteristics, reveals that the replacement units are effective throughout the three settings. This indicates that the replacement units are effective in enabling a variety of teachers in a diversity of settings to extend student

---

[6] The six research articles are published in a special issue of *Journal of Curriculum Studies*.

learning to more advanced mathematics. A second study analyzes the impact of participation in either reform-oriented or traditional curriculum programs in high school on the difficulty level of the first mathematics course students enroll in (Harwell et al., 2009). The findings indicate that reform-oriented curriculum programs do not prepare students for initially enrolling in more difficult university mathematics courses as well as traditional curriculum programs. The findings also reveal that, once enrolled, students earn similar grades. In a third study, Porter, Polikoff, Barghaus and Yang (2013) report about an algorithm, based on surveying the enacted curriculum, to construct tests aligned with, for instance, curriculum programs. As such, the study is informative with regard to measuring the effects of adopted curriculum programs.

Four articles relate to educative curriculum programs, and two of these seem to have an explicit focus on educative curriculum programs. Davis and colleagues (2014) report on a design process of educative curriculum programs that is both theoretically and empirically driven. Focusing on how prospective teachers can learn to read and use educative curriculum materials, Drake, Land, and Tyminski (2014) present a set of empirically based design principles for using educative curriculum programs in order to support the development of prospective teachers' knowledge. A study by Stein and Kaufman (2010) found that quality of instruction – measured in terms of levels of cognitive demand, attention to student thinking, and mathematical reasoning – was better when teachers were provided with greater support in understanding the big mathematical ideas of a lesson, implying that such support is one way curriculum programs can be educative for the teacher. Finally, Sleep's (2012) study of steering instruction toward the mathematical point has implications for the design of educative curriculum programs.

Two articles have an explicit focus on implementing curriculum programs at scale. Stein and Kaufman (2010) focus on how teacher capacity and teachers' use of curriculum programs influence the instructional quality. Their findings indicate that teacher capacity across school districts was not correlated with instructional quality, whereas teachers' lesson preparation that took into account the big mathematical ideas did correlate with instructional quality. The study also connected the support provided by curriculum programs in helping teachers understand the big mathematical ideas with the instructional quality of the lesson. In a related article, Stein and Coburn (2008) illustrate how different district reform efforts, centered on the adoption of curriculum programs, result in differing learning opportunities for teachers.

An article by Sleep (2012) describes what it takes to identify the goals of mathematical instruction and to use these goals to manage the teaching. The study identifies several central tasks for steering instruction toward the mathematical point, and as such adds to our understanding of the unfolding of the curriculum and contains information on teachers' planning for instruction.

One article questions the assumption, made in *Foundations for success: the final report of the national mathematics advisory panel*, that only experimental research studies can produce scientific evidence (Cobb & Jackson, 2008). The authors describe the limitations of developing implications for policy and practice by relying exclusively on research conducted using a single methodology.

Category: Economics journals

The screening of research articles in the category *Economics journals* resulted in two selected articles, both published in *Economics of Education Review*. Both articles describe the effectiveness of curriculum programs by including broad settings for curriculum implementation, such as municipalities and school districts. In their study, Bhatt, Koedel, and Lehmann (2013) describe that the effectiveness of curriculum programs might differ according to the specific subtopic, which holds implications in the decision process for certain curriculum programs. This suggests that one should take into account the topical performance for various curricular alternatives. Leme, Louzano, Ponczek, and Souza (2012) studied the impact on mathematics achievement of a particular method, which includes the structuring of curriculum content, the development of teacher and student textbooks, and the training and supervision of teachers and instructors. They found that the 4^{th}- and 8^{th}-grade students in municipalities that adopted the method performed better in mathematics than those in municipalities that did not.

Category: Sociology journals

One research article, published in *Sociology of education*, was selected after our screening of the *Sociology journals.* As teachers' social networks can be influential in teacher learning and organizational change, Coburn, Mata, and Choi (2013) study why some teachers have networks that are likely to support individual and organizational change, while others do not. They draw on a longitudinal, qualitative study of the implementation of a mathematics curriculum programs at four schools. Their findings show that district policy influenced the structure of networks and the benefits accrued through network exchanges, and introduced interaction routines that interrupted the conventional ways teachers talked together. As such, by including broader settings in the study of curriculum implementation, this study sheds light on how (district) policies can influence the implementation of curriculum programs.

Category: Technological resources journals

Two research articles, published in *British Journal of Educational Technology* and *Journal of Computer Assisted Learning*, were selected after screening the *Technological resources journals.* Although both articles deal with science education and not directly with curriculum programs, we believe they contain potentially interesting information relevant to the digitalization of curriculum programs. Kim and Olaciregui (2008) test the effects of a concept map-based electronic portfolio system in a 5^{th}-grade science class. They found that the group of students working with the electronic portfolio system performed better on both an information-processing performance test and a three-day delayed memory retention test. Thus, although not directly related to mathematics education and curriculum programs, this study reveals how an electronic mathematics curriculum program could be designed, and points at the potential benefits.

Building on previous findings concerning mathematics and science teachers' use of ICT, Voogt (2010) explores differences between extensive and non-extensive ICT-using science teachers with respect to pedagogical orientation, ICT competencies and professional engagement. The study reveals that both extensive and non-extensive ICT-using science teachers' pedagogical orientation reflected traditional as well as lifelong learning curriculum goals and practices. The two groups differed, however, in pursuing curriculum goals and orientation, ICT competencies and professional engagement: extensive ICT-using teachers reflected these to a higher extent than did non-extensive ICT-using teachers. We believe that these findings are relevant, as they shed light on potentially influential teacher factors in the digitalization of mathematics curriculum programs.

Overview of research on curriculum programs in mathematics education for Grades 1-9 or in preservice teacher education, since 2008 based on reading of abstracts

In this section, we summarize research related to, or relevant to, curriculum programs published since 2008. In line with the sections above, we do this per journal category. Most of the 34 retrieved research articles in the *Mathematics education journals* relate to textbook analyses ($n = 12$), the effectiveness of curriculum programs ($n = 11$) or changes in the curriculum as it unfolds throughout the temporal phases of curriculum use ($n = 9$).

Textbook analyses often compare textbooks from different countries – usually the US and Asian countries (e.g., Alajmi, 2012; Son & Senk, 2010). Objects of study vary from particular content areas such as fractions (e.g., Charalambous et al., 2010) or quadratic equations (Hong & Choi, 2014) to more general thinking skills such as reasoning and proof (e.g., Otten et al., 2014; Thompson et al., 2012) and the construction of deep knowledge (Shield & Dole, 2013).

Research articles describing the effect of ambitious, reform-oriented, curriculum programs on student performance often do this by comparing the performance of students working with (units) of ambitious curricula to that of students working with more traditional curricula (e.g., Gavin et al., 2013; Kramer & Keller, 2008). Some of the articles explicitly address both the effect of the curriculum program and that of its actual implementation (Grouws et al., 2013; Tarr et al., 2008). Some describe associations between high school curriculum programs and achievement in postsecondary education (Harwell et al., 2013; Post et al., 2010).

Further, regarding changes in the curriculum as it unfolds throughout the temporal phases of curriculum use, a number of articles explicitly address the interaction between teacher and curriculum program (e.g., Brown et

al., 2009; Lloyd, 2008). In addition, two articles focus on students' stances regarding reform-oriented curriculum programs (Jansen et al., 2012; Star et al., 2008).

Of the ten research articles in the *Curriculum research journals*, six focus explicitly on how teacher resources and curriculum resources interact, and how this might impact instruction (Charalambous & Hill, 2012; Charalambous et al., 2012; Hill & Charalambous, 2012a, 2012b; Lewis & Blunk, 2012; Sleep & Eskelson, 2012). One article sheds light on teachers' views of curriculum programs which might mediate the enactment of curriculum programs (Van Steenbrugge et al., 2013). Another article describes how teachers learn from successive enactment of curriculum programs (Choppin, 2009). One article also develops and presents a framework to capture teachers' use of curriculum programs (Sherin & Drake, 2009). Finally, one article analyzes language about children in curriculum programs as well as in other documents and environments (Parks, 2010).

The four selected articles in *Teacher education journals* focus on changes in the curriculum as it unfolds throughout the temporal phases, and aim to describe factors that explain these transformations. One article explicitly describes the resources on which the teachers draw when adapting the curriculum, and how this relates to particular curriculum adaptations (Davis et al., 2011). Stylianides and Stylianides (2008) present an analytical framework to describe the implementation of high demanding mathematical tasks. Two articles also include the school setting in their analysis of curriculum implementation, and in doing so exemplify the complexity associated with implementing curricular innovations (Haser, 2010; Marz & Kelchtermans, 2013).

Most of the articles in the *General education journals* relate to the effectiveness of curriculum programs (n = 3), educative curriculum programs (n = 4), or the implementation of curriculum programs at scale (n = 2). One article describes the impact of replacement units (integrating an interactive representational technology, paper curriculum and teacher professional development) on student learning of advanced mathematics in middle school (Roschelle et al., 2010). Replications of the study have revealed that these replacement units are effective across a variety of settings. A second article, analyzing the impact of participation in either reform-oriented or traditional curriculum programs on postsecondary education, finds that reform-oriented curriculum programs do not positively affect participation in more difficult university mathematics courses, but that once enrolled, students score similarly (Harwell et al., 2009). A third article describes a procedure for improving tests' alignment with, for instance, the curriculum program used in class (Porter et al., 2013).

As for educative curriculum programs, a first article describes a theoretically and empirically driven design process of educative curriculum programs (Davis et al., 2014). A second article presents a set of design principles for using educative curriculum programs in teacher education to support prospective teachers' development of knowledge (Drake et al., 2014). A third article connects the curriculum program's support in identifying the big mathematical ideas of the lesson to the quality of the enacted lesson (Stein & Kaufman, 2010). An article about steering instruction toward the mathematical point has implications for the design of educative curriculum programs (Sleep, 2012).

Two articles focus explicitly on implementing curriculum programs at scale. Stein and Kaufman (2010) find that, rather than teacher capacity, teachers' orientation to the big mathematical ideas in the lesson to be enacted had an impact on the quality of the lesson. In a related article, Stein and Coburn (2008) illustrate how different district reform efforts, centered on the adoption of curriculum programs, result in different learning opportunities for teachers. In both articles, findings are connected to the design of the curriculum programs. One article relates to implications for policy and practice, focusing solely on experimental studies (Cobb & Jackson, 2008).

The two selected articles in the category *Economics journals* describe the effectiveness of curriculum programs by including broad settings for curriculum implementation such as municipalities and school districts. Bhatt, Koedel, and Lehmann (2013) suggest that one should take into account the topical performance for various curricular alternatives. Leme, Louzano, Ponczek, and Souza (2012) find that students in municipalities that adopted methods including the structuring of curriculum content, the development of teacher and student textbooks, and the training and supervision of teachers and instructors outperformed students in municipalities that did not adopt the method.

The article selected from the *Sociology journals* studied how district policy affects the formation of social networks among teachers implementing a new curriculum program (Coburn et al., 2013). Their study showed

that district policy influences the structure of networks and the benefits accrued through network exchanges, and introduced interaction routines that interrupted the conventional ways teachers talked together.

The two selected articles in the *Technological resources journals* both relate to science education, but were selected because of their potential relevance concerning the digitalization of curriculum programs. Kim and Olaciregui (2008) describe and test the effects of a concept map-based electronic portfolio system, and find that students working with the electronic portfolio system scored higher on an information-processing performance test and a memory retention test. As such, this article reveals how an electronic mathematics curriculum program might be designed, and points to the potential benefits. The article by Voogt (2010) sheds light on how teachers' pedagogical orientations, ICT competencies and professional engagement might influence the use of ICT, and in doing so, points to potentially influential teacher factors in the digitalization of mathematics curriculum programs.

Research aim 2: Exemplify the research field, since 2008, with regard to selected areas for future research retrieved from Stein et al., 2007

Table 9 lists the research articles selected for a complete reading to exemplify the research field's state concerning four areas: 1) The design and use of educative curriculum materials, 2) Teachers' use of curriculum programs when planning for teaching, 3) Influencing factors that explain transformations between and within the phases of curriculum use, and 4) The effectiveness of curriculum programs. We describe each of these areas in more detail, based on our reading of the complete articles. At the end of this section, we also summarize these findings (See *Summarizing research*).

Tabell 9. Overview of research articles to exemplify the research field's state concerning selected areas for future research (based on Stein et al., 2007)

	Studies that:								
	1. Describe design principles of educative curricula - these are curricula that are also educative for teachers - and how teachers respond to or use these features.	2. Describe how teachers prepare for teaching.	3. Relate influencing factors that explain transformations between and within the phases of curriculum use. Particularly studies that:	*3.1. Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.*	*3.2. Combine the inclusion of schools and districts as settings for curriculum implementation.*	4. Describe the effectiveness of curriculum programs.	*4.1. Describe the effectiveness of curriculum programs and provide data on the degree of implementation (observations).*	*4.2. Describe what and how students learn from interacting with particular curriculum programs, rather than how much they learn.*
Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247-280.							**X**	*x*

Forts. Tabell 9

	Studies that:								
	4. Describe design principles of educative curricula - these are curricula that are also educative for teachers - and how teachers respond to or use these features.	5. Describe how teachers prepare for teaching.	6. Relate influencing factors that explain transformations between and within the phases of curriculum use. Particularly studies that:	*4.1. Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.*	*3.3. Combine the inclusion of schools and districts as settings for curriculum implementation.*	5. Describe the effectiveness of curriculum programs.	*5.1. Describe the effectiveness of curriculum programs and provide data on the degree of implementation (observations).*	*5.2. Describe what and how students learn from interacting with particular curriculum programs, rather than how much they learn.*
Brown, S. A., Pitvorec, K., Ditto, C., & Kelso, C. R. (2009). Reconceiving Fidelity of implementation: An Investigation of Elementary Whole-Number Lessons. *Journal for Research in Mathematics Education, 40*(4), 363-395.	*			*x*			*x*	
Cai, J. F., Ding, M. X., & Wang, T. (2014). How do exemplary Chinese and US mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265-280.		*x*						
Choppin, J. (2011). The Impact of Professional Noticing on Teachers' Adaptations of Challenging Tasks. *Mathematical Thinking and Learning, 13*(3), 175-197.	*			*x*				
Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban school districts. *American Journal of Education, 114*(4).				*x*	**X**			
Gavin, M. K., Casa, T. M., Adelson, J. L., & Firmender, J. M. (2013). The Impact of Challenging Geometry and Measurement Units on the Achievement of Grade 2 Students. *Journal for Research in Mathematics Education, 44*(3), 478-509.							*x*	**X**
Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies, 41*(4), 467-500.		**X**		*				
Hill, H. C., & Charalambous, C. Y. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Lessons learned and open issues. *Journal of Curriculum Studies, 44*(4), 559-576.	*			**X**				
Davis, E. A., Beyer, C., Forbes, C. T., & Stevens, S. (2011). Understanding pedagogical design capacity through teachers' narratives. *Teaching and Teacher Education, 27*(4), 797-810.	**X**			**X**				
Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using Educative Curriculum Materials to Support the Development of Prospective Teachers' Knowledge. *Educational Researcher, 43*(3).	**X**							

Forts. Tabell 9

	Studies that:							
	7. Describe design principles of educative curricula - these are curricula that are also educative for teachers - and how teachers respond to or use these features.	8. Describe how teachers prepare for teaching.	9. Relate influencing factors that explain transformations between and within the phases of curriculum use. Particularly studies that:	*5.1. Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.*	*3.4. Combine the inclusion of schools and districts as settings for curriculum implementation.*	6. Describe the effectiveness of curriculum programs.	*6.1. Describe the effectiveness of curriculum programs and provide data on the degree of implementation (observations).*	*6.2. Describe what and how students learn from interacting with particular curriculum programs, rather than how much they learn.*
Stein, M. K., & Kaufman, J. H. (2010). Selecting and Supporting the Use of Mathematics Curricula at Scale. *American Educational Research Journal, 47*(3), 663-693.	*x*	*x*		*	**X**			
Davis, E. A., Palincsar, A. S., Arias, A. M., Bismack, A. S., Marulis, L. M., & Iwashyna, S. K. (2014). Designing Educative Curriculum Materials: A Theoretically and Empirically Driven Process. *Harvard Educational Review, 84*(1), 24-52.	**X**							
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, 47*(4), 833-878.					*		*x*	*x*
Coburn, C. E., Mata, W. S., & Choi, L. (2013). The Embeddedness of Teachers' Social Networks: Evidence from a Study of Mathematics Reform. *Sociology of Education, 86*(4), 311-342.					**X**			
Voogt, J. (2010). Teacher factors associated with innovative curriculum goals and pedagogical practices: differences between extensive and non-extensive ICT-using science teachers. *Journal of Computer Assisted Learning, 26*(6), 453-464.				*x*				

Note. '**X**' indicates that the article was used as the main input to exemplify research in the related area; '*x'* indicates that the article was used as secondary input to exemplify research in the related area. An asterisk (*) indicates that the article contains related information but was not used to exemplify research in the related area.

Design principles of educative curricula and teachers' responses

To exemplify the research field's current state on this area, we draw mainly on our reading of three articles: Davis et al. (2011), Davis et al. (2014) and Drake et al. (2014). In addition, we also draw on our reading of Stein and Kaufman (2010).

Focusing on two teachers' use of a curriculum program, Davis et al. (2011) describe what kinds of changes these teachers make to the curriculum program, and the resources on which they draw when adapting the curriculum. Their study suggests that adaptations to the curriculum are more productive if they are based on the teacher's knowledge of the students, and if the teacher's goals for the lesson are aligned with those of the

curriculum authors. The study also reveals that even a subtle lack of alignment of goals can influence the quality of the adaptation. An implication, relating not only to teacher education and professional development but also to the design of educative curriculum programs, holds that teachers need support in understanding the intent of the lesson and of specific lesson activities or tasks. Related to this is the observation by Stein and Kaufman (2010) that the quality of instruction was higher when teachers used a curriculum program that supported them in identifying the big mathematical ideas. Likewise, teachers should be supported in developing a good understanding of their students' knowledge and capacities (Davis et al., 2011). We found that Davis et al. (2014) addressed these implications by focusing on how this support can be provided by curriculum programs, and that Drake et al. (2014) addressed these implications by focusing on the use of educative curriculum programs in teacher education.

Davis et al.'s (2014) article adds to the research field by describing a design-based research approach to designing educative curriculum programs. Arguing that the development of educative curriculum programs lacks a grounding in empirical observations of teachers' actual use of the curriculum programs, they ground their design process in both theoretical and empirical foundations. To do this, they first analyzed students' learning opportunities in existing curriculum programs. Second, they observed lessons to analyze how teachers implemented the lesson and how this influenced the students' learning opportunities in class. Third, they also studied students' outcomes. Informed by the information retrieved in these three steps and drawing on literature, the authors made specific design decisions. In total, five types of educative features were added to the existing curriculum program. After a testing of the "enhanced" curriculum program, the authors refined the educative features. Central in all these features is the provision of a rationale for recommendations, and guidance in how these recommendations can be implemented and adapted by teachers to fit the students in their class. As such, this study directly addresses Davis et al.'s (2011) stance that curriculum programs should support teachers in understanding the intent of a lesson. This study also supports the teacher in customizing the lesson to the students in class; this is also considered important for productive adaptations to the curriculum program (Davis et al., 2011).

Drake et al. (2014) build on related articles that have launched the idea that curriculum programs could be more supportive for the teacher if developed with a closer connection to the practice of teaching (Ball & Cohen, 1996), and that have proposed design principles for curriculum programs in order to do this (Davis & Krajcik, 2005). This addresses Davis et al.'s (2011) stance that teacher education should be supportive of prospective teachers' understanding of the goals of the curriculum and the development of strategies to know their students. Drake et al. (2014) describe that most studies on teacher learning from educative curriculum programs have focused on learning *about* these materials. Drawing on literature on the knowledge teachers need for teaching mathematics, Drake et al. (2014) add to this and argue that educative curriculum programs should also be used in teacher education programs to learn *from* these materials. To this end, and based on theory, prior research and their own empirical work, the authors describe five design principles to help prospective teachers use curriculum programs in educative ways.

A description of how teachers prepare for teaching

We draw on three articles to exemplify research on teachers' preparations for teaching. Sherin and Drake (2009) propose a curriculum strategy framework, developed and used to describe ten teachers' patterns of curriculum use. The framework focuses on three interpretative activities – reading, evaluating, and adapting the curriculum program – and does this before, during and after instruction. Sherin and Drake find patterns in teachers' curriculum strategy, or in when and how they read, evaluate and adjust the curriculum. Relevant for our focus, they distinguish three groups of teachers concerning the reading of the curriculum program. Whereas one group reads prior to instruction to get a sense of the broad overview of the lesson, the second group reads prior to instruction to know more about the details of the lesson, and the third group reads both before and during instruction, having done so before instruction to get a broad overview of the lesson, then adding this to more detailed information read during the lesson. Also, there was evidence that teachers do not read all information, and there were differences both in the nature of information the teachers read and in when they read this information.

The study also reveals that teachers who evaluated what they were expected to do with the curriculum program prior to instruction, and then shifted the focus of their evaluation to the students during instruction, usually adapted the curriculum by creating new activities rather than omitting or replacing them. It is also found that adaptation prior to instruction related to organizational aspects, while adaptation only during instruction also concerned content-related aspects. This is an interesting finding, given that Stein and Kaufman (2010) find that the quality of instruction was higher if teachers attended to the big mathematical ideas of the lesson while preparing for instruction.

In a study of how Chinese and US teachers view instructional coherence, Cai et al. (2014) find that both groups of teachers stressed the importance, while planning for the lesson, of attending to clear objectives and student thinking. Whereas not at all mentioned by the US teachers, most of the Chinese teachers also stressed the importance of studying the textbooks to understand the curriculum authors' intentions, and how they connected to prior and later pieces of knowledge.

Influencing factors that explain transformations between and within the phases of curriculum use

Research in this section is exemplified according to two particular foci. We first describe research that explicitly addresses or points to a need to study the interaction between curriculum resources and teacher resources. Second, we describe findings of research including the study of curriculum programs in broader settings such as schools and districts.

An integrated analysis of curriculum resources and teacher resources, and how they interact

Two studies explicitly center on an analysis of curriculum resources and teacher resources, and the interaction between the two. Hill and Charalambous (2012) report on a series of case studies set up to better understand both the unique and joint contribution of teachers' mathematical knowledge for teaching (MKT) and curriculum programs to the mathematical quality of instruction. With regard to the unique contribution of MKT, the four case studies provide information on how strong MKT might be beneficial for certain aspects of the mathematical quality of instruction. For instance, the most consistently observed affordance of MKT is related to teachers' use of a dense and precise mathematical language. The analysis also suggests that the association between MKT and student participation in meaning-making and reasoning might be weak. Due to the studies' design, more speculative findings are reported regarding the unique contribution of curriculum programs to the mathematical quality of instruction. The case studies suggest that the supports provided by the curriculum program matter, but also that the *Standards*-based curriculum program also placed a specific demand on the teacher that at times resulted in poor lesson implementation. Considering the joint contribution of MKT and curriculum programs, the study suggests that supportive curriculum programs might lead to high-quality instruction, even for low-MKT teachers, if the curriculum program is followed closely. It is also suggested that unsupportive curriculum programs might be problematic, particularly for low-MKT teachers. A third suggestion describes that high-MKT teachers might be able to compensate for some limitations in the curriculum program, enabling a high-quality instruction. The study further stresses the importance of teachers' orientation to curriculum programs and instruction, in addition to MKT and curriculum materials as important "contributors" to the mathematical quality of the lesson. Importantly, the study also reveals that the relationship between these contributors was more complex than straightforward for middle-rank-MKT teachers – which is the largest group of teachers.

Acknowledging that teachers need to adapt curriculum programs to the needs of particular students, Davis et al. (2011) studied how two teachers made adaptations to the curriculum. They studied the kinds of adaptations made by the teachers and the knowledge, experiences and resources they teachers draw when making adaptations. One of the teachers based her curricular adaptations on her knowledge of her students, the fact that the learning goals she aimed for aligned with the philosophy of the curriculum, her rich set of teaching experiences, and several curricular resources. The second teacher drew mainly on her own learning goals, which were usually not aligned with those of the curriculum, and additionally on her knowledge of her students. The first teacher's adaptations were more in line with the curriculum's intentions than those made by

the second teacher. Thus, by suggesting that teachers' changes to the curriculum are likely to be more productive if the teachers explicitly consider their knowledge of the students and the alignment between their own goals and those of the curriculum, the study sheds light on the complex interplay between teachers and curriculum programs in order to come to productive adaptations.

Choppin (2011) also studied teachers' adaptations to challenging curriculum programs. Of the five teachers included in the study, three made changes in a way that lowered the complexity. These teachers often evaluated the adapted tasks based on previous students' performance of them, which had shown that they needed additional support. Two teachers adapted the curriculum in a way that the tasks maintained their level of complexity. These teachers drew on specific features of student solutions, often interpreting them in terms of students' meaning-making of mathematical ideas. The author argues that the challenging tasks in the curriculum program were supportive for the second group of teachers, fostering their attention to student thinking. At the same time, the study shows that the first group of teachers likely need more support, indicating that adopting challenging curriculum programs does not necessarily result in increased teacher learning. By revealing how the two groups of teachers were (un)successful in maintaining the complexity of tasks when adapting the curriculum program to the students in class, this study sheds light on the interactive play between teachers and curriculum programs.

Brown et al.'s (2009) study of fidelity of implementation reveals that teachers' fidelity to the literal lesson did not necessarily match the intentions of the curriculum program authors. They find some patterns among teachers, as well as variations specific to particular lessons in which curriculum authors could have been more straightforward regarding their intentions. The authors contend that a study of curriculum implementation should thus include an analysis of the curriculum program in addition to the actual enactment by the teacher, in order not to miss how characteristics of the curriculum programs might be influential in this regard.

Stein and Coburn (2008) studied environments set up by the district to support teachers in implementing new curriculum programs. They did this in two districts, describing the learning environment in one as rich and focused on mathematical concepts and student thinking, whereas in the second it was characterized as focusing on management-related issues. Stein and Coburn (2008) connect this to the use of a particular curriculum program in each district. The first district adopted a curriculum program that provided opportunities to negotiate meanings about teaching and learning, whereas the program adopted by the second district was characterized as being limited in the provision of such opportunities. As such, the authors shed light on how features of the curriculum programs might play out on initiatives set up on a district level.

Voogt (2010) reports on teacher factors associated with extensive and non-extensive ICT use in Grade 8 in 22 educational systems. Among other findings, it was found that extensive ICT users considered themselves more confident in ICT competencies than did non-extensive users. This was particularly related to competencies related to the pedagogical use of ICT. It was also found that extensive ICT users more often collaborated with other teachers and attended more related professional development. These findings indicate the usefulness of professional development that accompanies ICT implementation to help teachers using ICT in an appropriate pedagogical way. Although the study was carried out in the context of science education, it seems likely that this could also be useful to attend to when digitalizing (parts of) the curriculum program.

Combine inclusion of schools and districts as settings for curriculum implementation

In this section, we draw on three studies. Stein and Coburn (2008), Stein and Kauffman (2010) and Coburn, Mata, and Choi (2013) are related in that they address the district-wide adoption of two curriculum programs in two districts. The three studies investigate the same districts and the same curriculum programs. Each of the two districts adopted one of the following two curriculum programs: *Everyday mathematics* (EM) and *Investigations* (INV). INV is characterized as more demanding for the teacher to implement than EM, but at the same time is also more supportive of the teacher.

Stein and Kaufman (2010) focus on the selection of mathematics curriculum programs at scale, i.e., at a district level, and argue that both teacher capacity and teachers' use of the curriculum programs influence the quality of instruction. Teacher capacity is determined by teachers' mathematical knowledge for teaching, experience, education and professional development. Teachers' use of the curriculum programs is

operationalized in their perceptions about the usefulness of the curriculum program, the percentage of time they used the curriculum programs, and what they reviewed and talked about when preparing the lesson (divided into non-mathematical, big mathematical ideas, materials). Quality of instruction centers on maintaining a high cognitive demand, attending to student thinking, and vesting intellectual authority in students' mathematical reasoning. Stein and Kaufman find that the quality of instruction was higher in districts that adopted INV. Related to teacher capacity at the district level, only the total amount of hours professional development influenced the quality of instruction in a significant, beneficial way; but only in the district that adopted INV. A more important determinant, playing out in both districts, seems to be how teachers used the curriculum programs. More particularly, the results suggest that reviewing the big mathematical ideas of a lesson while preparing is crucial to instructional quality. A further analysis of both curriculum programs revealed that INV is more supportive in this regard than EM. These findings imply that administrators concerned with the adoption of curriculum programs should consider the affordances provided by the programs to help teachers identify the big mathematical ideas of the lessons. The findings also suggest that the adoption of a curriculum program should be supplemented with a professional development program to help teachers identify the big mathematical ideas in the curriculum.

Stein and Coburn (2008), centering on understanding how districts can create organizational environments for teacher learning required for the implementation of new curriculum programs, focus on the same districts and curriculum programs as Stein and Kaufman (2010). To foster the implementation of the new curriculum program (EM), one district created a new organization outside the existing administrative line. This organization included one full-time mathematics coach per school, regional instructional specialists, and a central mathematics leadership team. Teachers mainly contacted the mathematics coach, and this connection was unidirectional. Professional development focused on how to manage and navigate through the curriculum material.

In the second district, which adopted the INV curriculum program, mathematics coaches were also the primary link between the district leaders and the teachers. Instead of full-time coaches, this district had two part-time coaches at every school who also taught part-time. Also, instead of regional instructional specialists, the school principals got involved in working with coaches and district leaders. In general, there was a richer flora of practices in the second district (INV) than in the first (EM), and the connections between the various actors were bidirectional in the second district whereas they were unidirectional in the first one. Professional development in the second district involved doing mathematics together and discussing the nature of the involved mathematics. The curriculum programs were viewed as a tool for meeting grade-level objectives. Also in the second district, the interactions between teachers were more likely to stretch over grade levels and go beyond management of the curriculum programs. Based on these findings, Stein and Coburn (2008) emphasize two important structural differences between the districts: the direction of the architecture, i.e. unidirectional (from district to schools and teachers) or bidirectional (input from both district, schools and teachers); and inclusive or exclusive of the existing school structure. The authors also point out that the INV curriculum program, more than EM, provides openings for the different actors to negotiate the meaning of reform for their practices.

As teachers' social networks can play an important role in teacher learning and organizational change, Coburn, Mata and Choi (2013) studied the influence of district policy on the formation of social networks. They did this in the context of the implementation of a new curriculum program (INV) in one of the two abovementioned districts. In Year 1, the district initiated multiple structures that required teachers to interact with new people regarding mathematics. The authors found that teachers' networks were small, and that they usually reached out to grade-level colleagues or coaches at the school. In Year 2 the district offered additional professional development for teachers in cross-school settings, and on a school level it shifted to cross-grade-level configurations. Consequently, the networks became larger and included more cross-level and outside-the-school connections. The teachers also became more successful in reaching out to find specific expertise. In Year 3 the professional development related to mathematics education diminished, and as a result the teachers again mostly reached out to same-level teachers. However, even in Year 3 teachers remained skillful in locating expertise in their network. As such, the study shows that teachers' social networks are sensitive to policy influence. It also reveals how the networks diminish when support is removed. Further, the study also reveals

that the exchange of district materials, including curriculum programs, in contrast to district information remained even when related professional development diminished. This indicates that both materials and expertise are, more than information, durable resources.

Effectiveness of implemented curriculum programs

We base our exemplification of this area on four research articles. Two deal explicitly with replacing existing units in curriculum programs with more ambitious ones, more in line with the philosophy of the NCTM *Standards*. Gavin et al. (2013) do this for geometry and measurement in the second grade of elementary school, while Roschelle et al. (2010) do it for proportionality and linear functions in Grades 7 and 8 of middle school. In both studies, professional development sessions accompanied the implementation of the new curriculum units, and both studies checked for implementation level as well. Both studies find that students working with the new materials did not fare worse on a more traditional test, and fared better on a test aligned with the ambitious philosophy. Whereas Gavin et al. (2013) do not concretely analyze which of the features of the curriculum unit influenced the students' scores, they conjecture two such features to be salient: complex tasks requiring reasoning at high levels; and instructional strategies encouraging students to explain their reasoning both verbally and in writing. Roschelle et al. (2010) replicated their findings across a variety of settings, which stresses that the combination of professional development and the integration of technology focusing on representations can result in robust improvements.

Particularly interesting in Tarr et al. (2008) is the analysis of the impact of both the curriculum program and the classroom learning environment on student performance. Student performance is measured through two tests: a more traditionally oriented one, and another more in line with the philosophy of the NCTM *Standards*. Two types of curriculum programs are included in the study: those developed by means of NSF funding to be aligned with the NCTM *Standards*, and commercially developed ones. The study measures both the degree of implementation of the curriculum programs and the extent to which the learning environment, or the actual teaching, was in line with the NCTM *Standards*. It is found that both groups of teachers utilized their curriculum program in a similar way, although those using NSF-funded curriculum programs reported higher frequencies of textbook use by their students. When taking into account student background and teacher variables, the authors find that both curriculum type and learning environment did not predict student achievement. Important, however, is the finding that the combination of NSF-funded curriculum programs and learning environments that were moderately to highly in line with NCTM *Standards* had a significant positive impact on the test that was in line with the reform philosophy. The same is not found for the scores on the more traditional test. Also, it is found that the combination of the use of an NSF-funded curriculum program and a learning environment with low alignment with the NCTM *Standards* even resulted in lower achievement. The findings that the use of ambitious curriculum programs does not automatically result in ambitious learning environments add to earlier findings about these ambitious curriculum programs being challenging to enact well. This also suggests that professional development aimed at yielding a strong implementation of such curriculum programs is needed; otherwise, there might even be negative effects on student learning.

Brown et al. (2009) studied fidelity of implementation in the context of whole-number lessons in Grades 1 and 2. Brown et al. (2009) distinguish between fidelity to the lesson activities as they appear in the curriculum program ("fidelity to the literal lesson") and fidelity to the learning opportunities intended by the curriculum program authors ("fidelity to the authors' intended lesson"). Interestingly, the results reveal that the level of fidelity to the literal lesson does not predict fidelity to the authors' intended lesson. This suggests that teachers likely need support in the curriculum program to identify learning opportunities, although Brown et al. (2010) mention that the analyzed curriculum program regularly provides such support. This in turn sheds light on the need to further study how curriculum programs can support teacher learning and how teachers make use of this support.

Summarizing research

To exemplify the current state of the research field regarding educative curriculum programs, we draw on our reading of four research articles: Davis et al. (2011), Davis et al. (2014), Drake et al. (2014), and Stein and Kaufman (2010). Central in all these articles is the assumption, stemming from framing the teacher-curriculum relationship as interactional, that curriculum programs can support the teacher in implementing curriculum programs in class. Following how two teachers adapt the curriculum, Davis et al. (2011) point out that one way curriculum programs could be educative for the teacher is by explicitly describing the intent of the lesson, lesson activities or tasks, and providing support in understanding their students. Stein and Kaufman (2010) describe that the quality of instruction is higher if the teachers attend to key mathematical ideas while preparing for the lesson. Furthermore, they connect this to the related support provided by the curriculum programs, and mention that administrators concerned with the selection of curriculum programs should consider the affordances provided by the programs to help teachers identify the big mathematical ideas of the lessons. Concerning the design of educative curriculum programs, Davis et al. (2014) propose a conceptualization of a principled design of educative curriculum programs, closely tied to the enactment of teaching. Drake et al. (2014) suggest that educative curriculum programs can be used in teacher education in a way that prospective teachers learn *from* in addition to *about* these materials, and propose design principles for doing this. As such, this study is interesting as it describes how educative curriculum programs can be useful in teacher education for building the knowledge prospective teachers need in order to start teaching.

We draw on three research articles to exemplify research on teachers' preparations for teaching. These studies reveal how teachers prepare for teaching and how this might influence their adaptation of the curriculum (Sherin & Drake, 2009). It is also found that attending to big mathematical ideas while preparing for the lesson is beneficial for the quality of instruction (Stein & Kaufman, 2010), and one study finds both similarities and differences between Chinese and US teachers' answers when asked about planning for teaching a mathematics lesson in a coherent way (Cai et al., 2014).

Our reading of research articles about the interaction between teacher resources and curriculum resources sheds light on a complex interplay between teachers and curriculum programs. Hill and Charalambous (2012) explicitly describe how teacher resources and curriculum resources uniquely and jointly contribute to the quality of instruction. Davis et al. (2011) describe the importance of aligning teachers' goals with the curriculum's goals, and of attending to student thinking in order to make productive changes to the curriculum. Also, Choppin (2011) describes attending to students' meaning-making of mathematical ideas as important in adjusting curriculum programs without lowering the complexity of tasks. Choppin (2011) also connects this to the presence of challenging tasks in the curriculum program, mentioning that some teachers likely need additional support in the program. Stein and Coburn (2008) connect features of the curriculum program to the character of initiatives set up by the district. Expanding the findings of Voogt (2010) to the context of digitalizing mathematics curriculum programs suggests that teachers would likely benefit from support in enacting such digitalized programs in a pedagogically appropriate way. Finally, Brown et al. (2009) reveal the necessity of taking into account features of the curriculum programs when studying the fidelity of implementation by the teacher.

To exemplify research that includes broader settings such as schools and districts in the study of curriculum implementation, we focused on three related studies. These studies shed light on a) important criteria to take into account when selecting a curriculum program (Stein and Kaufman, 2010), b) how districts can create organizational environments to foster the teacher learning required for the implementation of new curriculum programs (Stein & Coburn, 2008), and c) how district policy might influence the formation of social networks inside and outside schools supportive of teachers' implementation of the curriculum programs (Coburn et al., 2013).

We have drawn on four articles in our exemplification of research on the effectiveness of curriculum programs. Two focus on the effects of replacing existing units in curriculum programs with more ambitious ones (Gavin et al., 2013; Roschelle et al., 2010). Both find that students working with the new materials did not fare worse on a more traditional test, and fared better on a test aligned with the ambitious philosophy. Tarr et al. (2008) reveal that the adoption of ambitious curriculum programs does not necessarily result in better

performance by the students. Their study reveals that the learning environment in the class needed to be in line with the philosophy of these curriculum programs in order to positively impact student performance. The study also reveals that if the learning environment in the classroom did not align well with the curriculum programs, this could result in lower student achievement. One study focuses explicitly on fidelity of implementation, without taking into account the curriculum program's effectiveness (Brown et al., 2009). The study reveals that in order to measure curriculum implementation, one should also take into account the support provided by the curriculum program. In sum, stemming from these studies is that teachers likely need support accompanying the adoption of such ambitious curriculum programs. Also, although these studies describe that the adoption of these ambitious curriculum programs can result in achieving more ambitious learning goals without necessarily harming more traditional ones, it still remains rather vague how students actually learn from interacting with these curricula.

4.4 CURRICULUM PROGRAMS: DISCUSSION, LIMITATIONS AND CONCLUSION

Discussion

In the introductory section, we described that curriculum programs are widely used in mathematics classrooms (Grouws et al., 2004; Mullis et al., 2008; Pepin & Haggarty, 2001) and, as such, are powerful tools for impacting the teaching and learning of mathematics. We also mentioned that human interaction with curriculum programs influences the way the lessons are taught, and thus influences students' learning opportunities (Freeman & Porter, 1989; Remillard & Bryans, 2004; Stein et al., 1996). These findings underline that the impact of a curriculum program on teaching and student learning is not straightforward. Stein et al. (2007) have conceptualized this through the unfolding of the curriculum in a series of temporal phases. They distinguish between a written curriculum (for instance, the syllabus, teacher guide or student textbooks), an intended curriculum (or the teachers' plans for instruction) and an enacted curriculum (the actual teaching in the classroom). They also describe variables that mediate within and between these phases, such as teachers' orientations toward the curriculum program or organizational and policy contexts. All these variables – written curriculum, intended curriculum, enacted curriculum and mediating variables – interact and influence mathematics classroom practice and student learning.

Our reading of the abstracts of selected research articles has revealed that all of these variables are covered in recent research, although to different extents. More than a third of the 62 research articles address aspects that contribute to the unfolding of the curriculum, about a fourth focus on the effectiveness of curriculum programs, and about a fifth analyze textbooks. The inclusion of broader settings such as schools and districts, studies of how curriculum programs can be educative for the teacher, and studies of how teachers prepare for teaching are mentioned to a much lower extent: in 8%, 8% and 5% of the abstracts, respectively. It is remarkable that although mathematics curriculum programs are used in a variety of educational contexts, most (66%) of the research articles are set up exclusively within a US context.

A number of studies shed light on the complex interplay between teacher resources and curriculum resources and how this impacts classroom instruction. Particularly illustrative is Hill and Charalambous' (2012) report on a series of case studies seeking a better understanding of the unique and joint contribution of teachers' mathematical knowledge for teaching (MKT) and curriculum programs to the quality of instruction. The case studies suggest that 1) supportive curriculum programs can lead to high-quality instruction, even for low-MKT teachers, if the program is followed closely; 2) unsupportive curriculum programs might be problematic, particularly for low-MKT teachers; 3) high-MKT teachers might be able to compensate for some limitations in the curriculum program, enabling high-quality instruction; and 4) teachers' orientation toward curriculum programs and instruction, in addition to MKT and curriculum programs, are important contributors to the mathematical quality of the lesson. The studies also reveal that the relationship between these contributors was more complex than straightforward for middle-rank-MKT teachers, which is by definition the largest group of teachers. Other studies connect features of the newly adopted curriculum program to the character of initiatives taken by the district to support the implementation of the curriculum programs. For instance, Stein and Coburn (2008) relate the richness of an environment set up by a particular district to opportunities provided in the curriculum program to negotiate meanings about teaching and learning.

Addressing how curriculum programs might support the teacher in the complex interaction between teachers and curriculum programs, a number of studies focus on how curriculum programs can be educative for teachers; that is, how curriculum programs can support teacher learning to implement them in a way that fits with the students in class. Davis et al. (2011) focus on two teachers' adaptations to the curriculum. Their findings suggest that such adaptations are more productive if the teacher's goals are aligned with those of the curriculum, and if the teacher explicitly attends to the students' thinking. An implication for the design of educative curriculum programs entails the need to explicitly describe the intentions of the proposed lesson, and to provide support in attending to student thinking. Addressing this implication, Davis et al. (2014) present a

design-based research approach for the development of educative curriculum programs, which is based on both theoretical as empirical grounds. The proposed educative features center on the provision of a rationale for recommendations, and of guidance to implement and adapt these recommendations to the needs of the students in class. Drake et al. (2014) propose five design features for using educative curriculum programs in teacher education so that prospective teachers gain the knowledge needed to teach mathematics. As such, they relate the use of educative curriculum programs to teacher education.

Given the complex relationship between teachers, curriculum programs and the quality of instruction, and given that some research explicitly focuses on how curriculum programs can be educative for teachers, we found it surprising to retrieve only a small number of articles describing how teachers use these curriculum programs in planning for teaching. Underlining the importance of related research is a study by Stein and Kaufman (2010), stressing its importance for the instructional quality of teachers' focus on big mathematical ideas when preparing for teaching. Sherin and Drake (2009) found patterns in when and how teachers read, evaluate and adapt the curriculum program. They also found that teachers do not read all information, and that there were differences in the nature of the information the teachers read as well as when they read it. More related research is needed.

Some articles describe the effectiveness of ambitious curriculum programs. These articles tend to lay out that students working with reform-oriented curriculum programs generally do not fare worse on a traditional test than students working with traditional curriculum programs, and outperform their counterparts on a test aligned with the reform philosophy (Gavin et al., 2013; Roschelle et al., 2010). However, it remains rather vague how students learn from interacting with these curricula. Particularly interesting is a research article by Tarr et al. (2008), which reveals that the adoption of ambitious curriculum programs does not necessarily result in better performance by students. The study stresses the need for the learning environment in the class to be aligned with the philosophy of the curriculum program in order to positively impact student performance. This is particularly important as the study also reveals that if the learning environment in the classroom is not aligned with the ambitious curriculum program, this could result in even lower student achievement.

Pros and cons of the method

The screening based on the abstract of research articles indexed in Web of Science proved to be helpful in carrying out the review within the given time constraints. However, there are also a number of limitations tied to this method. First, our decision to focus solely on research articles indexed in Web of Science inevitably means that we excluded other potentially interesting research articles or book chapters not listed there. Also, intrinsic to the information provided in the abstract of research articles is its limitedness. For instance, some abstracts did not provide information on the context of study. For other articles that we read completely, we noted that some also included information exemplifying a research area other than the one for which the article had been selected, based on the information in the abstract. The selection of articles to read completely based on the selected area for future research proved to be helpful in specifically orienting this review.

Conclusion

A central finding stemming from this review is the complexity involved in how curriculum programs can support teachers in establishing classroom practices. Both curriculum resources and teacher resources, as well as other influencing factors, impact the quality of instruction, and studies have begun to point out how curriculum resources and teacher resources individually and jointly impact classroom practices. Multiple research articles have expressed the need for teacher support in implementing curriculum programs, by means of professional development, teacher education and support provided by the curriculum programs. Interesting in this regard is the state of the research field concerning the design of educative curriculum programs, and how teachers make use of such support. Studies have proposed design approaches, both for designing educative curriculum programs as well as for how to use educative curriculum programs in teacher education to support prospective teachers' development of knowledge. Further, although research has revealed that it is important to prepare in certain ways for teaching, we found very little research explicitly analyzing how teachers actually prepare for teaching a mathematics lesson. Also, we noted that it remains rather vague precisely how students

learn from interacting with ambitious curriculum programs, and research has stressed the need for a learning environment to be in line with the philosophy of the curriculum program to positively impact student learning. This stresses that simply introducing ambitious curriculum programs will not guarantee high quality instruction. Finally, the great majority of the reviewed studies were set up uniquely in a US context; given that curriculum programs stand as cultural artifacts, this suggests the need for more related research in other settings.

REFERENCES

Alajmi, A. H. (2012). How do elementary textbooks address fractions? A review of mathematics textbooks in the USA, Japan, and Kuwait. *Educational Studies in Mathematics, 79*(2), 239-261. doi: 10.1007/s10649-011-9342-1

Andersson, C. (2014). *The Impact of a Comprehensive Teacher Professional Development Program in Formative Assessment on Teachers' Practice and Students' Achievement*. Presentation at Nionde matematikdidaktiska forskningsseminariet (Madif-9), Sweden. Hämtad från http://ncm.gu.se/media/smdf/web_versions/shorts/S1_andersson_abstract%20madif_20140204.pdf

Andersson, C., Vingsle, L. & Palm, T. (2013). *The impact of a teacher professional development program in formative assessment on teachers' practice.* Poster presented at Eighth Congress of European Research in Mathematics Education (CERME 8), Turkey. Hämtad från http://umu.diva-portal.org/smash/get/diva2:746038/FULLTEXT01.pdf

Anthony, G. & Walshaw, M. (2007). *Effective pedagogy in mathematics: Best evidence synthesis iteration*. Wellington: Ministry of Education.

Brown, M. W. (2009). The teacher-tool relationship. Theorizing the design and use of curriculum materials. In J. T. Remillard, B. A. Herbel-Eisenman & G. M. Lloyd (Eds.), *Mathematics teachers at work: connecting curriculum materials and classroom instruction* (pp. 17-36). New York: Routledge.

Brown, S. A., Pitvorec, K., Ditto, C., & Kelso, C. R. (2009). Reconceiving Fidelity of implementation: An Investigation of Elementary Whole-Number Lessons. *Journal for Research in Mathematics Education, 40*(4), 363-395.

Baker, D., Knipe, H., Collins, J., Leon, J., Cummings, E., Blair, C., & Gamson, D. (2010). One Hundred Years of Elementary School Mathematics in the United States: A Content Analysis and Cognitive Assessment of Textbooks From 1900 to 2000. *Journal for Research in Mathematics Education, 41*(4), 383-423.

Bakker, A., & van Eerde, H. A. A. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.). *Doing qualitative research: Methodology and methods in mathematics education* (pp. 429-466). Berlin: Springer. doi: 10.1007/978-94-017-9181-6_16

Balan, A. (2012). *Assessment for learning: a case study in mathematics education.* (Doctoral dissertation, Malmö högskola, Fakulteten för lärande och samhälle).

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: what is – or might be – the role of curriculum materials in teacher learning and instructional reform. *Educational researcher, 25*(9), 6-8, 14.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389-407.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Bieda, K. N. (2010). Enacting Proof-Related Tasks in Middle School Mathematics: Challenges and Opportunities. *Journal for Research in Mathematics Education, 41*(4), 351-382.

Björklund Boistrup, L. (2013). *Bedömning i matematik pågår! : återkoppling för elevers engagemang och lärande*. Stockholm: Liber.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in education*, *5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the Black Box: Raising Standards through Classroom Assessment. *The Phi Delta Kappan, 80(2)*, 139- 148.

Black, P., & Wiliam, D. (2009). Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Boesen, J., Helenius, O., Bergqvist, E., Bergqvist, T., Lithner, J., Palm, T., & Palmberg, B. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *The Journal of Mathematical Behavior, 33*, 72-87.

Boström, E. (2014). *The Impact of a Teacher Professional Development Program in Formative Assessment on Mathematics Teachers' Classroom Practice*. Presentation at Nionde matematikdidaktiska forskningsseminariet (Madif-9), Sweden. Hämtad från http://ncm.gu.se/media/smdf/web_versions/shorts/S2_erika_bostrom-short_communication_madif9.pdf

Bouck, E. C., Joshi, G. S., & Johnson, L. (2013). Examining calculator use among students with and without disabilities educated with different mathematical curricula. [Article]. *Educational Studies in Mathematics, 83*(3), 369-385. doi: 10.1007/s10649-012-9461-3

Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-Analytic Methodology and Inferences About the Efficacy of Formative Assessment. *Educational Measurement: Issues and Practice, 31(4)*, 13-17.

Brodie, K. 2011. Working with learners' mathematical thinking: Towards a language of description for changing pedagogy. Teaching and Teacher Education, 27(1), 174-186

Brosvic, G. M., Dihoff, R. E., Epstein, M. L., & Cook, M. L. (2006). Feedback facilitates the acquisition and retention of numerical fact series by elementary school students with mathematics learning disabilities. *Psychological Record, 56*(1), 35-54.

Buric, I., & Soric, I. (2012). The role of test hope and hopelessness in self-regulated learning: Relations between volitional strategies, cognitive appraisals and academic achievement. *Learning and Individual Differences, 22*(4), 523-529. doi: 10.1016/j.lindif.2012.03.011

Burns, M. K., Klingbeil, D. A., & Ysseldyke, J. (2010). The effects of technology-enhanced formative evaluation on student performance on state accountability math tests. *Psychology in the Schools, 47*(6), 582-591. doi: 10.1002/pits.20492

Cai, J. F., Ding, M. X., & Wang, T. (2014). How do exemplary Chinese and US mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265-280. doi: 10.1007/s10649-013-9513-3

Cai, J. F., Moyer, J. C., Wang, N., Hwang, S., Nie, B. K., & Garber, T. (2013). Mathematical problem posing as a measure of curricular effect on students' learning. *Educational Studies in Mathematics, 83*(1), 57-69. doi: 10.1007/s10649-012-9429-3

Cates, G. L. (2005). Effects of peer versus computer-assisted drill on mathematics response rates. *Psychology in the Schools, 42*(6), 637-646. doi: 10.1002/pits.20105

Charalambous, C. Y., Delaney, S., Hsu, H. Y., & Mesa, V. (2010). A Comparative Analysis of the Addition and Subtraction of Fractions in Textbooks from Three Countries. *Mathematical Thinking and Learning, 12*(2), 117-151. doi: 10.1080/10986060903460070

Charalambous, C. Y., & Hill, H. C. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship. *Journal of Curriculum Studies, 44*(4), 443-466. doi: 10.1080/00220272.2011.650215

Charalambous, C. Y., Hill, H. C., & Mitchell, R. N. (2012). Two negatives don't always make a positive: Exploring how limitations in teacher knowledge and the curriculum contribute to instructional quality. *Journal of Curriculum Studies, 44*(4), 489-513. doi: 10.1080/00220272.2012.716974

Charalambous, C. Y., & Philippou, G. N. (2010). Teachers' concerns and efficacy beliefs about implementing a mathematics curriculum reform: integrating two lines of inquiry. *Educational Studies in Mathematics, 75*(1), 1-21. doi: 10.1007/s10649-010-9238-5

Choppin, J. M. (2009). Curriculum-Context Knowledge: Teacher Learning From Successive Enactments of a Standards-Based Mathematics Curriculum. *Curriculum Inquiry, 39*(2), 287-320. doi: 10.1111/j.1467-873X.2009.00444.x

Choppin, J. M. (2011). The Impact of Professional Noticing on Teachers' Adaptations of Challenging Tasks. *Mathematical Thinking and Learning, 13*(3), 175-197. doi: 10.1080/10986065.2010.495049

Clandinin, D. J., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 363-401). New York: Macmillan.

Clarke, D. (2006). Using international research to contest prevalent oppositional dichotomies. *ZDM- The International Journal of Mathematics Education, 38*(5), 376-387.

Clarke, D. (2013). Contingent conceptions of accomplished practice: the cultural specificity of discourse in and about the mathematics classroom. *ZDM: The International Journal on Mathematics Education*, *45*, 21-33.

Clarke, B., Doabler, C. T., Cary, M. S., Kosty, D., Baker, S., Fien, H., & Smolkowski, K. (2014). Preliminary Evaluation of a Tier 2 Mathematics Intervention for First-Grade Students: Using a Theory of Change to Guide Formative Evaluation Activities. *School Psychology Review, 43*(2), 160-178.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C., & Loef , M. (1989). Using Knowledge of Children's Mathematics Thinking in Classroom Teaching: An Experimental Study. *American Educational Research Journal, 26*, 499-531.

Cleary, T. J., & Callan, G. L. (2014). Student Self-Regulated Learning in an Urban High School: Predictive Validity and Relations Between Teacher Ratings and Student Self-Reports. *Journal of Psychoeducational Assessment, 32*(4), 295-305. doi: 10.1177/0734282913507653

Cobb, P. (2007). Putting philosophy to work: Coping with multiple theoretical perspectives. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 3-30). Charlotte, NC: Information Age Publishing.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*, 9–13.

Cobb, P., & Jackson, K. (2008). The Consequences of Experimentalism in Formulating Recommendations for Policy and Practice in Mathematics Education. *Educational Researcher, 37*(9), 573-581. doi: 10.3102/0013189x08327826

Coburn, C. E., Mata, W. S., & Choi, L. (2013). The Embeddedness of Teachers' Social Networks: Evidence from a Study of Mathematics Reform. *Sociology of Education, 86*(4), 311-342. doi: 10.1177/0038040713501147

Confrey, J., Strutchens, M. E., Battista, M. T., Smith, M. S., King, K. D., Sutton, J. T., . . . Reed, J. (2008). Situating research on curricular change. *Journal for Research in Mathematics Education, 39*(2), 102-112.

Davis, E. A., Beyer, C., Forbes, C. T., & Stevens, S. (2011). Understanding pedagogical design capacity through teachers' narratives. *Teaching and Teacher Education, 27*(4), 797-810. doi: 10.1016/j.tate.2011.01.005

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational researcher, 34*(3), 3-14.

Davis, E. A., Palincsar, A. S., Arias, A. M., Bismack, A. S., Marulis, L. M., & Iwashyna, S. K. (2014). Designing Educative Curriculum Materials: A Theoretically and Empirically Driven Process. *Harvard Educational Review, 84*(1), 24-52.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*(3), 231-264. doi: 10.1007/s11409-008-9029-x

Ding, M. X., & Li, X. B. (2014). Transition from concrete to abstract representations: the distributive property in a Chinese textbook series. *Educational Studies in Mathematics, 87*(1), 103-121. doi: 10.1007/s10649-014-9558-y

Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using Educative Curriculum Materials to Support the Development of Prospective Teachers' Knowledge. *Educational Researcher, 43*(3), 154-162. doi: 10.3102/0013189x14528039

Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *Journal of Experimental Education, 77*(1), 3-18. doi: 10.3200/jexe.77.1.3-20

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation, 14*(7), 1-11.

Ellerton, N. F. (2013). Engaging pre-service middle-school teacher-education students in mathematical problem posing: development of an active learning framework. *Educational Studies in Mathematics, 83*(1), 87-101. doi: 10.1007/s10649-012-9449-z

Franke, L. M., Kazemi, E., & Battey, D. (2007). Understanding teaching and classroom practice in mathematics. In F. K. Lester (Ed*.), Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 225-257). Charlotte, NC: Information Age Publishing.

Fägerstam, E. S., Joakim. (2012). Learning arithmetic outdoors in junior high school - influence on performance and self-regulating skills. *Education 3-13, 42*(4), 419-431.

Gavin, M. K., Casa, T. M., Adelson, J. L., & Firmender, J. M. (2013). The Impact of Challenging Geometry and Measurement Units on the Achievement of Grade 2 Students. *Journal for Research in Mathematics Education, 44*(3), 478-509.

Gough, D. Oliver, S., & Thomas, J. (2013). *Learning from research: systematic reviews for informing policy decisions.* London: University of London EPPI-Centre.

Gravemeijer, K. & Cobb, P. (2013). Design research from the learning design perspective. In: T. Plomp & N. Nieveen (Eds.). (2013). *Educational design research Part A: An introduction*, p.p. 72-113**,** Enschede: SLO

Gresalfi, M., Barnes, J., & Cross, D. (2012). When does an opportunity become an opportunity? Unpacking classroom practice through the lens of ecological psychology. *Educational Studies in Mathematics, 80*(1-2), 249-267. doi: 10.1007/s10649-011-9367-5

Grouws, D. A., Smith, M. S., & Sztajn, P. (2004). The preparation and teaching practices of U.S. mathematics teachers: Grades 4 and 8. In P. Kloosterman & F. Lester (Eds.), *The 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress: Results and interpretations* (pp. 221-269). Reston, VA: National Council of teachers of mathematics.

Grouws, D. A., Tarr, J. E., Chavez, O., Sears, R., Soria, V. M., & Taylan, R. D. (2013). Curriculum and Implementation Effects on High School Students' Mathematics Learning From Curricula Representing Subject-Specific and Integrated Content Organizations. *Journal for Research in Mathematics Education, 44*(2), 416-463.

Gursel, O., Tekin-Iftar, E., & Bozkurt, F. (2006). Effectiveness of simultaneous prompting in small group: The opportunity of acquiring non-target skills through observational learning and instructive feedback. *Education and Training in Developmental Disabilities, 41*(3), 225-243.

Harwell, M. R., Post, T. R., Cutler, A., Maeda, Y., Anderson, E., Norman, K. W., & Medhanie, A. (2009). The Preparation of Students From National Science Foundation-Funded and Commercially Developed High School Mathematics Curricula for Their First University Mathematics Course. *American Educational Research Journal, 46*(1), 203-231. doi: 10.3102/0002831208323368

Harwell, M. R., Post, T. R., Medhanie, A., Dupuis, D. N., & LeBeau, B. (2013). A Multi-Institutional Study of High School Mathematics Curricula and College Mathematics Achievement and Course Taking. *Journal for Research in Mathematics Education, 44*(5), 742-774.

Haser, C. (2010). Learning to teach in the national curriculum context. *European Journal of Teacher Education, 33*(3), 293-307. doi: 10.1080/02619761003713894

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research,* 77 (1), 81-112.

Hattie, J. (2009). Visible learning: a synthesis of over 800 meta-analyses relating to achievement. London: Routledge.Hattie, J. (2009).

Hemmi, K., & Ryve, A. (2014). Effective mathematics teaching in Finnish and Swedish teacher education discourses. *Journal of Mathematics Teacher Education*. 10.1007/s10857-014-9293-4

Hill, H. C., & Charalambous, C. Y. (2012a). Teacher knowledge, curriculum materials, and quality of instruction: Lessons learned and open issues. *Journal of Curriculum Studies, 44*(4), 559-576. doi: 10.1080/00220272.2012.716978

Hill, H. C., & Charalambous, C. Y. (2012b). Teaching (un)Connected Mathematics: Two teachers' enactment of the Pizza problem. *Journal of Curriculum Studies, 44*(4), 467-487. doi: 10.1080/00220272.2012.716972

Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts. *Second handbook of research on mathematics teaching and learning, 1*, 111-156.

Hiebert, J., & Grouws. D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 371-404). Charlotte, NC: Information Age Publishing.

Hiebert, J., & Wearne, D. (1992). Links Between Teaching and Learning Place Value With Understanding in First Grade. *Journal for Research in Mathematics Education, 23*, 98-122.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and instruction, 14*, 251-283.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *Curriculum journal*, *16*(2), 153-176.

Hong, D. S., & Choi, K. M. (2014). A comparison of Korean and American secondary school textbooks: the case of quadratic equations. *Educational Studies in Mathematics, 85*(2), 241-263. doi: 10.1007/s10649-013-9512-4

Huntley, M. A. (2009). Measuring Curriculum Implementation. *Journal for Research in Mathematics Education, 40*(4), 355-362.

Jansen, A., Herbel-Eisenmann, B., & Smith, J. P. (2012). Detecting Students' Experiences of Discontinuities Between Middle School and High School Mathematics Programs: Learning During Boundary Crossing. *Mathematical Thinking and Learning, 14*(4), 285-309. doi: 10.1080/10986065.2012.717379

Jones, M. H., Alexander, J. M., & Estell, D. B. (2010). Homophily Among Peer Groups Members' Perceived Self-Regulated Learning. *Journal of Experimental Education, 78*(3), 378-394. doi: 10.1080/00220970903548020

Kim, P., & Olaciregui, C. (2008). The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere. *British Journal of Educational Technology, 39*(4), 700-714. doi: 10.1111/j.1467-8535.2007.00763.x

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28-37.

Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., Büttner, G., & Klieme, E. (2010). Promotion of self-regulated learning in classrooms: investigating frequency, quality, and consequences for student performance. *Metacognition and Learning, 5*(2), 157-171. doi: 10.1007/s11409-010-9055-3

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research, 43*(4), 489-510. doi: 10.2190/EC.43.4.d

Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning, 22*(1), 24-33. doi: 10.1111/j.1365-2729.2006.00157.x

Kramarski, B., & Mizrachi, N. (2006). Online discussion and self-regulated learning: Effects of instructional methods on mathematical literacy. *Journal of Educational Research, 99*(4), 218-230. doi: 10.3200/joer.99.4.218-231

Kramarski, B., & Revach, T. (2009). The challenge of self-regulated learning in mathematics teachers' professional training. *Educational studies in mathematics*, *72*(3), 379-399.

Kramer, S. L., & Keller, R. (2008). An existence proof: Successful joint implementation of the IMP curriculum and a 4x4 block schedule at a suburban US high school. *Journal for Research in Mathematics Education, 39*(1), 2-8.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacognition and Learning, 5*(2), 173-194. doi: 10.1007/s11409-010-9056-2

Lazakidou, G., & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers & Education, 54*(1), 3-13. doi: 10.1016/j.compedu.2009.02.020

Lee, H., Feldman, A., & Beatty, I. D. (2012). Factors that Affect Science and Mathematics Teachers' Initial Implementation of Technology-Enhanced Formative Assessment Using a Classroom Response System. *Journal of Science Education and Technology, 21*(5), 523-539. doi: 10.1007/s10956-011-9344-x

Leme, M. C., Louzano, P., Ponczek, V., & Souza, A. P. (2012). The impact of structured teaching methods on the quality of education in Brazil. *Economics of Education Review, 31*(5), 850-860. doi: 10.1016/j.econedurev.2012.05.008

Lesh, R., & Zawojewski, J. S. (2007). Problem-solving and modeling. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 763-804). Charlotte, NC: Information Age Publishing.

Lewis, J. M., & Blunk, M. L. (2012). Reading between the lines: Teaching linear algebra. *Journal of Curriculum Studies, 44*(4), 515-536. doi: 10.1080/00220272.2012.716975

Lloyd, G. M. (2008). Curriculum use while learning to teach: One student teacher's appropriation of mathematics curriculum materials. *Journal for Research in Mathematics Education, 39*(1), 63-94.

Lloyd, G. M., Remillard, J. T., & Herbel-Eisenman, B. A. (2009). Teachers' use of curriculum materials: an emerging field. In J. T. Remillard, B. A. Herbel-Eisenman & G. M. Lloyd (Eds.), *Mathematics teachers at work: connecting curriculum materials and classroom instruction* (pp. 3-14). New York: Routledge.

Lundahl, C. (2011). *Bedömning för lärande*. Stockholm: Norstedts.

Marchis, I. (2012). Self-Regulated Learning and Mathematical Problem-solving. *New Educational Review, 27*(1), 195-208.

Marz, V., & Kelchtermans, G. (2013). Sense-making and structure in teachers' reception of educational reform. A case study on statistics in the mathematics curriculum. *Teaching and Teacher Education, 29*, 13-24. doi: 10.1016/j.tate.2012.08.004

McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the Effect of Formative Assessment on Student Achievement: So Much More is Needed. *Practical Assessment, Research & Evaluation*, *18*(2), 2.

Meaney, T., Trinick, T., & Fairhall, U. (2013). One Size Does NOT Fit All: Achieving Equity in Maori Mathematics Classrooms. *Journal for Research in Mathematics Education, 44*(1), 235-263.

Metallidou, P., & Vlachou, A. (2010). Childrens self-regulated learning profile in language and mathematics: The role of task value beliefs. *Psychology in the Schools, 47*(8), 776-788. doi: 10.1002/pits.20503

Montague, M. (2008). Self-regulation strategies to improve mathematical problem-solving for students with learning disabilities. *Learning Disability Quarterly, 31*(1), 37-44.

Mullis, I., Martin, M., & Foy, P. (2008). TIMSS 2007 International Mathematics Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Mathematics Advisory Panel. (2008). *Chapter 6: Report of the Task Group on Instructional Practices.* Hämtad 12 september 2014, från http://www.ed.gov/about/bdscomm/list/mathpanel/report/instructional-practices.pdf

National Research Council. (2001). In J. Kilpatrick, J. Swafford, and B. Findell (Eds.), *Adding it up: Helping children learn mathematics.* Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington DC: National Academy Press.

Ness, B. M., & Middleton, M. J. (2012). A Framework for Implementing Individualized Self-Regulated Learning Strategies in the Classroom. *Intervention in School and Clinic, 47*(5), 267-275. doi: 10.1177/1053451211430120

Niss, M. (2004). The Danish "KOM" Project and possible consequences for teacher education. In R. Strässer, G. Brandell, B. Grevholm, & O. Helenius (Eds.), *Educating for the future. Proceedings of an International Symposium on Mathematics Teacher Education*, (pp. 179-190). Gothenbutg, Seden: The Royal Swedish Academy of the Sciences, and National Center for Mathematics Education, University of Gothenburg.

Niss, M. (2007). Reflections on the State of and trends in research in mathematics teaching, learning. From here to utopia. In F.K. Lester, Frank K. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1293–1312). Reston: NCTM.

Ocak, G., & Yamac, A. (2013). Examination of the Relationships between Fifth Graders' Self-Regulated Learning Strategies, Motivational Beliefs, Attitudes, and Achievement. *Kuram Ve Uygulamada Egitim Bilimleri, 13*(1), 380-387.

Organisation for Economic Cooperation and Development. (2003). The PISA 2003 assessment framework: Mathematics, reading, science and problem-solving knowledge and skills. Paris: OECD PISA.

Organisation for economic co-operation and development. (2009). PISA 2009 Assessment Framework: Key competencies in reading, mathematics and science. Paris: OECD PISA.

Otten, S., Gilbertson, N. J., Males, L. M., & Clark, D. L. (2014). The Mathematical Nature of Reasoning-and-Proving Opportunities in Geometry Textbooks. *Mathematical Thinking and Learning, 16*(1), 51-79. doi: 10.1080/10986065.2014.857802

Parks, A. N. (2010). Metaphors of hierarchy in mathematics education discourse: the narrow path. *Journal of Curriculum Studies, 42*(1), 79-97. doi: 10.1080/00220270903167743

Pepin, B., Gueudet, G., & Trouche, L. (2013). Investigating textbooks as crucial interfaces between culture, policy and teacher curricular practice: two contrasted case studies in France and Norway. *ZDM, 45*(5), 685-698. doi: 10.1007/s11858-013-0526-2

Pepin, B., & Haggarty, L. (2001). Mathematics textbooks and their use in English, French, and German classrooms: a way to understand teaching and learning cultures. *The International Journal on Mathematics Education (ZDM), 33*(5), 158-175.

Phelan, J., Choi, K., Vendlinski, T., Baker, E., & Herman, J. (2011). Differential Improvement in Student Understanding of Mathematical Principles Following Formative Assessment Intervention. *Journal of Educational Research, 104*(5), 330-339. doi: 10.1080/00220671.2010.484030

Polya, G. (1945/57). *How to solve it*. Princeton: Princeton University Press.

Porter, A., Polikoff, M. S., Barghaus, K. M., & Yang, R. (2013). Constructing Aligned Assessments Using Automated Test Construction. *Educational Researcher, 42*(8), 415-423. doi: 10.3102/0013189x13503038

Post, T. R., Harwell, M. R., Davis, J. D., Maeda, Y., Cutler, A., Andersen, E., . . . Norman, K. W. (2008). Standards-based mathematics curricula and middle-grades students' performance on standardized achievement tests. *Journal for Research in Mathematics Education, 39*(2), 184-212.

Post, T. R., Medhanie, A., Harwell, M., Norman, K. W., Dupuis, D. N., Muchlinski, T., . . . Monson, D. (2010). The Impact of Prior Mathematics Achievement on the Relationship Between High School Mathematics Curricula and Postsecondary Mathematics Performance, Course-Taking, and Persistence. [Article]. *Journal for Research in Mathematics Education, 41*(3), 274-308.

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction, 27*, 63-73. doi: 10.1016/j.learninstruc.2013.03.002

Rakoczy, K., Klieme, E., Burgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction - Grading and feedback in mathematics classrooms. *Zeitschrift Fur Psychologie-Journal of Psychology, 216*(2), 111-124. doi: 10.1027/0044-3409.216.2.111

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211-246. doi: 10.3102/00346543075002211

Remillard, J. T., & Bryans, M. B. (2004). Teachers' orientations toward mathematics curriculum materials: implications for teacher learning. *Journal for Research in Mathematics Education, 35*(5), 352-388.

Rosario, P., Nunez, J. C., Valle, A., Gonzalez-Pienda, J., & Lourenco, A. (2013). Grade level, study time, and grade retention and their effects on motivation, self-regulated learning strategies, and mathematics achievement: a structural equation model. *European Journal of Psychology of Education, 28*(4), 1311-1331. doi: 10.1007/s10212-012-0167-9

Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., & Claro, S. (2010). Scaffolding group explanation and feedback with handheld technology: impact on students' mathematics learning. *Educational Technology Research and Development, 58*(4), 399-419. doi: 10.1007/s11423-009-9142-9

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. [Article]. *American Educational Research Journal, 47*(4), 833-878. doi: 10.3102/0002831210367426

Ruthven, K., Laborde, C., Leach, J., & Tiberghien, A. (2009). Design tools in didactical research: Instrumenting the epistemological and cognitive aspects of the design of teaching sequences. *Educational Researcher, 38*, 329-342.

Samuelsson, J. (2011). Development of self-regulated learning skills in mathematics in lower secondary school in Sweden, *Nordisk matematikkdidaktikk, 16*(3), 25-42.

Schoenfeld, A. (1985). *Mathematical Problem-solving*. New York: Academic Press.

Schoenfeld , A. (1992). Learning to think mathematically: Problem-solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed*.), Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 334-370). New York, NY: Macmillan Publishing Company.

Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies, 41*(4), 467-500. doi: 10.1080/00220270802696115

Shield, M., & Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics, 82*(2), 183-199. doi: 10.1007/s10649-012-9415-9

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Edcational Researcher, 15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Skolverket (2011). Lgr 11. Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011. Stockholm: Skolverket.

Sleep, L. (2012). The Work of Steering Instruction Toward the Mathematical Point: A Decomposition of Teaching Practice. *American Educational Research Journal, 49*(5), 935-970. doi: 10.3102/0002831212448095

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

Smith, M. S. & Stein, M. K. (2011). *5 practices for orchestrating productive mathematical discussions*. Reston, VA: NCTM.

Son, J. W., & Senk, S. L. (2010). How reform curricula in the USA and Korea present multiplication and division of fractions. *Educational Studies in Mathematics, 74*(2), 117-142. doi: 10.1007/s10649-010-9229-6

Stacey, K., & Vincent, J. (2009). Modes of reasoning in explanations in Australian eighth-grade mathematics textbooks. *Educational Studies in Mathematics, 72*(3), 271-288. doi: 10.1007/s10649-009-9193-1

Star, J. R., Smith, J. P., & Jansen, A. (2008). What students notice as different between reform and traditional mathematics programs. *Journal for Research in Mathematics Education, 39*(1), 9-32.

Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban school districts. *American Journal of Education, 114*(4), 583-626. doi: 10.1086/589315

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488. doi: 10.3102/00028312033002455

Stein, M. K., & Kaufman, J. H. (2010). Selecting and Supporting the Use of Mathematics Curricula at Scale. *American Educational Research Journal, 47*(3), 663-693. doi: 10.3102/0002831209361210

Stein, M. K., Remillard, J. T., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319-370). Greenwich, CT: Information Age Publishing.

Stylianides, A. J., & Stylianides, G. J. (2008). Studying the classroom implementation of tasks: High-level mathematical tasks embedded in 'real-life' contexts. *Teaching and Teacher Education, 24*(4), 859-875. doi: 10.1016/j.tate.2007.11.015

Stylianides, G. J. (2009). Reasoning-and-Proving in School Mathematics Textbooks. [Article]. *Mathematical Thinking and Learning, 11*(4), 258-288. doi: 10.1080/10986060903253954

Sun, X. H. (2011). "Variation problems" and their roles in the topic of fraction division in Chinese mathematics textbook examples. *Educational Studies in Mathematics, 76*(1), 65-85. doi: 10.1007/s10649-010-9263-4

Swedish National Agency for Education. (2012). Curriculum for the compulsory school system, the pre-school class and the leisure-time centre 2011

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247-280.

Taylor, M. W. (2013). Replacing the 'teacher-proof' curriculum with the 'curriculum-proof' teacher: Toward more effective interactions with mathematics textbooks. *Journal of Curriculum Studies, 45*(3), 295-321. doi: 10.1080/00220272.2012.710253

Thompson, D. R., Senk, S. L., & Johnson, G. J. (2012). Opportunities to Learn Reasoning and Proof in High School Mathematics Textbooks. *Journal for Research in Mathematics Education, 43*(3), 253-295.

Throndsen, I. (2011). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology, 81*(4), 558-578. doi: 10.1348/2044-8279.002008

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houng, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht: Kluwer Academic Publishers.

Van Steenbrugge, H., Valcke, M., & Desoete, A. (2013). Teachers' views of mathematics textbook series in Flanders: Does it (not) matter which mathematics textbook series schools choose? *Journal of Curriculum Studies, 45*(3), 322-353. doi: 10.1080/00220272.2012.713995

Villegas-Reimers, E. (2003). *Teacher professional development: an international review of the literature*. Paris: International Institute for Educational Planning.

Vingsle, C. (2014a). *Formative assessment: Teacher knowledge and skills to make it happen*. (Licentiatavhandling, Umeå universitet, Institutionen för naturvetenskapernas och matematikens didaktik). Från http://umu.diva-portal.org/smash/get/diva2:735415/FULLTEXT01.pdf

Vingsle, C. (2014b). *How hard can it be? What knowledge and skills does a teacher practicing formative assessment use?* Presentation at Nionde matematikdidaktiska forskningsseminariet (Madif-9), Sweden. Från:http://ncm.gu.se/media/smdf/web_versions/shorts/S17_shortcom_Madif9_%20Lotta_Vingsle_revised.pdf

Voogt, J. (2010). Teacher factors associated with innovative curriculum goals and pedagogical practices: differences between extensive and non-extensive ICT-using science teachers. *Journal of Computer Assisted Learning, 26*(6), 453-464. doi: 10.1111/j.1365-2729.2010.00373.x

Walshaw, M, & Anthony, G. (2008). The teacher's role in classroom discourse: a review of recent research into mathematics classrooms. *Review of Educational Research, 78*, pp. 516-551.

Wiliam, D. (2013). *Att följa lärande–formativ bedömning i praktiken.* Lund: Studentlitteratur.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, P. S., Cooney, T. J., & Stinson, D. W. (2005). What constitutes good mathematics teaching and how it develops: Nine high school teachers' perspectives. *Journal of Mathematics Teacher Education, 8*, 83–111.

Yackel, E. (2000): *Creating a mathematics classroom environment that fosters the development of mathematical argumentation*. Paper presented at the Ninth International Congress of Mathematical Education, Tokyo.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*, 458-477.

Yeh, S. S. (2009). Class size reduction or rapid formative assessment? A comparison of cost-effectiveness. *Educational Research Review, 4*(1), 7-15. doi: 10.1016/j.edurev.2008.09.001

Xu, L., & Clarke, D. (2013). Meta-rules of discursive practice in mathematics classrooms from Seoul, Shanghai and Tokyo. ZDM - The International Journal on Mathematics Education, 45(1), 61–72.

APPENDIX A – SEARCH STRINGS

Search strings used in Web of Science Core Collection:

#16 #15 OR #14 OR #13 OR #12 OR #11 OR #10 DocType=All document types; Language=All languages;

#15 ((TS=(math* AND ("teach* move*" OR "instruction* move*" OR "pedagogic* move*")))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Review) DocType=All document types; Language=All languages;

#14 ((TS=(math* AND ("teach* model*" OR "instruction* model*" OR "pedagogic* model*")))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Review) DocType=All document types; Language=All languages;

#13 (TS=(math* AND ("teach* approach*" OR "instruction* approach*" OR "pedagogic* approach*"))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Review) DocType=All document types; Language=All languages;

#12 (TS=(math* AND ("teach* method*" OR "instruction* method*" OR "pedagogic* method*"))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Review) DocType=All document types; Language=All languages;

#11 (TS=(math* AND ("teach* strateg*" OR "instruction* strateg*" OR "pedagogic* strateg*"))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Review) DocType=All document types; Language=All languages;

#10 (TS=(math* AND ("teach* practice*" OR "instruction* practice*" OR "pedagogic* practice*"))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Review) DocType=All document types; Language=All languages;

APPENDIX B – CODING MATRIX OF ABSTRACTS

Promoting the learning of mathematical processes

IS+Proc&SA	Gresalfi, M., Barnes, J. and Cross, D.	2012	When does an opportunity become an opportunity? Unpacking classroom practice through the lens of ecological psychology	Educational Studies in Mathematics, 80(1-2), 249-267
IS+Proc	Matsumura, L. C., Slater, S. C. and Crosson, A.	2008	Classroom climate, rigorous instruction and curriculum, and students' interactions in urban middle schools	Elementary School Journal, 108(4), 293-312
IS+Proc	Webb, N. M., Franke, M. L., Ing, M., Chan, A., De, T., Freund, D. and Battey, D.	2008	The role of teacher instructional practices in student collaboration	Contemporary Educational Psychology, 33(3), 360-381
TA+Proc&Prod	Carbonneau, K. J., Marley, S. C. and Selig, J. P.	2013	A Meta-Analysis of the Efficacy of Teaching Mathematics With Concrete Manipulatives	Journal of Educational Psychology, 105(2), 380-400
TA+Proc&SA	Cotic, M. and Zuljan, M. V.	2009	Problem-based instruction in mathematics and its impact on the cognitive results of the students and on affective-motivational aspects	Educational Studies, 35(3), 297-310
TA+Proc	Jitendra, A. K. and Star, J. R.	2011	Meeting the Needs of Students With Learning Disabilities in Inclusive Mathematics Classrooms: The Role of Schema-Based Instruction on Mathematical Problem-Solving	Theory into Practice, 50(1), 12-19
TA+Proc	Kotsopoulos, D. and Lee, J.	2012	An Analysis of Math Congress in an Eighth Grade Classroom	Mathematical Thinking and Learning, 14(3), 181-198

TA+Proc&Prod	Le, V. N., Lockwood, J. R., Stecher, B. M., Hamilton, L. S. and Martinez, J. F.	2009	A Longitudinal Investigation of the Relationship between Teachers' Self-Reports of Reform-Oriented Instruction and Mathematics and Science Achievement	Educational Evaluation and Policy Analysis, 31(3), 200-220
TA+Proc	Montague, M.	2008	Self-regulation strategies to improve mathematical problem-solving for students with learning disabilities	Learning Disability Quarterly, 31(1), 37-44
TA+Proc&Prod	Rittle-Johnson, B., Star, J. R. and Durkin, K.	2012	Developing procedural flexibility: Are novices prepared to learn from comparing procedures?	British Journal of Educational Psychology, 82(3), 436-455
TA+Proc och BV+P	Song, Y. J. and Looi, C. K.	2012	Linking teacher beliefs, practices and student inquiry-based learning in a CSCL environment: A tale of two teachers	International Journal of Computer-Supported Collaborative Learning, 7(1), 129-159
TA+Proc&Prod BV+Proc&Prod	Thorvaldsen, S., Vavik, L. and Salomon, G.	2012	The Use of ICT Tools in Mathematics: A Case-control Study of Best Practice in 9th Grade Classrooms	Scandinavian Journal of Educational Research, 56(2), 213-228
TA+Proc	Xin, Y. P., Zhang, D. K., Park, J. Y., Tom, K., Whipple, A. and Si, L.	2011	A Comparison of Two Mathematics Problem-Solving Strategies: Facilitate Algebra-Readiness	Journal of Educational Research, 104(6), 381-395
LM+Proc&Prod	Hattikudur, S. and Alibali, M. W.	2010	Learning about the equal sign: Does comparing with inequality symbols help?	Journal of Experimental Child Psychology, 107(1), 15-30
LM+Proc	Krawec, J., Huang, J., Montague, M., Kressler, B. and de Alba, A. M.	2013	The Effects of Cognitive Strategy Instruction on Knowledge of Math Problem-Solving Processes of Middle School Students With Learning Disabilities	Learning Disability Quarterly, 36(2), 80-92
LM+Proc	Orosco, M. J.	2014	Word Problem Strategy for Latino English Language Learners at Risk for Math Disabilities	Learning Disability Quarterly, 37(1), 45-53

LM+Proc	Scheiter, K., Gerjets, P. and Schuh, J.	2010	The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures	Instructional Science, 38(5), 487-502
LM+Proc	Star, J. R. and Rittle-Johnson, B.	2009	It pays to compare: An experimental study on computational estimation	Journal of Experimental Child Psychology, 102(4), 408-426

Promoting the learning of mathematical products

IS+Prod	Adams, A. E. and Pegg, J.	2012	Teachers' Enactment of Content Literacy Strategies in Secondary Science and Mathematics Classes	Journal of Adolescent & Adult Literacy, 56(2), 151-161
IS+Prod	Akyuz, D., Stephan, M. and Dixon, J. K.	2012	The Role of the Teacher in Supporting Imagery in Understanding Integers	Egitim Ve Bilim-Education and Science, 37(163), 268-282
IS+Prod&SA	Davis, N., Sheldon, L. and Colmar, S.	2014	Memory Mates: A Classroom-Based Intervention to Improve Attention and Working Memory	Australian Journal of Guidance and Counselling, 24(1), 111-120
IS+Prod	Grossman, P., Loeb, S., Cohen, J. and Wyckoff, J.	2013	Measure for Measure: The Relationship between Measures of Instructional Practice in Middle School English Language Arts and Teachers' Value-Added Scores	American Journal of Education, 119(3), 445-470
IS+Prod	Panayiotou, A., Kyriakides, L., Creemers, B. P. M., McMahon, L., Vanlaar, G., Pfeifer, M., Rekalidou, G. and Bren, M.	2014	Teacher behavior and student outcomes: Results of a European study	Educational Assessment Evaluation and Accountability, 26(1), 73-93
IS+Prod&SA	Rakoczy, K., Klieme, E., Burgermeister, A. and Harks, B.	2008	The interplay between student evaluation and instruction - Grading and feedback in mathematics classrooms	Zeitschrift Fur Psychologie-Journal of Psychology, 216(2), 111-124
IS+Prod	Richland, L. E., Stigler, J. W. and Holyoak, K. J.	2012	Teaching the Conceptual Structure of Mathematics	Educational Psychologist, 47(3), 189-203

IS+Prod	Shein, P. P.	2012	Seeing With Two Eyes: A Teacher's Use of Gestures in Questioning and Revoicing to Engage English Language Learners in the Repair of Mathematical Errors	Journal for Research in Mathematics Education, 43(2), 182-222
IS+Prod	Stronge, J. H., Ward, T. J. and Grant, L. W.	2011	What Makes Good Teachers Good? A Cross-Case Analysis of the Connection Between Teacher Effectiveness and Student Achievement	Journal of Teacher Education, 62(4), 339-355
IS+Prod and TA+Prod	Misquitta, R.	2011	A Review of the Literature: Fraction Instruction for Struggling Learners in Mathematics	Learning Disabilities Research & Practice, 26(2), 109-119
TA+Prod and Char	Allsopp, D. H., McHatton, P. A. and Farmer, J. L.	2010	TECHNOLOGY, MATHEMATICS PS/RTI, AND STUDENTS WITH LD: WHAT DO WE KNOW, WHAT HAVE WE TRIED, AND WHAT CAN WE DO TO IMPROVE OUTCOMES NOW AND IN THE FUTURE?	Learning Disability Quarterly, 33(4), 273-288
TA+Prod	Andrews, P., Ryve, A., Hemmi, K. and Sayers, J.	2014	PISA, TIMSS and Finnish mathematics teaching: an enigma in search of an explanation	Educational Studies in Mathematics, 87(1), 7-26
TA+Prod	Ayres, P.	2013	Can the isolated-elements strategy be improved by targeting points of high cognitive load for additional practice?	Learning and Instruction, 23, 115-124
TA+Prod	Boaler, J.	2008	Promoting 'relational equity' and high mathematics achievement through an innovative mixed-ability approach	British Educational Research Journal, 34(2), 167-194
TA+Prod och TA+SA	Boaler, J. and Staples, M.	2008	Creating Mathematical Futures through an Equitable Teaching Approach: The Case of Railside School	Teachers College Record, 110(3), 608-645

TA+Prod	Bragg, L. A.	2012	TESTING THE EFFECTIVENESS OF MATHEMATICAL GAMES AS A PEDAGOGICAL TOOL FOR CHILDREN'S LEARNING	International Journal of Science and Mathematics Education, 10(6), 1445-1467
TA+Prod	Braithwaite, D. W. and Goldstone, R. L.	2013	Integrating Formal and Grounded Representations in Combinatorics Learning	Journal of Educational Psychology, 105(3), 666-682
TA+Prod	Brophy, S., Klein, S., Portsmore, M. and Rogers, C.	2008	Advancing Engineering Education in P-12 Classrooms	Journal of Engineering Education, 97(3), 369-387
TA+Prod	Chang, M.	2008	Teacher Instructional Practices and Language Minority Students: A Longitudinal Model	Journal of Educational Research, 102(2), 83-97
TA+Prod	Cho, S. J., Bottge, B. A., Cohen, A. S. and Kim, S. H.	2011	Detecting Cognitive Change in the Math Skills of Low-Achieving Adolescents	Journal of Special Education, 45(2), 67-76
TA+Prod and Char	Croninger, R. G., Buese, D. and Larson, J.	2012	A Mixed-Methods Look at Teaching Quality: Challenges and Possibilities From One Study	Teachers College Record, 114(4), 36
TA+Prod and BV+P	Desimone, L. M., Smith, T. M. and Phillips, K. J. R.	2013	Linking Student Achievement Growth to Professional Development Participation and Changes in Instruction: A Longitudinal Study of Elementary Students and Teachers in Title I Schools	Teachers College Record, 115(5), 46
TA+Prod	Even, R. and Kvatinsky, T.	2010	What mathematics do teachers with contrasting teaching approaches address in probability lessons?	Educational Studies in Mathematics, 74(3), 207-222
TA+Prod	Flores, M. M., Hinton, V. and Strozier, S. D.	2014	Teaching Subtraction and Multiplication with Regrouping Using the Concrete-Representational-Abstract Sequence and Strategic Instruction Model	Learning Disabilities Research & Practice, 29(2), 75-88
TA+Prod and IS + Prod	Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P. and Flojo, J.	2009	Mathematics Instruction for Students With Learning Disabilities: A Meta-Analysis of Instructional Components	Review of Educational Research, 79(3), 1202-1242

TA+Prod	Groth, R. E.	2009	Characteristics of teachers' conversations about teaching mean, median, and mode	Teaching and Teacher Education, 25(5), 707-716
TA+Prod and BV+Prod	Guarino, C., Dieterle, S. G., Bargagliotti, A. E. and Mason, W. M.	2013	What Can We Learn About Effective Early Mathematics Teaching? A Framework for Estimating Causal Effects Using Longitudinal Survey Data	Journal of Research on Educational Effectiveness, 6(2), 164-198
TA+Prod	Haeck, C., Lefebvre, P. and Merrigan, P.	2014	The distributional impacts of a universal school reform on mathematical achievements: A natural experiment from Canada	Economics of Education Review, 41(137-160
TA+Prod	Hlas, A. C. and Hlas, C. S.	2012	A Review of High-Leverage Teaching Practices: Making Connections Between Mathematics and Foreign Languages	Foreign Language Annals, 45, 76-97.
TA+Prod and Char	Jitendra, A. K.	2013	Understanding and Accessing Standards-Based Mathematics for Students With Mathematics Difficulties	Learning Disability Quarterly, 36(1), 4-8
TA+Prod	Kay, R. H. and Knaack, L.	2008	An examination of the impact of learning objects in secondary school	Journal of Computer Assisted Learning, 24(6), 447-461
TA+Prod&SA and BV+Prod&SA	Kikas, E., Peets, K. and Hodges, E. V. E.	2014	Collective student characteristics alter the effects of teaching practices on academic outcomes	Journal of Applied Developmental Psychology, 35(4), 273-283
TA+Prod	Mathur, S. R. and Schoenfeld, N.	2010	Effective Instructional Practices in Juvenile Justice Facilities	Behavioral Disorders, 36(1), 20-27
TA+Prod	Ngware, M. W., Oketch, M. and Mutisya, M.	2014	Does teaching style explain differences in learner achievement in low and high performing schools in Kenya?	International Journal of Educational Development, 36, 3-12
TA+Prod	Rakes, C. R., Valentine, J. C., McGatha, M. B. and Ronau, R. N.	2010	Methods of Instructional Improvement in Algebra: A Systematic Review and Meta-Analysis	Review of Educational Research, 80(3), 372-400

TA+Prod&SA	Ridlon, C. L.	2009	Learning Mathematics via a Problem-Centered Approach: A Two-Year Study	Mathematical Thinking and Learning, 11(4), 188-225
TA+Prod	Rimm-Kaufman, S. E., Larsen, R. A. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., Merritt, E. G., Abry, T. and DeCoster, J.	2014	Efficacy of the Responsive Classroom Approach: Results From a 3-Year, Longitudinal Randomized Controlled Trial	American Educational Research Journal, 51(3), 567-603
TA+Prod&SA	Ruthven, K.	2011	USING INTERNATIONAL STUDY SERIES AND META-ANALYTIC RESEARCH SYNTHESES TO SCOPE PEDAGOGICAL DEVELOPMENT AIMED AT IMPROVING STUDENT ATTITUDE AND ACHIEVEMENT IN SCHOOL MATHEMATICS AND SCIENCE	International Journal of Science and Mathematics Education, 9(2), 419-458
TA+Prod	Scheuermann, A. M., Deshler, D. D. and Schumaker, J. B.	2009	THE EFFECTS OF THE EXPLICIT INQUIRY ROUTINE ON THE PERFORMANCE OF STUDENTS WITH LEARNING DISABILITIES ON ONE-VARIABLE EQUATIONS	Learning Disability Quarterly, 32(2), 103-120
TA+Prod	Serow, P. and Callingham, R.	2011	Levels of use of Interactive Whiteboard technology in the primary mathematics classroom	Technology Pedagogy and Education, 20(2), 161-173
TA+Prod	Slavin, R. E. and Lake, C.	2008	Effective programs in elementary mathematics: A best-evidence synthesis	Review of Educational Research, 78(3), 427-515
TA+Prod	Slavin, R. E., Lake, C. and Groff, C.	2009	Effective Programs in Middle and High School Mathematics: A Best-Evidence Synthesis	Review of Educational Research, 79(2), 839-911
TA+Prod	Sleep, L.	2012	The Work of Steering Instruction Toward the Mathematical Point: A Decomposition of Teaching Practice	American Educational Research Journal, 49(5), 935-970
TA+Prod	Sztajn, P., Confrey, J., Wilson, P. H. and Edgington, C.	2012	Learning Trajectory Based Instruction: Toward a Theory of Teaching	Educational Researcher, 41(5), 147-156

TA+Prod	Valle, M. S., Diaz, Z., Waxman, H. C. and Padron, Y. N.	2013	Classroom Instruction and the Mathematics Achievement of Non-English Learners and English Learners	Journal of Educational Research, 106(3), 173-182
TA+Prod&SA	Walker, J. M. T.	2008	Looking at teacher practices through the lens of parenting style	Journal of Experimental Education, 76(2), 218-240
TA+Prod	Wiley, C. R. H., Good, T. L. and McCaslin, M.	2008	Comprehensive School Reform Instructional Practices Throughout a School Year: The Role of Subject Matter, Grade Level, and Time of Year	Teachers College Record, 110(11), 2361-2388
TA+Prod	Woolley, M. E., Rose, R. A., Orthner, D. K., Akos, P. T. and Jones-Sanpei, H.	2013	Advancing Academic Achievement Through Career Relevance in the Middle Grades: A Longitudinal Evaluation of CareerStart	American Educational Research Journal, 50(6), 1309-1335
TA+Prod	Xin, Y. P.	2008	The Effect of Schema-Based Instruction in Solving Mathematics Word Problems: An Emphasis on Prealgebraic Conceptualization of Multiplicative Relations	Journal for Research in Mathematics Education, 39(5), 526-551
TA+Prod	Xin, Y. P. and Zhang, D.	2009	Exploring a Conceptual Model-Based Approach to Teaching Situated Word Problems	Journal of Educational Research, 102(6), 427-441
TA+Prod	Zakelj, A. and Grmek, M. I.	2013	Ability Grouping and Pupils' Results on the National Assessment of Knowledge	Croatian Journal of Education-Hrvatski Casopis Za Odgoj I Obrazovanje, 15(2), 439-463
TA+Prod	Zakharov, A., Carnoy, M. and Loyalka, P.	2014	Which teaching practices improve student perform-ance on high-stakes exams? Evidence from Russia	International Journal of Educational Development, 36, 13-21
LM+Prod	Allsopp, D. H., McHatton, P. A. and Farmer, J. L.	2010	TECHNOLOGY, MATHEMATICS PS/RTI, AND STUDENTS WITH LD: WHAT DO WE KNOW, WHAT HAVE WE TRIED, AND WHAT CAN WE DO TO IMPROVE OUTCOMES NOW AND IN THE FUTURE?	Learning Disability Quarterly, 33(4), 273-288

LM+Prod	Chiu, M. S.	2009	APPROACHES TO THE TEACHING OF CREATIVE AND NON-CREATIVE MATHEMATICAL PROBLEMS	International Journal of Science and Mathematics Education, 7(1), 55-79
LM+Prod	David, M. M. and Tomaz, V. S.	2012	The role of visual representations for structuring classroom mathematical activity	Educational Studies in Mathematics, 80(3), 413-431
LM+Prod	Fuchs, L. S., Fuchs, D. and Compton, D. L.	2013	Intervention Effects for Students With Comorbid Forms of Learning Disability: Understanding the Needs of Nonresponders	Journal of Learning Disabilities, 46(6), 534-548
LM+Prod	Gravemeijer, K. and van Eerde, D.	2009	Design Research as a Means for Building a Knowledge Base for Teachers and Teaching in Mathematics Education	Elementary School Journal, 109(5), 510-524
LM+Prod	Ng, S. F. and Lee, K.	2009	The Model Method: Singapore Children's Tool for Representing and Solving Algebraic Word Problems	Journal for Research in Mathematics Education, 40(3), 282-313
LM+Prod	Ngu, B. H., Yeung, A. S. and Tobias, S.	2014	Cognitive load in percentage change problems: unitary, pictorial, and equation approaches to instruction	Instructional Science, 42(5), 685-713
LM+Prod	Obersteiner, A., Reiss, K. and Ufer, S.	2013	How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills	Learning and Instruction, 23, 125-135
LM+Prod	Powell, S. R., Fuchs, L. S. and Fuchs, D.	2010	Embedding Number-Combinations Practice Within Word-Problem Tutoring	Intervention in School and Clinic, 46(1), 22-30
LM+Prod	Ruthven, K.	2008	THE INTERPRETATIVE FLEXIBILITY, INSTRUMENTAL EVOLUTION, AND INSTITUTIONAL ADOPTION OF MATHEMATICAL SOFTWARE IN EDUCATIONAL PRACTICE: THE EXAMPLES OF COMPUTER ALGEBRA AND DYNAMIC GEOMETRY	Journal of Educational Computing Research, 39(4), 379-394

LM+Prod	Silver, E. A., Mesa, V. M., Morris, K. A., Star, J. R. and Benken, B. M.	2009	Teaching Mathematics for Understanding: An Analysis of Lessons Submitted by Teachers Seeking NBPTS Certification	American Educational Research Journal, 46(2), 501-531
LM+Prod	Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J. and Osterlind, S. J.	2008	The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement	Journal for Research in Mathematics Education, 39(3), 247-280
LM+Prod	Tsai, Y. L. and Chang, C. K.	2009	USING COMBINATORIAL APPROACH TO IMPROVE STUDENTS' LEARNING OF THE DISTRIBUTIVE LAW AND MULTIPLICATIVE IDENTITIES	International Journal of Science and Mathematics Education, 7(3), 501-531
LM+Prod	van Loon-Hillen, N., van Gog, T. and Brand-Gruwel, S.	2012	Effects of worked examples in a primary school mathematics curriculum	Interactive Learning Environments, 20(1), 89-99
BV+Prod	Chernoff, E. J. and Zazkis, R.	2011	From personal to conventional probabilities: from sample set to sample space	Educational Studies in Mathematics, 77(1), 15-33
BV+Prod&SA	Argentin, G., Pennisi, A., Vidoni, D., Abbiati, G. and Caputo, A.	2014	Trying to Raise (Low) Math Achievement and to Promote (Rigorous) Policy Evaluation in Italy: Evidence From a Large-Scale Randomized Trial	Evaluation Review, 38(2), 99-132
BV+Prod	Aslam, M. and Kingdon, G.	2011	What can teachers do to raise pupil achievement?	Economics of Education Review, 30(3), 559-574
BV+Prod	Boonen, T., Van Damme, J. and Onghena, P.	2014	Teacher effects on student achievement in first grade: which aspects matter most?	School Effectiveness and School Improvement, 25(1), 126-152
BV+Prod	Brown, S. A., Pitvorec, K., Ditto, C. and Kelso, C. R.	2009	Reconceiving Fidelity of implementation: An Investigation of Elementary Whole-Number Lessons	Journal for Research in Mathematics Education, 40(4), 363-395
BV+Prod	Desimone, L. M. and Long, D.	2010	Teacher Effects and the Achievement Gap: Do Teacher and Teaching Quality Influence the Achievement Gap Between Black and White and High- and Low-SES Students in the Early Grades?	Teachers College Record, 112(12), 3024-3073

BV+Prod	Imbo, I. and LeFevre, J. A.	2009	Cultural Differences in Complex Addition: Efficient Chinese Versus Adaptive Belgians and Canadians	Journal of Experimental Psychology-Learning Memory and Cognition, 35(6), 1465-1476
BV+Prod	Kikas, E., Peets, K., Palu, A. and Afanasjev, J.	2009	The role of individual and contextual factors in the development of maths skills	Educational Psychology, 29(5), 541-560
BV+Prod	Kuzniak, A. and Rauscher, J. C.	2011	How do teachers' approaches to geometric work relate to geometry students' learning difficulties?	Educational Studies in Mathematics, 77(1), 129-147
BV+Prod	Leme, M. C., Louzano, P., Ponczek, V. and Souza, A. P.	2012	The impact of structured teaching methods on the quality of education in Brazil	Economics of Education Review, 31(5), 850-860
BV+Prod	Lubienski, S. T., Lubienski, C. and Crane, C. C.	2008	Achievement Differences and School Type: The Role of School Climate, Teacher Certification, and Instruction	American Journal of Education, 115(1), 97-138
BV+Prod	Maheady, L. and Gard, J.	2010	Classwide Peer Tutoring: Practice, Theory, Research, and Personal Narrative	Intervention in School and Clinic, 46(2), 71-78
BV+Prod and TA+Prod	Palardy, G. J. and Rumberger, R. W.	2008	Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning	Educational Evaluation and Policy Analysis, 30(2), 111-140
BV+Prod	Tchoshanov, M. A.	2011	Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics	Educational Studies in Mathematics, 76(2), 141-164
BV+Prod	Throndsen, I. and Turmo, A.	2013	Primary mathematics teachers' goal orientations and student achievement	Instructional Science, 41(2), 307-322
BV+Prod and BV+P	Wallace, M. R.	2009	Making Sense of the Links: Professional Development, Teacher Practices, and Student Achievement	Teachers College Record, 111(2), 573-596

Student attributes

IS+SA	Battey, D.	2013	Good mathematics teaching for students of color and those in poverty: the importance of relational interactions within instruction	Educational Studies in Mathematics, 82(1), 125-144
IS+SA	Berry, R. A. W. and Kim, N.	2008	Exploring teacher talk during mathematics instruction in an inclusion classroom	Journal of Educational Research, 101(6), 363-377
IS+SA	Rattan, A., Good, C. and Dweck, C. S.	2012	It's ok - Not everyone can be good at math: Instructors with an entity theory comfort (and demotivate) students	Journal of Experimental Social Psychology, 48(3), 731-737
IS+SA	Williams-Johnson, M., Cross, D., Hong, J., Aultman, L., Osbon, J. and Schutz, P.	2008	There Are No Emotions in Math: How Teachers Approach Emotions in the Classroom	Teachers College Record, 110(8), 1574-1610
TA+SA	Baker, J. A., Clark, T. P., Maier, K. S. and Viger, S.	2008	The differential influence of instructional context on the academic engagement of students with behavior problems	Teaching and Teacher Education, 24(7), 1876-1883
TA+SA	Baki, A. and Cakiroglu, U.	2010	Learning objects in high school mathematics classrooms: Implementation and evaluation	Computers & Education, 55(4), 1459-1469
TA+SA	Beausaert, S. A. J., Segers, M. S. R. and Wiltink, D. P. A.	2013	The influence of teachers' teaching approaches on students' learning approaches: the student perspective	Educational Research, 55(1), 1-15
TA+SA	Billingsley, G., Scheuermann, B. and Webber, J.	2009	A Comparison of Three Instructional Methods for Teaching Math Skills to Secondary Students With Emotional/Behavioral Disorders	Behavioral Disorders, 35(1), 4-18
TA+SA	Jansen, A., Herbel-Eisenmann, B. and Smith, J. P.	2012	Detecting Students' Experiences of Discontinuities Between Middle School and High School Mathematics Programs: Learning During Boundary Crossing	Mathematical Thinking and Learning, 14(4), 285-309

TA+SA	Leonard, J., Brooks, W., Barnes-Johnson, J. and Berry, R. Q.	2010	The Nuances and Complexities of Teaching Mathematics for Cultural Relevance and Social Justice	Journal of Teacher Education, 61(3), 261-270
TA+SA	Lynch, K. and Star, J. R.	2014	Views of Struggling Students on Instruction Incorporating Multiple Strategies in Algebra I: An Exploratory Study	Journal for Research in Mathematics Education, 45(1), 6-18
TA+SA	Thoonen, E. E. J., Sleegers, P. J. C., Peetsma, T. T. D. and Oort, F. J.	2011	Can teachers motivate students to learn?	Educational Studies, 37(3), 345-360
LM+SA	Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Muller, M. and Messner, R.	2012	Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations	Educational Studies in Mathematics, 79(2), 215-237
LM+SA	Sengul, S. and Dereli, M.	2013	The Effect of Learning Integers Using Cartoons on 7th Grade Students' Attitude to Mathematics	Kuram Ve Uygulamada Egitim Bilimleri, 13(4), 2526-2534
TA+SA&Prod and BV+SA&Prod	Riconscente, M. M.	2014	Effects of Perceived Teacher Practices on Latino High School Students' Interest, Self-Efficacy, and Achievement in Mathematics	Journal of Experimental Education, 82(1), 51-73
BV+SA	Birgin, O., Baloglu, M., Catlioglu, H. and Gurbuz, R.	2010	An investigation of mathematics anxiety among sixth through eighth grade students in Turkey	Learning and Individual Differences, 20(6), 654-658

Practices

IS+P	Akkus, R.	2013	Following Students' Ideas: How Much to Let Go?	Egitim Ve Bilim-Education and Science, 38(169), 96-108
IS+P	Akkus, R. and Hand, B.	2011	EXAMINING TEACHERS' STRUGGLES AS THEY ATTEMPT TO IMPLEMENT DIALOGICAL INTERACTION AS PART OF PROMOTING MATHEMATICAL REASONING WITHIN THEIR CLASSROOMS	International Journal of Science and Mathematics Education, 9(4), 975-998

IS+P	Bozack, A. R., Vega, R., McCaslin, M. and Good, T. L.	2008	Teacher Support of Student Autonomy in Comprehensive School Reform Classrooms	Teachers College Record, 110(11), 2389-2407
IS+P	Brodie, K.	2011	Working with learners' mathematical thinking: Towards a language of description for changing pedagogy	Teaching and Teacher Education, 27(1), 174-186
IS+P	Harel, G. and Rabin, J. M.	2010	Teaching Practices Associated With the Authoritative Proof Scheme	Journal for Research in Mathematics Education, 41(1), 14-19
IS+P	Harkness, S. S.	2009	Social constructivism and the Believing Game: a mathematics teacher's practice and its implications	Educational Studies in Mathematics, 70(3), 243-258
IS+P	Murata, A.	2013	Diversity and High Academic Expectations Without Tracking: Inclusively Responsive Instruction	Journal of the Learning Sciences, 22(2), 312-335
IS+P	Ottmar, E. R., Rimm-Kaufman, S. E., Berry, R. Q. and Larsen, R. A.	2013	DOES THE RESPONSIVE CLASSROOM APPROACH AFFECT THE USE OF STANDARDS-BASED MATHEMATICS TEACHING PRACTICES? Results from a Randomized Controlled Trial	Elementary School Journal, 113(3), 434-457
IS+P	Radford, J., Blatchford, P. and Webster, R.	2011	Opening up and closing down: How teachers and TAs manage turn-taking, topic and repair in mathematics lessons	Learning and Instruction, 21(5), 625-635
IS+P	Truxaw, M. P. and DeFranco, T. C.	2008	Mapping Mathematics Classroom Discourse and Its Implications for Models of Teaching	Journal for Research in Mathematics Education, 39(5), 489-525
IS+P	Webb, N. M.	2009	The teacher's role in promoting collaborative dialogue in the classroom	British Journal of Educational Psychology, 79, 1-28
LM+P	Bieda, K. N.	2010	Enacting Proof-Related Tasks in Middle School Mathematics: Challenges and Opportunities	Journal for Research in Mathematics Education, 41(4), 351-382

LM+P	Bonotto, C.	2013	Artifacts as sources for problem-posing activities	Educational Studies in Mathematics, 83(1), 37-55
LM+P	Boston, M. D. and Smith, M. S.	2009	Transforming Secondary Mathematics Teaching: Increasing the Cognitive Demands of Instructional Tasks Used in Teachers' Classrooms	Journal for Research in Mathematics Education, 40(2), 119-156
LM+P	Dolonen, J. A. and Ludvigsen, S.	2012	Analyzing students' interaction with a 3D geometry learning tool and their teacher	Learning Culture and Social Interaction, 1(3-4), 167-182
LM+P and IS+P	Ni, Y. J., Zhou, D. H., Li, X. Q. and Li, Q.	2014	Relations of Instructional Tasks to Teacher-Student Discourse in Mathematics Classrooms of Chinese Primary Schools	Cognition and Instruction, 32(1), 2-43
LM+P	Tay, L. Y., Lim, S. K., Lim, C. P. and Koh, J. H. L.	2012	Pedagogical approaches for ICT integration into primary school English and mathematics: A Singapore case study	Australasian Journal of Educational Technology, 28(4), 740-754
LM+P	Wood, R. and Ashfield, J.	2008	The use of the interactive whiteboard for creative teaching and learning in literacy and mathematics: a case study	British Journal of Educational Technology, 39(1), 84-96
BV+P	Amador, J. and Lamberg, T.	2013	Learning Trajectories, Lesson Planning, Affordances, and Constraints in the Design and Enactment of Mathematics Teaching	Mathematical Thinking and Learning, 15(2), 146-170
BV+P	Boerst, T., Sleep, L., Ball, D. and Bass, H.	2011	Preparing Teachers to Lead Mathematics Discussions	Teachers College Record, 113(12), 2844-2877
BV+P	Bray, W. S.	2011	A Collective Case Study of the Influence of Teachers' Beliefs and Knowledge on Error-Handling Practices During Class Discussion of Mathematics	Journal for Research in Mathematics Education, 42(1), 2-38
BV+P	Bonner, E. P.	2014	Investigating practices of highly successful mathematics teachers of traditionally underserved students	Educational Studies in Mathematics, 86(3), 377-399

BV+P	Chai, C. S., Koh, J. H. L. and Tsai, C. C.	2013	A Review of Technological Pedagogical Content Knowledge	Educational Technology & Society, 16(2), 31-51
BV+P	Charalambous, C. Y. and Philippou, G. N.	2010	Teachers' concerns and efficacy beliefs about implementing a mathematics curriculum reform: integrating two lines of inquiry	Educational Studies in Mathematics, 75(1), 1-21
BV+P	Chen, J. J., Brown, G. T. L., Hattie, J. A. C. and Millward, P.	2012	Teachers' conceptions of excellent teaching and its relationships to self-reported teaching practices	Teaching and Teacher Education, 28(7), 936-947
BV+P	Chen, X. and Li, Y. P.	2010	INSTRUCTIONAL COHERENCE IN CHINESE MATHEMATICS CLASSROOM-A CASE STUDY OF LESSONS ON FRACTION DIVISION	International Journal of Science and Mathematics Education, 8(4), 711-735
BV+P	Choppin, J. M.	2009	Curriculum-Context Knowledge: Teacher Learning From Successive Enactments of a Standards-Based Mathematics Curriculum	Curriculum Inquiry, 39(2), 287-320
BV+P	Cobb, P. and Jackson, K.	2013	Lessons for Mathematics Education From the Practices of African American Mathematics Teachers	Teachers College Record, 115(2), 14
BV+P	Drijvers, P., Doorman, M., Boon, P., Reed, H. and Gravemeijer, K.	2010	The teacher and the tool: instrumental orchestrations in the technology-rich mathematics classroom	Educational Studies in Mathematics, 75(2), 213-234
BV+P	Fulmer, S. M. and Turner, J. C.	2014	THE PERCEPTION AND IMPLEMENTATION OF CHALLENGING INSTRUCTION BY MIDDLE SCHOOL TEACHERS Overcoming Pressures from Students	Elementary School Journal, 114(3), 303-326
BV+P	Gur, G., Cakiroglu, J. and Aydin, Y. C.	2012	Investigating Predictors of Sense of Efficacy Beliefs of Classroom, Science, and Mathematics Teachers	Egitim Ve Bilim-Education and Science, 37(166), 68-76

BV+P and Char	Hansson, A.	2010	Instructional responsibility in mathematics education: modelling classroom teaching using Swedish data	Educational Studies in Mathematics, 75(2), 171-189
BV+P	Hardman, F., Abd-Kadir, J., Agg, C., Migwi, J., Ndambuku, J. and Smith, F.	2009	Changing pedagogical practice in Kenyan primary schools: the impact of school-based training	Comparative Education, 45(1), 65-86
BV+P	Heck, R. H. and Moriyama, K.	2010	Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: a regression discontinuity approach	School Effectiveness and School Improvement, 21(4), 377-408
BV+P	Jang, S. J. and Tsai, M. F.	2012	Reasons for using or not using interactive whiteboards: Perspectives of Taiwanese elementary mathematics and science teachers	Australasian Journal of Educational Technology, 28(8), 1451-1465
BV+P	Lampert, M., Boerst, T. and Graziani, F.	2011	Organizational Resources in the Service of School-Wide Ambitious Teaching Practice	Teachers College Record, 113(7), 1361-1400
BV+P	Land, T. J. and Drake, C.	2014	Enhancing and Enacting Curricular Progressions in Elementary Mathematics	Mathematical Thinking and Learning, 16(2), 109-134
BV+P	Louw, J., Muller, J. and Tredoux, C.	2008	Time-on-task, technology and mathematics achievement	Evaluation and Program Planning, 31(1), 41-50
BV+P	Mayrowetz, D.	2009	Instructional Practice in the Context of Converging Policies Teaching Mathematics in Inclusive Elementary Classrooms in the Standards Reform Era	Educational Policy, 23(4), 554-588
BV+P	McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D. and Senk, S. L.	2012	Knowledge of Algebra for Teaching: A Framework of Knowledge and Practices	Journal for Research in Mathematics Education, 43(5), 584-615
BV+P	Meaney, T., Trinick, T. and Fairhall, U.	2013	One Size Does NOT Fit All: Achieving Equity in Maori Mathematics Classrooms	Journal for Research in Mathematics Education, 44(1), 235-263

BV+P	Olteanu, C. and Olteanu, L.	2012	IMPROVEMENT OF EFFECTIVE COMMUNICATION-THE CASE OF SUBTRACTION	International Journal of Science and Mathematics Education, 10(4), 803-826
BV+P	Parise, L. M. and Spillane, J. P.	2010	Teacher Learning and Instructional Change: How Formal and On-the-Job Learning Opportunities Predict Change in Elementary School Teachers' Practice	Elementary School Journal, 110(3), 323-346
BV+P	Retelsdorf, J. and Gunther, C.	2011	Achievement goals for teaching and teachers' reference norms: Relations with instructional practices	Teaching and Teacher Education, 27(7), 1111-1119
BV+P	Rohrer, D. and Pashler, H.	2010	Recent Research on Human Learning Challenges Conventional Instructional Strategies	Educational Researcher, 39(5), 406-412
BV+P	Ruthven, K., Deaney, R. and Hennessy, S.	2009	Using graphing software to teach about algebraic forms: a study of technology-supported practice in secondary-school mathematics	Educational Studies in Mathematics, 71(3), 279-297
BV+P	Ruthven, K., Hennessy, S. and Deaney, R.	2008	Constructions of dynamic geometry: A study of the interpretative flexibility of educational software in classroom practice	Computers & Education, 51(1), 297-317
BV+P	Stein, M. K. and Kaufman, J. H.	2010	Selecting and Supporting the Use of Mathematics Curricula at Scale	American Educational Research Journal, 47(3), 663-693
BV+P	Tran, N. A., Schneider, S., Duran, L., Conley, A., Richland, L., Burchinal, M., Rutherford, T., Kibrick, M., Osborne, K., Coulson, A., Antenore, F., Daniels, A. and Martinez, M. E.	2012	The effects of mathematics instruction using spatial temporal cognition on teacher efficacy and instructional practices	Computers in Human Behavior, 28(2), 340-349
BV+P	Valli, L., Croninger, R. G. and Buese, D.	2012	Studying High-Quality Teaching in a Highly Charged Policy Environment	Teachers College Record, 114(4), 33

BV+P	Zazkis, R. and Leikin, R.	2010	Advanced Mathematical Knowledge in Teaching Practice: Perceptions of Secondary Mathematics Teachers	Mathematical Thinking and Learning, 12(4), 263-281

Characterization

Char	Beauchamp, G., Kennewell, S., Tanner, H. and Jones, S.	2010	Interactive whiteboards and all that jazz: the contribution of musical metaphors to the analysis of classroom activity with interactive technologies	Technology Pedagogy and Education, 19(2), 143-157
Char	Bryant, B. R., Bryant, D. P., Kethley, C., Kim, S. A., Pool, C. and Seo, Y. J.	2008	Preventing mathematics difficulties in the primary grades: The critical features of instruction in textbooks as part of the equation	Learning Disability Quarterly, 31(1), 21-35
Char	Corey, D. L., Peterson, B. E., Lewis, B. M. and Bukarau, J.	2010	Are There Any Places That Students Use Their Heads? Principles of High-Quality Japanese Mathematics Instruction	Journal for Research in Mathematics Education, 41(5), 438-478
Char	Depaepe, F., De Corte, E. and Verschaffel, L.	2010	Teachers' approaches towards word-problem-solving: Elaborating or restricting the problem context	Teaching and Teacher Education, 26(2), 152-160
Char	Doabler, C. T. and Fien, H.	2013	Explicit Mathematics Instruction: What Teachers Can Do for Teaching Students With Mathematics Difficulties	Intervention in School and Clinic, 48(5), 276-285
Char	Even, R. and Kvatinsky, T.	2009	APPROACHES TO TEACHING MATHEMATICS IN LOWER-ACHIEVING CLASSES	International Journal of Science and Mathematics Education, 7(5), 957-985
Char or/and TA+Prod	Foegen, A.	2008	Algebra progress monitoring and interventions for students with learning disabilities	Learning Disability Quarterly, 31(2), 65-78

Char	Fuson, K. C.	2009	Avoiding misinterpretations of Piaget and Vygotsky: Mathematical teaching without learning, learning without teaching, or helpful learning-path teaching?	Cognitive Development, 24(4), 343-361
Char	Grant, L. W., Stronge, J. H. and Xu, X. X.	2013	A cross-cultural comparative study of teacher effectiveness: Analyses of award-winning teachers in the United States and China	Educational Assessment Evaluation and Accountability, 25(3), 251-276
Char	Gregson, S. A.	2013	Negotiating Social Justice Teaching: One Full-Time Teacher's Practice Viewed From the Trenches	Journal for Research in Mathematics Education, 44(1), 164-198
Char	Hand, V.	2012	Seeing culture and power in mathematical learning: toward a model of equitable instruction	Educational Studies in Mathematics, 80(1-2), 233-247
Char	Hennessey, M. N., Higley, K. and Chesnut, S. R.	2012	Persuasive Pedagogy: A New Paradigm for Mathematics Education	Educational Psychology Review, 24(2), 187-204
Char	Herbel-Eisenmann, B. A. and Otten, S.	2011	Mapping Mathematics in Classroom Discourse	Journal for Research in Mathematics Education, 42(5), 451-485
Char	Hobbs, L. and Davis, R.	2013	Narrative Pedagogies in Science, Mathematics and Technology	Research in Science Education, 43(3), 1289-1305
Char	Jacobs, V. R., Lamb, L. L. C. and Philipp, R. A.	2010	Professional Noticing of Children's Mathematical Thinking	Journal for Research in Mathematics Education, 41(2), 169-202
Char	Kaufmann, L.	2008	Dyscalculia: neuroscience and education	Educational Research, 50(2), 163-175
Char	Lee, S. J., Brown, R. E. and Orrill, C. H.	2011	Mathematics Teachers' Reasoning About Fractions and Decimals Using Drawn Representations	Mathematical Thinking and Learning, 13(3), 198-220
Char	Lee, K. and Ng, S. F.	2011	Neuroscience and the Teaching of Mathematics	Educational Philosophy and Theory, 43(1), 81-86
Char	Lembke, E. S., Hampton, D. and Beyers, S. J.	2012	Response to intervention in mathematics: Critical elements	Psychology in the Schools, 49(3), 257-272

Char	Lin, Y. H., Wilson, M. and Cheng, C. L.	2013	An investigation of the nature of the influences of item stem and option representation on student responses to a mathematics test	European Journal of Psychology of Education, 28(4), 1141-1161
Char	Macknight, V.	2011	Maths Teaching Methods: Relational Abstracting as a Hidden Metaphysics	Science as Culture, 20(4), 455-470
Char	Marshall, J. C., Smart, J. and Horton, R. M.	2010	THE DESIGN AND VALIDATION OF EQUIP: AN INSTRUMENT TO ASSESS INQUIRY-BASED INSTRUCTION	International Journal of Science and Mathematics Education, 8(2), 299-321
Char	Niess, M. L.	2013	CENTRAL COMPONENT DESCRIPTORS FOR LEVELS OF TECHNOLOGICAL PEDAGOGICAL CONTENT KNOWLEDGE	Journal of Educational Computing Research, 48(2), 173-198
Char	Rosales, J., Vicente, S., Chamoso, J. M., Munez, D. and Orrantia, J.	2012	Teacher-student interaction in joint word-problem-solving. The role of situational and mathematical knowledge in mainstream classrooms	Teaching and Teacher Education, 28(8), 1185-1195
Char	Schneider, M. and Stern, E.	2009	The Inverse Relation of Addition and Subtraction: A Knowledge Integration Perspective	Mathematical Thinking and Learning, 11(1-2), 92-101
Char	Sherin, M. G. and Drake, C.	2009	Curriculum strategy framework: investigating patterns in teachers' use of a reform-based elementary mathematics curriculum	Journal of Curriculum Studies, 41(4), 467-500
Char	Shirley, M. L., Irving, K. E., Sanalan, V. A., Pape, S. J. and Owens, D. T.	2011	THE PRACTICALITY OF IMPLEMENTING CONNECTED CLASSROOM TECHNOLOGY IN SECONDARY MATHEMATICS AND SCIENCE CLASSROOMS	International Journal of Science and Mathematics Education, 9(2), 459-481
Char	Singer, F. M. and Moscovic, H.	2008	Teaching and learning cycles in a constructivist approach to instruction	Teaching and Teacher Education, 24(6), 1613-1634

Char	So, W. W. M.	2013	CONNECTING MATHEMATICS IN PRIMARY SCIENCE INQUIRY PROJECTS	International Journal of Science and Mathematics Education, 11(2), 385-406
Char	Wagner, D. and Davis, B.	2010	Feeling number: grounding number sense in a sense of quantity	Educational Studies in Mathematics, 74(1), 39-51
Char	Walkowiak, T. A., Berry, R. Q., Meyer, J. P., Rimm-Kaufman, S. E. and Ottmar, E. R.	2014	Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability	Educational Studies in Mathematics, 85(1), 109-128
Char	Walshaw, M.	2013	Explorations Into Pedagogy Within Mathematics Classrooms: Insights From Contemporary Inquiries	Curriculum Inquiry, 43(1), 71-94
Char	Williams, R. T., Swanlund, A., Miller, S., Konstantopoulos, S., Eno, J., van der Ploeg, A. and Meyers, C.	2014	Measuring Instructional Differentiation in a Large-Scale Experiment	Educational and Psychological Measurement, 74(2), 263-279

APPENDIX C – CODING OF FULL-TEXT ARTICLES

Code: IS+P

Specify the reference
Akkus, R. (2013). Following Students' Ideas: How Much to Let Go? *Egitim ve Bilim. Education and Science, 38*, 96-108.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The object of study is the teacher's role in creating argument-based class discussions. The purpose of the study is to shed light on a classroom environment in which a mathematics teacher incorporated an argument-based approach. The research questions of the study are: What happens when a teacher follows his students' alternative ideas in his mathematics classroom? What is the limit of letting go in a problem-solving process?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
A constructivist view of learning is taken, with students' mathematical ideas as the base for a challenging and supportive classroom environment. The teacher's role is to identify potential mathematics and to create an environment based on students' ideas where students can construct their own understandings. Dialogical interaction and negotiation of mathematical ideas and solutions gives students ownership of problem-solving.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
One ninth-grade teacher was selected for the study. He participated in a year-long professional development. First, the teachers experienced argument-based lessons as students and discussed these. Then, the teachers prepared lessons based on argumentation and taught the lessons. Feedback on the lessons was given by the author. George, the selected teacher, was videotaped three times. One transcribed lesson midway through the project was selected for detailed analysis. This lesson was best representative of his struggles when changing his teaching. The author talked to George before and after the lesson.

The quality of classroom discourse and George's struggles were the two focal points for the lesson analysis. An observation matrix adapted from Gunel (2006) was used for analysis. The matrix captures three areas of pedagogical practice: creating dialogical interaction, controlling knowledge and the problem-solving process, and unit preparation and making connections. For each pedagogical area, segments of the lesson were identified. In relation to creating dialogical interaction, four types of questions were identified. According to the criteria matrix, the teacher is placed at one of four levels (exploring, developing, transitioning, or practicing).

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results showed that this teacher had difficulties in letting students follow their own problem-solving process. Due to the nature of his questions (mostly short answer, yes-no, or rhetorical questions), the interaction was only between the teacher and one student at a time. Very few questions required the students to give richer answers to their reasoning and the teacher often interrupted students when they tried to provide their reasoning. The teacher rarely asked other students what they thought about their peer's ideas. The students were hereby given limited opportunities to express and negotiate their ideas. Since the teacher seemed to be the only authority of knowledge, the students were not given the responsibility for the problem-solving process.
The teacher controlled the conversation and took over the responsibility for the thinking process. He thereby lost several opportunities to both realize and address students' misconceptions. Further, he had difficulties in connecting students' ideas to the big mathematical ideas.

Are the results derived from the data analysis?

Yes.

Which practical implications are put forward? In which ways are they substantiated?

Practical implications are not put forward in the article.

The study's relevance to the review question?

The study is of relevance for the review question of teachers' instructional strategies to establish classroom practices in mathematics.

The appropriateness of its methods in the context of this specific review

Useful for a configurative map.

The quality of the execution of these methods

The criteria and levels are described vaguely in the article.

Code: TA+Prod

Specify the reference
Ayres, P. 2013. Can the isolated-elements strategy be improved by targeting points of high cognitive load for additional practice? Learning and Instruction. 23, 115-124.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
Materials high in element interactivity require considerably more working memory resources and are harder to learn. In order to deal with high cognitive load caused by problem complexity, specific instructional methods are required. Isolating elements is one such strategy, as it reduces element interactivity by initially presenting partial tasks before progressing to whole tasks (Ayres, 2006a; Pollock, Chandler, & Sweller, 2002). Learners initially develop partial schemas, which they thenbuild on to construct full schemas at a later time. The main aim of this study was to investigate whether the isolated-elements strategy can be improved further by providing additional practice within problems where cognitive load is highest.

The mathematical task of simplifying the expression 4(3x-6)-5(7-2x) is central to the study.

In the current study, which used the same bracket-expansion tasks as Ayres (2006a) the second and forth calculations were targeted for extra practice. It was expected that by spending more time practicing the most cognitively demanding components, error rates would decrease due to this extended practice. Therefore it was predicted that a strategy that combines both an isolated-elements approach with extended practice at key locations (referred to as a Targeted-isolated approach) would lead to greater learning than both an Equal-isolated approach where equal practice is spent on all four calculations, and a Full-worked example approach where no elements are isolated.

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
In Cognitive load theory (CLT) it is argued that for successful learning to occur the total demands on working memory (total cognitive load) should not exceed the working memory capacity of the learner. For novice learners, with little prior knowledge, chunked information is not readily available and therefore cognitive capacity is easily challenged.

This implies, for instance, that Ayres claims that "Asking novice learners to solve problems creates extraneous cognitive load, because much of the learner's working memory resources are taken up by solving the problem rather than learning about the key features of the problem (see Kirschner, Sweller, & Clark, 2006). In contrast, studying solutions to problems reduces problem-solving search and frees up more working memory resources for schema acquisition (germane cognitive load)."

Instead of exposing students initially to a complex task, containing many interacting elements, students are given related tasks that contain fewer interacting elements. Such a strategy is consistent with a simple-to complex sequencing approach.

Mayer et al. (2002) showed that superior learning (problem-solving) could be obtained by pre-training on the component model in contrast to learning the component and causal models simultaneously.

However, this approach (sub-goals) is also consistent with intrinsic load reduction. By focussing on subgoals, element interactivity is lowered because only the elements within a subgoal are considered at any one time rather than all of the elements in the task.

The results from both experiments were consistent, regardless of whether procedures or concepts were emphasised, learners who were provided with initial materials containing reduced element interactivity (isolatedelements group) outperformed students who were provided fully interacting elements throughout.

On subsequent tests, which required students to solve fully interacting problems there was an expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). Students with low levels of prior knowledge, benefited from the isolated-elements approach (partial worked examples), whereas students with higher levels of prior knowledge benefited most from fully interacting elements (full-worked examples).

Extended practice is a well-known technique to strengthen automation of procedures (Cooper & Sweller, 1987; Logan & Klapp, 1991; Shiffrin & Schneider, 1977; Van Galen & Reitsma, 2010). In order to master bracket expansion tasks learners must become highly proficient in all aspects of the task,

including individual components that are high in element interactivity. Consequently, extended practice on Calculations 2 and 4 was expected to help automation and calculation proficiency.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
In this study three strategies were compared in learning how to expand bracket-expansion tasks. One strategy, called the Equalisolated strategy, isolated the elements by breaking down the brackets into four separate calculations. Learners were provided an equal amount of practice on each calculation. The second strategy, called the Targeted-isolated strategy, also isolated the elements by breaking down the brackets into four separate calculations. However, learners spent three times the amount of time studying the second and fourth calculations compared with the first and third calculations. Both these isolated-elements strategies ensured that only one calculation at a time was presented. In contrast, the third strategy, The Full-worked example strategy, provided full-worked example where all four calculations were presented each time. To provide adequate instruction, all three strategies were embedded within worked examples (Cooper & Sweller, 1987).

Fifty-four Year 8 (mean age of 13.9 years) students from a Sydney high school participated, and were randomly assigned to a group receiving the full-worked strategy (N = 20), or the equalisolated strategy (N = 18) or the targeted-isolated strategy (N = 16). All students had some prior knowledge of the tasks presented.

The experiment consisted of a prior-knowledge test, an acquisition phase and two test phases.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results found in the study indicated that the Targeted isolated strategy was most effective when compared with the Full-worked example strategy (Hypotheses 1). A significant advantage was found in completing the acquisition problem-solving tasks. Students in the Targeted-isolated group also rated the learning phase easier than those in the Full-worked example group.

In comparing the Targeted-isolated strategy with the Equal isolated strategy (Hypotheses 2), no overall advantage was found for the targeted approach on test measures, the cognitive load measure, or the efficiency measures. However, there was a significant interaction found on the Transfer test. Students with low prior knowledge were disadvantaged by the targeted strategy. In contrast, students with higher levels of prior knowledge were not disadvantaged or advantaged.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
These findings may have important implications for the use of worked examples. Research into worked examples has tended to focus on how they can be moderated to help the transition to expertise. Strategies such as variation (Paas & Van Merriënboer, 1994), completion problems (see Van Merriënboer, 1990; Van Merriënboer & de Croock, 1992), and fading (see Renkl & Atkinson, 2003) have all led to significant learning gains by building upon the standard worked-example model. This present study, as well as other studies into isolatedelements, has shown that full-worked examples may not necessarily be the best starting point for novices.

The study's relevance to the review question?
The study is conducted in a laboratory setting and does not include teachers. The ideas of Cognitive load theory and the results might be useful in the sense that it gives teachers opportunities to think but quite much work is needed in order to operationalize this in classrooms.

The appropriateness of its methods in the context of this specific review
We have to discuss as to whether and in which ways laboratory experiments are of relevance for guiding teachers' actions.

The quality of the execution of these methods
Methodologically solid.

Code: TA+SA

Specify the reference
Baki, A. and Cakiroglu, U. (2010). Learning objects in high school mathematics classrooms: Implementation and evaluation. *Computers & Education, 55*(4), 1459-1469.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The goal of this study is to examine the use of learning objects (LOs) in high school mathematics classrooms. This includes designing a LOR (Learning Object Repository: an environment to manage, store and organize the LOs that are known as) about mathematics curricula, planning to apply LOs in the classroom and evaluating the benefits of LOs. The key research questions can be summarized as;

- How did the students use the LOR?
- How did the students evaluate the use of LOs?
- What were students' attitudes and views towards LOR? How and for what purpose should the teacher use LOs?

LOs are computer-based (web-based) learning labs. LOs can be designed to support exploration and investigation of mathematical ideas and relations, and help conceptual understanding.

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
No theory of teaching or learning.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
In the study, 30 students at 9th grade at a high school in Trabzon practiced with LO based learning environment about the topic through 9th grade mathematics curriculum.

Both quantitative and qualitative data was gathered. Quantitative data included a web-based evaluation Form and Internet records of working time period and the frequency of study with an object during study. The qualitative data was gathered from interviews conducted with six students and the teacher. The interviews were used to gain further information about students' opinions and attitudes of working with LOs. The views and attitudes of the students were analyzed by categorizing them as themes. Data from the interviews were compared with data received from observations made by the researchers.

In order to understand the reasons why teacher preferred LOs and the aims for using them, the views of teacher were obtained at semi-structured interviews within certain time intervals. The qualitative data obtained from the teacher was categorized and presented according to frequency of usage. In the interviews; the students were asked to introduce their experiences with LOs throughout the study as liked and disliked cases. They were asked to refer to LOs as being or not being engaging, enjoyable or interesting. They compared their lessons with LOs to their previous lessons without LOs. They also specified their preferences for using LOs.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The students' views and attitudes about the computer-based learning environment of LOs indicate that the innovations in courses by using LOs have influenced the students in a positive way. Students have made positive comments about the motivational and learning themes. For instance, they expressed that LOs were not boring, that they enjoyed working with them and that they found the LOs interesting and motivating. They also added that they were pleased with their overall performance; they helped them study and learn more and encouraged them to take more responsibility in their learning process. Students liked the LOs, as it was considered to have had enhanced interactivity and visual potentials. Being in a learning environment under the control of themselves with the usage of LOs did increase the students' interest in the course. Games or competitions, which are two important factors in LOs, have enabled to create an interesting learning environment for the students. Overall, students evaluated the LOR as *highly* sufficient in terms of Learning

Value, Added Value and Usability of the Design. The Technology Function was considered sufficient. Negative views expressed were generally about similar examples, such as getting tired and worrying about exams.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
No discussion of practical implication.

The study's relevance to the review question?
High, in terms of its focus on a particular teaching approach, in relation to students' views and attitudes of the teaching approach.

The appropriateness of its methods in the context of this specific review
The methods are appropriate for the study conducted and for our review

The quality of the execution of these methods
The execution of the method lacks in quality. The study lacks in theoretical grounding. The language is unclear throughout the entire text. The analysis of the qualitative data is unclear.

Code: IS+SA

Specify the reference
Berry, R. A. W. Kim, N. 2010. Exploring teacher talk during mathematics instruction in an inclusion classroom. Journal of Educational Research. 101, 363-377.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
In this observational study, our primary purpose was to explore the nature of classroom verbal interaction and the congruence of instructional approaches with standards-based mathematics pedagogy in a first-grade inclusion classroom. The exploration was motivated by our belief that students with disabilities would benefit from mathematics instruction that encourages knowing as well as doing (Cawley, 2002) and that interactive dialogue is crucial for the development.

In the present study, we investigated teacher discourse during daily mathematics lessons led by multiple teachers in an inclusion classroom. We were guided by these questions:

What were the characteristics of teacher talk during mathematics lessons with students with disabilities? What was the extent of congruence of the discourse both with the NCTM Communication Standard and among the teachers of knowing.

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
Sociocultural theories of learning view social interaction as constitutive for learning. In this view, learning is thought to be both individual and social, so that participation in social exchanges is necessary to the internalization of learning and knowledge (Rogoff, 1995; Wells, 2000). It then follows that social discourse functions as a tool for teachers to use in constructing effective teaching strategies and developing active learning roles (Palincsar, 1998).

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
For this study, we selected the examination of *discourse*, defined broadly as teacher talk that may or may not support student learning (Burbules, 1993; Cazden, 1988), as the most useful approach for investigating teacher expectations and communication norms for mathematics lessons in an inclusion classroom where four adults – two credentialed teachers, one prospective teacher, and one paraprofessional – shared teaching responsibilities.

Primary data included four videotaped mathematics lessons. We videotaped an entire unit on money, capturing (a) as many of the whole-class lessons as possible and (b) small pull-out groups that met in the hallway or occasionally in the resource room. The corpus of videotapes included 37 whole- or small-group lessons or portions of lessons taped for a period of two weeks for a total of eight hours of videotaped data.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
We identified six categories of teacher utterances: (a) question/elicit, (b) respond to students' contributions, (c) organize/give instructions, (d) present/explain, (e) evaluate, and (f) sociate.

We did not find differences among the teachers regarding frequency of use for three categories of utterances: organizing, presenting/explaining, and sociating. We found statistically significant differences in frequency of use for three categories: eliciting/questioning, responding, and evaluating. However, no explanatory patterns of teacher use were evident. Overall, mathematics lessons were chiefly recitational.

Content-related teacher talk comprised elicitations of recall-based responses, preceding acknowledgement of those responses. This pattern was similar across the four teachers. They rarely asked students to provide explanations, share ideas, or assist peers.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
Develop complementary instructional strategies. Teachers of students with disabilities need to have confident, knowledgeable, and masterful use of multiple approaches, whether traditional or problem centered. They must be able to skillfully combine the use of different approaches at the appropriate times.

In this view, teachers are advised to use the most effective approach to achieve a specific objective. For example, they should use direct instruction to teach basic mathematics knowledge, such as mathematics facts; use self-instruction (i.e., verbal prompts to mediate cognitive and metacognitive operations) to teach processes (Kroesbergen et al., 2004); and use interactive instruction involving "discussion of strategies (i.e., general metacognitive and domain-specific heuristics), multiple solutions to problems, and a 'debriefing' component" (Woodward & Montague, 2002, p. 97) to teach problem-solving.

In general, the implications seem to be speculations that are rather far from the results of the study.

The study's relevance to the review question?
The study is of high relevance for the review questions of teachers' strategies to establish classroom practices.

The appropriateness of its methods in the context of this specific review
Useful for a configurative map.

The quality of the execution of these methods
Acceptable

Code: LM+P

Specify the reference
Bieda, K. N. (2010). Enacting proof-related tasks in middle school mathematics: Challenges and opportunities. *Journal for Research in Mathematics Education, 41*(4), 351–382.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
What is the nature of the students' and teacher's actions and discourse during the enactment of tasks written to engage middle school students in justifying and proving?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
No background theory in use. The approach is stoff-didactical in terms of being specific to conceptualizations of proofs and proof-activities.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
Qualitative analysis of classrooms observations. Data were collected in seven middle schools. The author constructed field notes guided by an observation protocol. The observations were supplemented with teacher interviews to get the teacher's perspective on how he or she acted.

The tasks were taken from the connected mathematics project (CMP).

The data-analysis followed a four step structure.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results show that students regularly responded to proof-related tasks by offering conjectures that created opportunities to prove, and students provided justifications for nearly half of the conjectures, resulting in proving events. Although the students were regularly involved in responding to proof-related tasks, teacher in the classroom observed did not provide sufficient feedback to sustain Discussions about students' and/or conjectures justifications. Findings also indicate that when a teacher feedback to students' provided justification, it was not sufficient to establish standards for proofs in a mathematics classroom. For instance, teachers were as likely to sanction a justification with a positive appraisal if it was based on a non-proof argument as justification based on a general argument.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
Points to the importance of setting standards for proof-related instructions. It is suggested that teachers need to take an active role during proof-related activities in order to establish the standards and guide the students' proof-abilities.

The study's relevance to the review question?
High.

The appropriateness of its methods in the context of this specific review
The study lacks in theoretical grounding. The method is structured but the many choices mace are not clearly motivated and justified in theory. Field notes are used to collect data from the classroom observations. No video-recordings.

It would be expected to see a deeper elaboration of design principles of tasks, aiming to support proof-related activities.

The quality of the execution of these methods
Acceptable.

Code: BV+SA

Specify the reference
Birgin, O., Baloglu, M., Cathlioglu, H. & Gurbuz, R. (2010). An investigation of mathematics anxiety among sixth through eighth grade students in Turkey. *Learning and Individual Differences, 20*, 654–658.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The purpose of the present study is to investigate mathematics anxiety on a sample of sixth through eighth grade Turkish students in terms of mathematics achievement levels, perceived enjoyment of the mathematics teaching method (PET), perceived enjoyment of mathematics (PEM), and perceived help with mathematics from parents (PPM).

a) What are the mathematics anxiety levels of sixth through eighth grade Turkish students?

b) Are there any significant differences on mathematics anxiety across gender and grade level?

c) Are there any significant relationships between mathematics anxiety and mathematics achievement, PET, PEM, and PPM and can these variables be used in the prediction of mathematics anxiety?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
No theory of teaching or learning.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
The study group consisted of 220 sixth through eighth grade Turkish students. Participants represent culture and socioeconomic status of a suburban community. It is considered that this geographic region of Turkey will reflect the overall picture of the country in terms of mathematics anxiety; due to the fact that, according to the PISA results, the mean mathematics anxiety score of the region is the same as the mean mathematics anxiety score of the entire country.A Personal Information Form and the Mathematics Anxiety Scale for Elementary School Students (MASESS) were used to collect the data in the study. The form included demographic items such as gender and grade level and questions about mathematics achievement level (assessed by the mathematics grade in the previous year), PET, PEM, and PPM.

Three standard multiple regressions were computed between mathematics anxiety as the dependent variable and mathematics achievement, PET, PEM, and PPM as independent variables for three grade levels.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
For all three grade levels, the highest zero-order correlations were between mathematics anxiety and mathematics achievement. Whereas the lowest anxiety was measured among the sixth graders (M = 28.32; SD = 8.24), the highest anxiety was among the eighth graders (M = 33.49; SD = 7.57), which shows that mathematics anxiety levels increase as grades increase.

Gender difference in mathematics anxiety is the single most studied environmental antecedent. In the present study no significant difference between boys and girls was found. However, the inconsistency of the results is still present. This inconsistency might be due to the fact that mathematics anxiety is not a well-defined and measured construct. The present study found stronger and more prevalent grade level differences than gender differences in mathematics anxiety.

For all grade levels, PEM had a significant negative effect on mathematics anxiety. In addition, PEM was the second highest contributor to mathematics anxiety following mathematics achievement for the eighth graders.

The current study found that PET had a significant negative effect on mathematics anxiety on sixth and seventh graders but not on eighth graders. After mathematics achievement, it was the second highest contributor to mathematics anxiety for the sixth graders.

It is concluded that students' own perceptions of mathematics and the teaching method are more prevalent than their parents' help with mathematics.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
Results indicate that the impact of PET on mathematics anxiety seems to diminish with age, whereas the impact of PEM seems to be increasing. Therefore, it can be recommended that mathematics anxiety remediation programs focus more on attitudes towards teaching for younger children and towards mathematics as a whole subject for older children.

Results from this study could be used by researchers to plan further studies of mathematics anxiety and to compare their results with the results of the present study. The results could also be used by educators or counselors to plan remedial programs or to revise current course requirements and teaching strategies. Future experimental studies are needed to determine the causes of mathematics anxiety as most studies on the topic are correlational. In addition, further investigation of mathematics anxiety with younger populations is always encouraged.

The study's relevance to the review question?
Medium. It does not get into any details about the classroom interaction and the role of teaching approach and teacher strategies.

The appropriateness of its methods in the context of this specific review
Not very good. The method is based on self-reported information.

The quality of the execution of these methods
The quality is good. No explicit conceptualization of central concepts such as anxiety, However, the data is collected from previous developed and used instruments for gathering information about the object of study. The statistical analysis is elaborated with good quality.

Code: BV+Prod

Specify the reference
Boonen, T., Van Damme, J., and Onghena, P. (2014). Teacher effects on student achievement in first grade: which aspects matter most? *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 24*(1), 126-152.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
How large are the effects of teacher background qualifications, teacher attitudes and beliefs, and instructional practices on student achievement in mathematics, reading, and spelling in first grade?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
No background theory. Several foreground theories presented, based on a discussion on previous research according teacher background qualifications, teacher attitudes and beliefs, and instructional practices.

The study makes particular use of a framework developed by Palardy and Rumberger (2008). This framework provides for a comprehensive multilevel theoretical investigation of teacher effectiveness, including teacher background, teacher attitudes, and instructional practices.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
The data were collected in the context of the SiBO Project (the Dutch acronym for School Careers in Primary Education). A random stratified sample of 122 primary schools was drawn. In the SiBO project a single cohort of approximately 4,000 students was followed from kindergarten (aged five to six years) until the end of sixth grade (aged 11 to 12 years) and beyond.

In the present study, data were used from 3476 first-grade students (aged 6–7 years) nested within 196 classes and teachers (i.e., all classes were taught by a single teacher) nested within 111 primary schools. The mathematics achievement test consisted of 35 items measuring quantity comparison, seriation, classification, measuring, and counting.

The data were analyzed by means of multilevel regression analysis, which takes the hierarchical structure of the data into account.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results are aggregative.

The results of the present study on teacher effectiveness suggest that teachers have a modest to strong effect on student achievement.

A significant effect of teachers' background was found for student achievement in mathematics. Students with more experienced teachers tended to perform better, whereas students with teachers doing more in-service training tended to perform worse. Of the attitudes and beliefs variables examined, job satisfaction was the only variable significantly associated with math achievement. No variables measuring instructional time or modality were associated with math achievement in first grade. Findings revealed that instructional practices had the greatest effect on student achievement in reading and spelling, but not on student achievement in mathematics.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
No practical implications are presented. The only implications reflected on is for future research.

The study's relevance to the review question?
High.

The appropriateness of its methods in the context of this specific review
The method is well outlined. It is based on a new framework, developed in order to provide for a comprehensive multilevel theoretical investigation of teacher effectiveness, including teacher background, teacher attitudes, and instructional practices. The paper ends with presenting limitations of the method.
The appropriateness of the method is good. The result and argumentation are based on a qualitative analysis, which is grounded in theory.

The quality of the execution of these methods
Good

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

Code: IS+P

Specify the reference
Bozack, A. R., Vega, R., McCaslin, M. and Good, T. L. 2008. Teacher Support of Student Autonomy in Comprehensive School Reform Classrooms. Teachers College Record, 110(11), 2389-2407.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The object of study is autonomy-supportive teacher behaviors. The research questions of the study are: Are autonomy-supportive teaching practices present? And if so, what is the *nature* of the teacher-student interactions in these classrooms?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
Research suggestions from self-determination theory (SDT) on how teachers can foster autonomy in the classroom is used to develop the Autonomy Supportive Behavior Instrument (ASBI) for classroom observations. SDT defines autonomy as the "psychological need to experience one's behavior as emanating from or endorsed by the self rather than being initiated by forces or events that feel alien or with which they do not identify" (Reeve, Deci, & Ryan, 2004, p. 34). SDT rests on the assumption that only when a behavior has been completely internalized will a person perceive his or her behavior as autonomous (Deci & Ryan, 1991). Eight teaching practices are suggested by SDT that can foster autonomy in the classroom:

1. listening carefully,
2. creating opportunity for students to work in their own way,
3. creating opportunity for students to talk,
4. arranging learning material and seating patterns so that students manipulate objects rather than passively watch and listen,
5. offering encouragement when students show effort and persistence,
6. giving hints and praising mastery and progress,
7. replying to student-generated questions in a contingent, satisfying way, and
8. acknowledging students' perspectives.

It is through participation in the classroom – as individuals and as social beings – that students begin to learn their roles within the classroom community (Rogoff, 1994). Classrooms in which teachers model respectful interactions, focus on success for every student, and engage students in help-giving and help-seeking behaviors, can provide the safety net that students need to engage in autonomous, self-regulated behaviors. Student autonomy does not exist in isolation, but is a component of the relationship between self and classroom contexts.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
Data for the study was collected from five elementary schools (Grades 3, 4, and 5) that were a subset of the authors' larger Comprehensive School Reform (CSR) sample. Field notes from 106 classroom observations were analyzed using the Autonomy Supportive Behavior Instrument (ASBI) consisting of eight questions derived from the eight teaching practices suggested by SDT. Four additional items to help identify classroom contexts were included.

As the authors point out, there are limitations to consider when interpreting the results of this study. There were lack of uniform writing of the field notes because the outsider observers were instructed to "note the noteworthy" in the CSR field notes.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results show that all eight characteristics suggested by SDT were present in the field notes, but their prevalence and the nature in which they were evident differed. There was little opportunity for students to choose how they wanted to work with objects. Students had many opportunities to talk, but teachers rarely helped students to relate ideas and concepts between topics or between lessons. Explicit instances of encouragement or teachers engaging the experiences, expertise, or perspective of students were rarely identified. The results are of configurative nature.

Are the results derived from the data analysis?
Yes.

Which practical implications are put forward? In which ways are they substantiated?
The findings suggest that teachers in these settings could extend their autonomy-supportive behavior by scaffolding relationships among concepts and ideas across lessons and subjects. Also, teachers could further extend their autonomy-supportive behaviors by not only responding to, but also elaborating on, the things their students say and by engaging students' own experiences, expertise, and perspectives in the learning process.

The study's relevance to the review question?
The study is of relevance for the review question of teachers' instructional strategies to establish classroom practices in mathematics.

The appropriateness of its methods in the context of this specific review
Useful for a configurative map.

The quality of the execution of these methods
Acceptable

Code: BV+P

Specify the reference
Bray, W. S. (2011). A Collective Case Study of the Influence of Teachers' Beliefs and Knowledge on Error-Handling practices During Class Discussion of Mathematics. *Journal for Research in Mathematics Education, 42*(1), 2-38.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The object of study is teachers' error-handling practices during whole-class mathematical discussions. The aim is to tease out the relative influence of particular types of teacher knowledge and beliefs when student errors arise during class discussion of mathematics. The research questions of the study are: In what ways do teachers respond to student errors in the context of class discussion in mathematics? How do teachers' beliefs and knowledge influence their error-handling practices during class discussion in mathematics?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
A constructivist view of learning is taken, with students' existing knowledge constructions as the base for conceptually supportive, reform-based mathematics teaching. Students' mathematical thinking in solving problems and mathematical discussions are seen as central features of instruction. In these discussions, students' errors are seen particularly useful to stimulate students' thinking.

In the study, an interactive perspective on teachers' knowledge, beliefs, and experiences provides the underlying conceptual framework. An underlying assumption is that teachers' beliefs and knowledge influence, i.e. shape, their actions during mathematics teaching.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
The research design is a collective case study design. A cross-case analysis was made to see patterns across cases and to increase the potential to generalize. Four third-grade teachers at an urban school, in which many students live in poverty and are English language learners, participated in the study. The teachers have varying teaching experience and beliefs. The author (and others) was hired to implement a professional development program during one year at the school with monthly meetings. The program, Everyday mathematics, is reform-based.

Analyses were made of 16 lessons (4 per teacher) with class discussions of students' different solutions to mathematical problems. Pre- and postinterviews were made for every lesson. A teacher knowledge interview (with teaching scenarios) and a web-based beliefs survey (with video-clips) were conducted in the beginning and in the end of the year to measure changes in knowledge and beliefs.

In the analysis, seven teacher beliefs central to reform-oriented mathematics teaching were measured with the help of a rubric to the survey. Four facets of teacher knowledge were explored: knowledge of key mathematical concepts, knowledge of student strategies, knowledge of teaching strategies shown to support development of conceptual understanding, and ability to use mathematical knowledge to interpret student work.

First, the author identified sequences in transcribed discussions when teachers encountered student errors and coded how the teachers responded. Constant comparative method (Glaser & Strauss, 1967) was used to characterize each teacher's typical response patterns to student errors. Then, data from pre- and postinterviews and measures of beliefs and knowledge (in the beginning and end of the year) were used to analyze each teacher's response patterns in relation to her beliefs and knowledge. The author particularly looked for patterns of change in practice that corresponded to patterns of change in beliefs and knowledge. The outcome was case stories giving a theoretical explanation of how each teacher's response patterns were related to her beliefs and knowledge. Also, a cross-case analysis was made of the case stories to identify dimensions of teachers' error-handling practices that captured the teachers' response patterns. Finally, theoretical explanations for how beliefs and knowledge influence a teacher's actions were examined for each dimension.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results on the teachers' beliefs and knowledge show that three of the four teachers (all except Ms. Larsano) made significant shifts toward reform-oriented beliefs during the year. Two of the teachers (Ms. Aria and Ms. Rosena) had strong conceptual knowledge in both beginning and end of the year, but even stronger in the end. The other two teachers (Ms. Jarmin and Ms. Larsano) had weaker knowledge for teaching mathematics in conceptually ways, but their knowledge grew over the year. This is consistent with trends in other data sources, including formal as well as informal classroom observations and monthly meetings.

Two contrasting case story excerpts illustrate teachers' error-handling practices. Ms. Rosena intentionally incorporated erroneous student solutions into the discussions. Her beliefs were that students' errors need to be confronted and that students learn a lot from their own and others' mistakes. Towards the end of the year, students collaborated to understand correct as well as incorrect solutions. Ms. Rosena had students comment on each other's solutions, but she struggled throughout the year to develop routines for greater student-student interaction. Ms Larsano was not intentional about incorporating erroneous student solutions. She often did not notice errors that still came up or did not clear out correct and incorrect aspects of student solutions. She had tight control over the discussion and asked closed questions. Ms. Larsano held beliefs that incorrect solutions can hurt students' feelings and that her students had limited abilities to support each other.

The results on the cross-case analysis reveal three dimensions of error-handling practices across the four teachers. These dimensions are: 1) Intentional focus on flawed solutions in whole-class discussions, 2) Promotion of conceptual understanding through discussion of errors, and 3) Mobilization of a community of learners to address errors.

Are the results derived from the data analysis?
Yes.

Which practical implications are put forward? In which ways are they substantiated?
The findings suggest that teachers would benefit from greater awareness of common student errors and how these errors relate to key mathematics concepts. Teachers need support to envision how student errors can be used as springboards for inquiry. This informs professional development. Reform-based mathematics teaching that builds on student thinking is complex and teachers need time and support to change their beliefs. They also need support to develop knowledge about school mathematics, students' mathematical thinking, and routines to enact such pedagogy.

The study's relevance to the review question?
The study is of high relevance for the review question of teachers' instructional strategies to establish classroom practices in mathematics.

The appropriateness of its methods in the context of this specific review
Useful for a configurative map.

The quality of the execution of these methods
The article is based on the author's doctoral dissertation and has a solid methodological base.

Code: IS+P

Specify the reference
Brodie, K. 2011. Working with learners' mathematical thinking: Towards a language of description for changing pedagogy. Teaching and Teacher Education, 27(1), 174-186

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
How can we describe the hybrid nature of the shifts in teacher practices as they implement their interpretations of reform visions? This paper aims to describe more closely teachers' *hybrid practices* as they take up aspects of reform practice, particularly becoming more responsive to learners' thinking.

Drawing on the literature on classroom talk, this study develops a *coding scheme for teachers' responses to learners' contribution*s, relates the codes back to the literature, and uses the codes to describe the practices of four teachers who deliberately tried to focus their teaching more on learners' thinking. The codes provide for the beginnings of a language of description for changing practices as teachers work with learners' mathematical thinking.

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
Theories of learning, both concerning students and teachers are not explicitly presented. However, the active participation of students in classrooms, student-centered practices, are suggested as a preferred practice. However, the article in nuanced in discussing student-centered practices making explicit the complexities of the teachers' role and moves in establishing such practices.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
I investigate the practices of four South African secondary school teachers as they deliberately tried to shift their practices to focus on learners' thinking through classroom talk. The teachers were enrolled in a post-graduate degree programme and as part of this programme worked together to prepare a set of tasks that would engage learners' mathematical thinking, and planned ways to support learners' talk and thinking in the classroom.

The participants in this study were four secondary school mathematics teachers, teaching for two weeks each, approximately two months apart. The teachers were observed and videotaped in one Grade 10 or 11 classroom each. The main difference between the two weeks was in the first week the teachers taught according to their own teaching plan in their usual way; while for the second week, they jointly prepared a set of tasks that would support mathematical reasoning and deliberately attempted to engage learners' mathematical thinking. The teachers were all enrolled on an in-service post-graduate degree program at a university in Johannesburg, which had a strong component of mathematics content knowledge for teaching.

All the lessons of each teacher in each week were videotaped and the videotapes were transcribed. I did all the coding on the transcripts while watching the videos, so that the context of the utterances as well as non-verbal actions could be taken into account. The codes were developed through an interactive process between theory and data.

Unit of coding is the teachers' turn, or sub-turn, when the teacher made more than one move in a turn.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The paper develops a method for investigating their teaching across the two weeks and how their teaching changed as they began to take account of their learners' thinking.

The shifts were similar among teachers who used the same or similar tasks and did not show clear differences across classrooms of different SES, further adding to evidence that suggests that a teacher with the interest and resources to change teaching in lower SES classrooms can do so.

Are the results derived from the data analysis?
Yes, clearly anchored in both relevant research and data analysis.

Which practical implications are put forward? In which ways are they substantiated? Whom might benefit from the results and the practical implications (students, parents, teachers, mathematics coaches, principals, municipality leaders, politicians,…). Why and how could the results and conclusions be useful?
While a language of description is essentially a research tool, it can also be a tool for teachers and educators. Being able to describe what you are doing is a first step in understanding and improving on it. While the language that I have provided is certainly not the only possibility, I suggest that it can and will give teachers a way in which they can talk about and reflect on their own practices of responding to and interacting with learners' contributions.

The study's relevance to the review question?
The study is of high relevance for the review questions of teachers' strategies to establish classroom practices.

The appropriateness of its methods in the context of this specific review
The study contributes, substantially, to a configuration of subsets of classroom practices and teachers' moves within these practices.

The quality of the execution of these methods
The execution is convincing.

Code: LM+Prod

Specify the reference
David, M. M. and Tomaz, V. S. (2012). The role of visual representations for structuring classroom mathematical activity. *Educational Studies in Mathematics, 80*(3), 413-431.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
In the classroom practice that we (the authors) observed, there emerged some situations that led the teacher to make the drawings the object of his teaching, without intentionally having planned it in advance. We discuss these situations with the following aims:

1. To investigate how drawings of geometrical figures can structure the geometry activity in the classroom;

2. To discuss teaching actions that can facilitate the students' visualization of mathematical objects.

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
The theoretical grounding is made by the framework of cultural–historical Activity Theory and the theory of expansive learning of Engeström.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
The analysis is focused on one activity system of interconnected activities, according to the third generation of Activity Theory. This activity system, Calculation of areas, is formed by a constellation of four activities: (1) Calculation of areas by reconfiguration of geometrical figures; (2) The teacher and momentarily two girls focus on the interpretation of the drawing; (3) The majority of students keep following the previous pattern of calculation; (4) Returning to the calculation of areas with a new procedure.

Observation of 23 lessons throughout one academic year of a 5th grade classroom (25 students, 11-12 years old). Analysing what mathematics is being taught and learned and how it is being taught.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The results are configurative.

The authors conclude that the activity, which capitalized on visualization, created potential learning related to the following mathematical products: new procedures to do the calculation of areas; characterization of geometrical figures; the norms and rules for the use and interpretation of a drawing in mathematics; the representation of a right-angled triangle by a drawing and the notion that not all triangles are right-angled; the notation for a right angle and conservation of areas.

The analysis describes how the drawings structured the mathematical activity in the classroom. At some moments they played the role of an artifact; at others, even if not previously planned by the teacher, they became the focus and the very object of the activity.

Are the results derived from the data analysis?
Yes

Which practical implications are put forward? In which ways are they substantiated?
It is suggested that teachers should explore situations that arise in class, even if the situations were not prepared with this objective, to discuss those rules with the students. Moreover, that s/he become more conscious, on the one hand, of the importance of sometimes focusing the object of his/her teaching activities on those rules; and, on the other hand, of the great variety of possibilities of learning (beyond the rules) that emerge for the students in this process of discussion.

The study's relevance to the review question?
Highly relevant as it deals with mathematical learning and conditions for learning.

The appropriateness of its methods in the context of this specific review
The appropriateness of the method is good. The result and argumentation are based on a qualitative analysis, which is grounded in theory.

The quality of the execution of these methods
Convincing

Code: LM+SA

Specify the reference
Schukajlow, S., Leiss, D., Pekrun, R. Blum, W., Müller, M., & Messner, R. (2011). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics, 79*, 215–237.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The present study was designed to answer the following research questions:

1. Do students' enjoyment, boredom, value, interest and self-efficacy expectations differ according to the type of problem (intra-mathematical problems, "dressed up" word problems, modelling problems)?

2. What are the effects of classroom instruction using modelling problems on students' enjoyment, value, interest and self-efficacy expectations related to these types of problems?

3. Are there any differences in the effects of different teaching methods, including "directive", teacher-centred instruction and "operative-strategic", student-centred instruction involving student

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
No explicit use of a background theory. Use od several foreground theories such as frameworks of emotions, values and interest, Self-efficacy expectations and mathematical problems and modeling.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
Self-reported pre-test and pre-test to measure students' affect about three types of problems: five modelling problems, four "dressed up" word problems and four intra-mathematical problems. For the pre- and post-test assessment, the same problems and scales are used, in order to make it possible to analyse change in student's affective constructs. These items ask respondents to report about their enjoyment, value, interest and self-efficacy expectation related to the problems. During the teaching unit, students were taught using modelling problems about "Pythagoras' theorem" and "linear functions". Uses parametric tests (t tests, analyses of variance (ANOVAs) and MANOVA) to answer our research questions.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
No differences among students' enjoyment, value, interest and self-efficacy towards intra- mathematical problems, word problems and modelling problems.

A comparison of post- and pre-test means shows that the teaching of modelling problems did have positive effects on students' affect with regard to all three types of problems.

The third research question asks if there were any differences between the effects of teacher-cenatred ("directive") versus student-centred ("operative-strategic") teaching. Clearly, the "directive" teaching unit produced weaker positive changes in students' enjoyment and interest than the operative-strategic" form of teaching did.

Are the results derived from the data analysis?
Yes. Statistical data.

Which practical implications are put forward? In which ways are they substantiated?
Teachers cannot assume that it is sufficient to simply select reality-related problems for triggering students' positive emotions and interest.

The study's relevance to the review question?
The study is relevant to the. Review. It deals with task design. However, the focus is not on students' learning of mathematics. The design variable is related to student attributes.

The appropriateness of its methods in the context of this specific review
The appropriateness of the method is good. The result and argumentation are based on statistical methods, connecting to evidence-based research.

The quality of the execution of these methods
The statistics are convincing.

Code: IS+Prod

Specify the reference
Shein, P. (2012). Seeing with two eyes: A teacher's use of gestures in questioning and revoicing to engage English language learners in the repair of mathematical errors. *Journal for Research in Mathematics Education, 43(2)*, 182-222.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
This revelatory case study examines a 5th-grade teacher's orchestration of discourse and interaction to create opportunities for English language learners to participate in the repair of mathematical errors during a unit on finding the area of geometric shapes.

The study addresses the research question *How do a teacher's gestures support his or her practices of questioning and revoicing in creating opportunities for English language learners to participate in mathematical discussions in the face of errors?* Two sub-questions emerged to streamline the investigation in questioning and revoicing: *How does the teacher use gesture in her questioning and notice students' gestures as they respond to the questioning? and How does the teacher use gestures in her revoicing of students' responses?*

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
This research was guided by a social learning theory that characterizes learning as active and interactive participation in communities of practice (Lave & Wenger, 1991; Wenger, 1998). Also, aspects of Embodied cognition are highlighted.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
This research is a *revelatory* case study (Yin, 2009) of a teacher whose gestures played a vital role in promoting the participation of her ELLs in the repair of errors. The site selected for the study was a K–5 school of approximately 1000 students located in South Central Los Angeles, California. There were 25 students in the fifth-grade classroom. Six students were followed closely to obtain an in-depth look at their interactions with the teacher and the mathematics content. Observations and video-recordings of a mathematics class during an entire nine-day unit on finding the area of geometric shapes.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The findings detail how the teacher used gestures in grounding her questioning, revoicing students' strategies, and narrating the meaning of geometric features. Configurative.

Are the results derived from the data analysis?
Yes. Three episodes are presented to illustrate how teacher gestures matters.

Which practical implications are put forward? In which ways are they substantiated?
When gestures are considered as a legitimate communicative modality within a community of practice, students who rely on gesture, especially the ELLs whose proficiency in the English language is underdeveloped, have opportunities to negotiate mathematical meanings with others and foster a positive identity with the subject area in the academic community.

The study's relevance to the review question?
The study is of high relevance for the present review as it explicitly deals with teachers' interactional strategies.

The appropriateness of its methods in the context of this specific review
The result and argumentation are based on an in-depth analysis of one teachers' interaction with her class.

The quality of the execution of these methods
The language is clear and the argumentation is convincing.

Code: Char

Specify the reference
Singer, F. & Moscovic, H. (2008). Teaching and learning cycles in a constructivist approach to instruction. *Teaching and Teacher Education, 24*, 1613-1634.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
This study attempts to analyze and synthesize the knowledge collected in the area of conceptual models used in teaching and learning during inquiry-based projects, and to propose a new frame for organizing the classroom interactions within a constructivist approach.

Argues for yet another new model due to continuously changing learning environment and available resources of a learning environment (such as the Internet).

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
Elaborates on a wide range of Grand theories, e.g. Constructivism and Symbolic interactionism, and conceptual frameworks such as Polya's and Schoenfeld's problem-solving frameworks.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
Mostly a theoretical paper. Two case studies (Grade 9 and Grade 3) are used to illustrate the proposed model. In the Grade 9 case study the topic was mathematics. Later, the working versions of the proposed model were improved during a teacher-training program.

Limited information is provided of data collection and method of analysis.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
The teaching assumption of the study is inquiry-based teaching and the learning focuses on process-oriented competences such as problem-solving ability. Based on thjis assumption the paper develops a model for structuring and analyzing teaching and learning in science and mathematics. The model centers on three axis: Immersion, Structuring and Applying (IMSTRA).

The results are configurative!

Are the results derived from the data analysis?
Yes, to some extent. But the IMSTRA model is very much derived from and grounded in theories and previous models.

Which practical implications are put forward? In which ways are they substantiated?
The Researchers put forward arguments for how the IMSTRA-model can be implemented in teaching practice. However, the most visible application of the model they argue is for curriculum development and for teacher training programs.

The study's relevance to the review question?
The study has relevance for the review as it deals with how to structure (inquiry-based) teaching in order to increase its value for the learning of mathematics (and science).

The appropriateness of its methods in the context of this specific review
The study contributes, substantially, to a configuration of key variables of critical features of teaching cycles.

The quality of the execution of these methods
The argumentation is convincing.

Code: LM+Proc

Specify the reference
Star and Rittle-Johnson (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology 102*, 408–426.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
Comparing and contrasting examples. This experimental study evaluates the benefits of supporting comparison in a classroom context for children learning about computational estimation.

No explicit aim or research question is stated.

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
No learning theory outlined. The only theory elaborated is the theory of mathematics and particularly the concepts or methods of comparisons and computational estimation.

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
Students from two schools participated in the study. School A is a private urban school where 69 fifth-grade students participated (32 girls and 37 boys). School B is a small rural school where 45 fifth graders and 46 sixth graders participated. The research design followed the structure of a pre-test–intervention–post-test design, including a retention test.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
"Semi-aggregated nature". This study contributes to a growing body of research demonstrating that comparing multiple strategies for solving the same problem facilitates learning. The focus here is on estimation, which is both a critically important real-world skill and a mathematical domain that is significantly more complex than equation solving, which has been the target of prior work. Comparison helped students develop a larger repertoire of estimation strategies, improved their ability to select the most appropriate strategies for computing an easy estimate, and increased some students' retention of conceptual knowledge. The current results provide experimental evidence that, when learning how to estimate, it pays to compare.

Are the results derived from the data analysis?
Yes. Three episodes are presented to illustrate how teacher gestures matters.

Which practical implications are put forward? In which ways are they substantiated?
The use of random assignment of students to conditions within their regular classroom context, along with maintenance of a fairly typical classroom environment, provided causal evidence for the benefit of comparing solution strategies while maintaining fairly good external validity. The study provides several suggestions for using comparison effectively in mathematics classrooms. First, teachers must choose problems and solution strategies carefully. The problems should highlight important and meaningful concepts for students to learn and be solvable using multiple strategies. In addition, students may need some familiarity with one of the strategies before comparing two different strategies.

The study's relevance to the review question?
The teacher is not at the front. Focus is on learning material and, particularly, task design and learning activity.

The appropriateness of its methods in the context of this specific review
The study reflects much of the methods of evidence-based research methodology.

The quality of the execution of these methods
The language is clear and the argumentation is convincing.

Code: Char

Specify the reference
Walkowiak, T. A., Berry, R. Q., Meyer, J. P., Rimm-Kaufman, S. E. & Ottmar, E. R. (2014). Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability. *Educational Studies in Mathematics, 85*(1), 109-128.

What is the exact object of study/objects of studies? Hence which phenomena are studied? Which research aims and questions are posted?
The purpose of this study is to introduce a measure of standards-based mathematics teaching practices, the Mathematics Scan (M-Scan), and to examine its validity and score reliability. The object of study is the M-Scan tool as a quantitative tool for measuring teaching practices that are widely agreed upon to be effective in mathematics teaching. The research questions are:

1. Does the content of the M-Scan align with classroom processes that characterize standards-based mathematics teaching practices?
2. Are the dimensions on the M-Scan well defined such that coders' responses are aligned with the dimension descriptors?
3. What are the convergent and discriminant patterns of validity between the M-Scan and two existing measures of classroom instruction, the RTOP and the CLASS?
4. To what extent was our research team able to achieve high score reliability on M-Scan scores, suggesting that the measure holds promise for broader use in mathematics education?

Which theories or underlying assumptions are implicitly or explicitly used to conceptualize these objects and processes into phenomena?
Standards-based teaching practices rest on the NCTM (2000, 2007) principles and standards. The authors point out that there is general agreement internationally about what constitutes effective mathematics teaching and the processes in which students should be engaged. They refer to European, Chinese, and Australian standards besides the US NCTM Standards. Borko et al.'s (2005) organization of teaching practices across eight dimensions forms the theoretical basis for the development of the M-Scan tool. The eight dimensions are:

1. Structure of the lesson
2. Multiple representations
3. Students' use of mathematical tools
4. Cognitive demand
5. Mathematical discourse community
6. Explanation and justification
7. Problem-solving
8. Connections and applications

Method. Which kind of data is used? What is the total data scope used and analyzed? Which kinds of analyses are used to process data? Which other aspects of the research design are worth pointing out?
60 video-taped lesson observations was made in 60 third- and fourth-grade classrooms. The 60 videotapes were coded using the M-Scan as well as two other measures: RTOP and CLASS. The validity of the M-Scan tool was analyzed in three ways: by content review by experts, by response processes of coders and by relating M-Scan to the other two measures. The reliability of M-Scan was also analyzed.

What are the results and conclusions of the study? Are the results of configurative or aggregative nature?
Firstly, results show that the experts agreed that the dimensions of the M-Scan represent components of standards-based mathematics teaching practices (research question 1). Secondly, 87.7% of the coders' rationales were aligned with coding guide descriptors (research question 2). Thirdly, expected correlations were rather high between M-Scan and RTOP, but low between M-Scan and CLASS (research question 3). In sum, it appears that valid inferences can be made from the M-Scan scores about the extent to which standards-based mathematics teaching practices are present in classrooms. Regarding reliability (research question 4), assessing inter-coder reliability indicates that M-Scan shows promise for broader use for large-scale research focused on mathematics teaching practices.

Are the results derived from the data analysis?
Yes.

Which practical implications are put forward? In which ways are they substantiated?
Practical implications are that the M-Scan tool can be a cost-effective protocol to analyze standards-based teaching practices over a large number of entire lessons, since it can offer efficiency in data collection and analysis when using direct observation. M-Scan also provides a framework for understanding the dimensions of standards-based mathematics teaching practices.

The study's relevance to the review question?
The study is about the validity and reliability of the tool M-Scan as a quantitative tool for measuring standards-based teaching practices. It can be helpful for understanding the dimensions of standards-based mathematics teaching practices.

The appropriateness of its methods in the context of this specific review
Useful for a configurative map.

The quality of the execution of these methods
Acceptable.

APPENDIX D. APPLIED PROCEDURE TO CONSTRUCT JOURNAL CATEGORIES

Although the WoS search resulted in 1,100 unique research articles, the total number of research articles listed in Table 3 add up to more than 1,100 (n = 1145). Since a journal can be listed in multiple WoS journal categories, there exists some overlap between certain categories. As we followed the categorization by WoS, this implies that some journals (and the research articles published in these journals) in our sample are listed in multiple categories. For instance, the *American Journal of Economics and Sociology* is listed both in the *Sociology* journal category as well as in the *Economics* journal category.

We allowed for one exception. We did not include journals in the category *Education and educational research* that were listed also in another category.

Our decision not to include research articles in this journal category that also are listed in other journal categories was informed by practical considerations. The WoS journal category *Education and educational research* is by far the largest category in our sample and a proper analysis of the research articles in this category revealed the necessity of a further categorization within this WoS journal category. We decided not to categorize journals in this category if they were already categorized in another WoS category.

We also decided to merge the WoS categories that related to psychology, except for the WoS category *Psychology educational* (n = 109). The merged group *Psychology* included 99 research articles (See Table 7).

Next, we refined the categorization mainly by merging categories with a high overlap between journals. For instance, all journals listed in the WoS journal category *Language linguistics* were also listed in the WoS journal category *Linguistics.* Therefore, we merged these two WoS categories. In this specific case, we also merged these two WoS categories with *Language, reading and literacy*, a category within the WoS category *Education and educational research*, given the similarities between the three categories.

APPENDIX D. APPLIED PROCEDURE TO CONSTRUCT JOURNAL CATEGORIES

Although the WoS search resulted in 1,100 unique research articles, the total number of research articles listed in Table 3 add up to more than 1,100 (n = 1145). Since a journal can be listed in multiple WoS journal categories, there exists some overlap between certain categories. As we followed the categorization by WoS, this implies that some journals (and the research articles published in these journals) in our sample are listed in multiple categories. For instance, the *American Journal of Economics and Sociology* is listed both in the *Sociology* journal category as well as in the *Economics* journal category.

We allowed for one exception. We did not include journals in the category *Education and educational research* that were listed also in another category.

Our decision not to include research articles in this journal category that also are listed in other journal categories was informed by practical considerations. The WoS journal category *Education and educational research* is by far the largest category in our sample and a proper analysis of the research articles in this category revealed the necessity of a further categorization within this WoS journal category. We decided not to categorize journals in this category if they were already categorized in another WoS category.

We also decided to merge the WoS categories that related to psychology, except for the WoS category *Psychology educational* (n = 109). The merged group *Psychology* included 99 research articles (See Table 7).

Next, we refined the categorization mainly by merging categories with a high overlap between journals. For instance, all journals listed in the WoS journal category *Language linguistics* were also listed in the WoS journal category *Linguistics.* Therefore, we merged these two WoS categories. In this specific case, we also merged these two WoS categories with *Language, reading and literacy*, a category within the WoS category *Education and educational research*, given the similarities between the three categories.

APPENDIX E. CODING OF ABSTRACTS OF SELECTED RESEARCH ARTICLES

Tabell 10. Coding of selected research articles for category "Mathematics education journals"

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
Alajmi, A. H. (2012). How do elementary textbooks address fractions? A review of mathematics textbooks in the USA, Japan, and Kuwait. *Educational Studies in Mathematics, 79*(2), 239-261. doi: 10.1007/s10649-011-9342-1	Presentation of fractions in elementary textbooks	Textbook analysis of physical characteristics, lesson structure, nature of math problems	3(?) textbook series	Kuwait, Japan, USA	USA and Kuwaiti textbooks are larger and introduce fractions in the first grade. The Japanese introduce fractions in the third grade; use linear models and connect with measurement. The Harcourt text (USA?) uses concrete material, the Kuwaiti depend on pictorial representation of the area model. All focus on standard algorithms as the main computational methods.	/
Bouck, E. C., Joshi, G. S., & Johnson, L. (2013). Examining calculator use among students with and without disabilities educated with different mathematical curricula. *Educational Studies in Mathematics, 83*(3), 369-385. doi: 10.1007/s10649-012-9461-3	If students with and without disabilities used calculators to solve mathematics assessment problems and whether using calculators improved their performance	Math tests (with items from released state tests) for 6th and 7th grade students following NSF- or traditional curricula. Calculator use was self-reported, unclear how. Statistical analysis of correlations.	/	US	Only time is reported to be a main factor impacting calculator use and students who self-reported using a calculator were more likely to answer questions correctly.	Implications for practice given the controversy over calculator use
Cai, J. F., Ding, M. X., & Wang, T. (2014). How do exemplary Chinese and US mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265-280. doi: 10.1007/s10649-013-9513-3	Teachers' views about the meaning of, and the ways to achieve, instructional coherence.	/	/	US, China	Meaning: US: connections between teaching activities, lessons, or topics. Chinese: the interconnected nature of mathematical knowledge beyond the teaching flow. Ways to achieve: US: managing a complete lesson structure. Chinese: pre-design of teaching sequences, transitional language, questioning based on the study of textbooks and students beforehand, addressing student thinking,	Contribute to our understanding about the meaning of instructional coherence and ways to achieve instructional coherence in different cultural contexts.

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
					dealing with emerging events in order to achieve "real" coherence.	
Cai, J. F., Moyer, J. C., Wang, N., Hwang, S., Nie, B. K., & Garber, T. (2013). Mathematical problem posing as a measure of curricular effect on students' learning. *Educational Studies in Mathematics, 83*(1), 57-69. doi: 10.1007/s10649-012-9429-3	Effect of middle school curriculum on high school learning	Problem posing tasks. High school student performance is analysed by use of a qualitative rubric. Unclear from abstract if/how degree of implementation of curriculum is accounted for.	/	/ (Probably US)	Students who had used a standards-based curriculum in middle school performed equally well or better in high school than students who had used more traditional curricula.	Problem posing as a valid measure of curriculum effect, and qualitative rubrics as useful assessment tools. Instructional and methodological implications are discussed
Charalambous, C. Y., & Philippou, G. N. (2010). Teachers' concerns and efficacy beliefs about implementing a mathematics curriculum reform: integrating two lines of inquiry. *Educational Studies in Mathematics, 75*(1), 1-21. doi: 10.1007/s10649-010-9238-5	Teachers' concerns and efficacy beliefs toward mathematics curriculum reforms.	Partly: Qualitative analysis of teacher logs.	151 elementary mathematics teachers. 53 teacher logs.	/	A model integrating teachers' concern and efficacy beliefs: - concerns of preceding stages inform concerns of succeeding stages; - efficacy beliefs about using the reform affect task and impact concerns and are informed by their self concerns; - efficacy beliefs about employing pre-reform instructional approaches influence all types of teacher concerns	Methodological implications and directions for future studies are discussed
Ding, M. X., & Li, X. B. (2014). Transition from concrete to abstract representations: the distributive property in a Chinese textbook series. *Educational Studies in Mathematics, 87*(1), 103-121. doi: 10.1007/s10649-014-9558-y	How mathematics curriculum may structure representations in ways that facilitate the transition from concrete to abstract .	Textbook analysis. Instances of the distributive property analyzed (1) within worked example, (2) from worked example to practice problems within a topic, (3) across multiple topics over grades.	1 textbook series, 319 instances of the distributive property	China	Four features that facilitate the transition: (1) situates initial learning in a word problem context, (2) sets up abstract representations as an ultimate goal, (3) incorporates problem variations with connections, (4) it engages students in constant sense making of the transition process	Implementations and future research directions are discussed
Ellerton, N. F. (2013). Engaging pre-service middle-school teacher-education students in mathematical problem posing: development of an active learning framework. *Educational Studies in Mathematics, 83*(1), 87-101. doi: 10.1007/s10649-012-	How problem posing can be made an integral part of mathematics teacher education programs	/	/	/	Teacher education students can recognize the need for problem posing both in their own programs and in school mathematics curricula. A framework for interpreting the role of problem posing.	Teacher education programs are a good place to start the process of redesigning mathematics curricula in terms of problem posing.

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
9449-z						
Gresalfi, M., Barnes, J., & Cross, D. (2012). When does an opportunity become an opportunity? Unpacking classroom practice through the lens of ecological psychology. *Educational Studies in Mathematics, 80*(1-2), 249-267. doi: 10.1007/s10649-011-9367-5	How student engagement is an accomplishment of the classroom system; the role of the teacher	/	2 teachers and their students.	/ (Identical curricular materials)	/	Contributes to understanding of how teachers' framing of activity significantly impacts the ways that students are likely to engage tasks. Highlights the interactional nature of learning, with the goal of clarifying why learning is not simply an individual accomplishment.
Hong, D. S., & Choi, K. M. (2014). A comparison of Korean and American secondary school textbooks: the case of quadratic equations. *Educational Studies in Mathematics, 85*(2), 241-263. doi: 10.1007/s10649-013-9512-4	Quadratic equations sections of textbooks	Textbook comparison/ analysis. Number of topics, contents and mathematics items were analyzed.	/	Korean and American secondary textbooks	Korean students learn some topics earlier than American students. American textbooks include more problems requiring explanations, various representations and problems requiring higher level cognitive demand. This indicates that textbooks might not be the reason for American and Korean students' performances in international comparative studies.	/
Shield, M., & Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics, 82*(2), 183-199. doi: 10.1007/s10649-012-9415-9	To assess junior secondary mathematics textbooks' potential to assist in teaching and learning aimed at building and applying deep mathematical knowledge.	Textbook analysis according to specific curriculum goals and associated indicators from the field of proportional reasoning	5 textbook series.	/	The five textbook series provided limited support for the development of multiplicative structures required for proportional reasoning	The study demonstrated a method that could be applied to the analysis of junior secondary mathematics in many parts of the world.
Son, J. W., & Senk, S. L. (2010). How reform curricula in the USA and Korea present multiplication and division of fractions. *Educational Studies in Mathematics, 74*(2), 117-142. doi: 10.1007/s10649-010-9229-6	Cross-national differences in schooling; development of multiplication and division of fractions in curricula	Analyses of content and problems in textbooks	2 curricula	US (Everyday mathematics) & Korea (7th Korean mathematics curriculum)	Multiplication of fractions is developed one semester earlier in KM than in EM, but division at the same time. They devote the same number of lessons to multiplication, but KM devotes five times as many as EM to division. Conceptual understanding is developed before	Insights into cross-national differences in schooling

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
					procedural fluency in EM, whereas in KM they are developed simultaneously. Multistep computational problems are more common in KM than in EM, and the response types are also more varied.	
Stacey, K., & Vincent, J. (2009). Modes of reasoning in explanations in Australian eighth-grade mathematics textbooks. *Educational Studies in Mathematics, 72*(3), 271-288. doi: 10.1007/s10649-009-9193-1	Reasoning in textbooks	Textbook analysis. Classification of explanations according to the mode of reasoning used.	Seven topics in ninetextbooks	Australia, Grade 8	Seven modes of reasoning were found. Most textbooks provided explanations for most topics but the main purpose appeared to be rule derivation or justification in preparation for practice exercises, rather than using explanations as thinking tools. Textbooks generally did not distinguish between the legitimacies of deductive and other modes of reasoning.	/
Sun, X. H. (2011). "Variation problems" and their roles in the topic of fraction division in Chinese mathematics textbook examples. *Educational Studies in Mathematics, 76*(1), 65-85. doi: 10.1007/s10649-010-9263-4	Roles of variation problems as used in textbooks	Textbook comparison;	(Probably) one Chinese and one American textbook	China, America	No information about the results of the textbook analysis. A framework to understand variation practice is introduced.	Educational implications are suggested
Baker, D. K., H.; Collins, J.; Leon, J.; Cummings, E.; Blair, C.; Gamson, D. (2010). One Hundred Years of Elementary School Mathematics in the United States: A Content Analysis and Cognitive Assessment of Textbooks From 1900 to 2000. *Journal for Research in Mathematics Education, 41*(4), 383-423	A content analysis of mathematics textbooks from 1900 to 2000 in elementary school.	Content analysis, document analysis.	141 elementary school mathematics textbooks published between 1900 and 2000	US	The study revealed patterns in content covered by the textbooks over the years.	Discussion of implications in terms of the historical study of mathematics and curriculum in US schools.
Bieda, K. N. (2010). Enacting Proof-Related Tasks in Middle School Mathematics: Challenges and Opportunities. *Journal for Research in Mathematics Education, 41*(4), 351-382.	An analysis of the processes and outcomes of implementing proof-related tasks.	Multiple case study of classrooms engaged in doing proof-related tasks with experienced teachers	7 middle school classes	US	The findings suggest that students' experiences with such tasks are insufficient for developing an understanding of what constitutes valid mathematical justification	/

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
		familiar with the curriculum materials. Document analysis, classroom observations, teacher interviews				
Brown, S. A., Pitvorec, K., Ditto, C., & Kelso, C. R. (2009). (2009). Reconceiving Fidelity of implementation: An Investigation of Elementary Whole-Number Lessons. *Journal for Research in Mathematics Education, 40*(4), 363-395.	The study examines interactions between teachers and curricula as evidenced by their enactments of whole-number lessons from a Standards-based curriculum.	Classroom observations	1st- and 2nd-grade classroom lessons	US	The level of fidelity to the written curriculum differs from the level of fidelity to the authors' intended curriculum during lesson enactments.	An exploration of how curricula support and hinder teachers as they engage students in opportunities to learn mathematics and how teachers' instructional moves and choices impact the enactment of curricula.
Confrey, J. S., M. E.; Battista, M. T.; Smith, M. S.; King, K. D.; Sutton, J. T.; Boerst, T. A.; Reed, J. (2008). Situating research on curricular change. *Journal for Research in Mathematics Education, 39*(2), 102-112.	The purpose of this article is to highlight advances related to curricular research; pose questions that require further investigation; and describe related, emerging subfields	Review	/	US	A discussion of the promise of new methodologies, an examination of evidence on how standards-based curricula have impacted different groups of students, a discussion of how researchers from teacher education have opened new subfields, report of the policy implications of the combination of research on curricular effectiveness in general, on issues of diversity and equity, and on the interaction of teacher knowledge and curricular reform.	/
Gavin, M. K., Casa, T. M., Adelson, J. L., & Firmender, J. M. (2013). The Impact of Challenging Geometry and Measurement Units on the Achievement of Grade 2 Students. *Journal for Research in Mathematics Education, 44*(3), 478-509.	The study reports about the impact of developed geometry and measurement units on students' achievement.	12 week intervention study; Experimental & comparison group, hierarchical linear modeling, mathematics concepts subtest & open-response assessment.	Grade 2 students: 193 in experimental group, 192 in control group.	US	The experimental group exhibited a deeper understanding of geometry and measurement concepts as measured by the open-response assessment while still performing as well on a traditional measure covering all mathematics content.	/
Grouws, D. A., Tarr, J. E., Chavez, O., Sears, R., Soria, V. M., & Taylan, R. D. (2013). Curriculum and Implementation Effects	The study analyzes what curriculum, curriculum implementation and teacher	One group consisted of students studying an integrated textbook; a	2161 students (first year high school) from ten schools in five states.	US	Students who studied from the integrated curriculum were significantly advantaged over students who studied	/

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
on High School Students' Mathematics Learning From Curricula Representing Subject-Specific and Integrated Content Organizations. *Journal for Research in Mathematics Education, 44*(2), 416-463.	characteristics are associated with students' learning.	second group of students studied from a subject-specific textbook; hierarchical linear modeling.			from a subject-specific curriculum; Opportunity to learn and teaching experience were significant moderating factors.	
Harwell, M. R., Post, T. R., Medhanie, A., Dupuis, D. N., & LeBeau, B. (2013). A Multi-Institutional Study of High School Mathematics Curricula and College Mathematics Achievement and Course Taking. *Journal for Research in Mathematics Education, 44*(5), 742-774.	The study examined the relationship between high school mathematics curricula and student achievement and course-taking patterns over four years of college course-taking.	Retrospective study; Three types of curricula were studied: National Science Foundation (NSF) funded curricula, the University of Chicago School Mathematics Project curriculum, and commercially developed curricula.	Over 10,000 students from 32 postsecondary four-year institutions.	US	High school mathematics curricula were unrelated to college mathematics achievement or students' course-taking patterns for students who began college with precalculus (college algebra) or a more difficult course. Students of the NSF-funded curricula were statistically more likely to begin their college mathematics at the developmental level.	/
Huntley, M. A. (2009). Measuring Curriculum Implementation. *Journal for Research in Mathematics Education, 40*(4), 355-362.	The study describes the variety of ways that curriculum implementation is measured.	Review	/	US	There is a need to develop curriculum-sensitive tools for analyzing classroom practice. The study outlines the use of the Concerns-Based Adoption Model (CBAM) theory to develop analytical tools for measuring implementation of middle-grades reform mathematics curricula. The study presents steps for future research.	/
Kramer, S. L., & Keller, R. (2008). An existence proof: Successful joint implementation of the IMP curriculum and a 4x4 block schedule at a suburban US high school. *Journal for Research in Mathematics Education, 39*(1), 2-8.	The article reports results from a study that investigated joint effects of two innovations adopted at a high school: 4 x 4 block scheduling and the Standards-based curriculum (IMP).	/	A high school in an affluent suburban community in the northeast United States	US	By the end of 12th grade, cohorts of students who had studied IMP under a block schedule scored higher than had earlier cohorts of students who had studied a traditional high school mathematics curriculum under a traditional schedule. The article also describes actions taken by the school to build capacity before adopting the reforms.	The results can be seen as an 'existence proof' of what can happen when such reforms are adopted jointly at a site that has put considerable effort into building capacity to implement them well.
Lloyd, G. M. (2008). Curriculum use while	The article describes a	One student teacher using	1 student teacher	US	Several key factors appeared to contribute	Directions for future research

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
learning to teach: One student teacher's appropriation of mathematics curriculum materials. *Journal for Research in Mathematics Education, 39*(1), 63-94.	student teacher's interactions with mathematics curriculum materials during her internship in a kindergarten classroom	two curricula teaching lessons multiple times to different groups of students.			to Anne's particular ways of using the curriculum materials: features of her student-teaching placement, her personal resources and background, and characteristics of the materials based on this student teacher's use	about student teachers' and other teachers' curriculum use are suggested in accord with the outcomes of the study.
Meaney, T., Trinick, T., & Fairhall, U. (2013). One Size Does NOT Fit All: Achieving Equity in Maori Mathematics Classrooms. *Journal for Research in Mathematics Education, 44*(1), 235-263.	The study explores how a school infuses the identity of Indigenous students into the school-based curriculum through the promotion of their language and culture in mathematics lessons.	Application of Bernstein's pedagogic device	1 teacher	New Zealand	For equity to be achieved regarding students' mathematics learning; parents' and the community's aspirations for students' education need to be infused into debates about the knowledge that teachers are expected to include in their teaching. This enables the local context to make a positive contribution to students' learning.	Programs for improvement should not be imposed on schools unless there are opportunities for them to be adapted to the needs of individual schools.
Post, T. R. H., M. R.; Davis, J. D.; Maeda, Y.; Cutler, A.; Andersen, E.; Kahan, J. A.; Norman, K. W. (2008). Standards-based mathematics curricula and middle-grades students' performance on standardized achievement tests. *Journal for Research in Mathematics Education, 39*(2), 184-212.	An analysis of the connection between standards-based curricula and student achievement.	Hierarchical linear modeling	Approximately 1400 middle-grades students who had used a standards-based curriculum for at least three years.	US	Students' achievement levels on the Open Ended and Problem Solving subtests were greater than those on the Procedures subtest. This finding is consistent with several other studies.	/
Post, T. R., Medhanie, A., Harwell, M., Norman, K. W., Dupuis, D. N., Muchlinski, T., . . . Monson, D. (2010). The Impact of Prior Mathematics Achievement on the Relationship Between High School Mathematics Curricula and Postsecondary Mathematics Performance, Course-Taking, and Persistence. *Journal for Research in Mathematics Education, 41*(3), 274-308.	The study examined the impact of prior mathematics achievement on the relationship between high school mathematics curricula and student postsecondary mathematics performance.	Retrospective study; Three types of curricula were studied: National Science Foundation (NSF) funded curricula, the University of Chicago School Mathematics Project curriculum, and commercially developed curricula.	4144 students from 266 high schools	US	High school curriculum was not differentially related to the pattern of mathematics grades that students earned over time or to the difficulty levels of the students' mathematics course-taking patterns. There also was no relationship between high school curricula and the number of college mathematics courses completed.	/
Star, J. R. S., J. P.; Jansen, A. (2008). What students notice as different between reform and traditional mathematics programs.	An analysis of how high school and college students perceived Standards-based	/	93 high school and college students (Grade 9 & university students)	US	Nearly all students reported differences. High-impact differences, like Content, were not always related to	The results show that students' responses to reform programs can be quite diverse and only

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
Journal for Research in Mathematics Education, 39(1), 9-32.	and reform calculus programs as different from traditional ones.				curriculum type. Students' perceptions aligned moderately well with those of reform curriculum authors.	partially aligned with adults' views.
Tarr, J. E. R., R. E.; Reys, B. J.; Chavez, O.; Shih, J.; Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247-280	A study of the impact of middle-grades mathematics curricula and the classroom learning environment on student achievement	Hierarchical linear modelling	2,533 students at ten middle schools adopting NSF funded or publisher-developed textbooks	US	Curriculum type was not a significant predictor of student performance after controlling for student variables, but the standards-based learning environment moderated the effect of curriculum type.	/
Thompson, D. R., Senk, S. L., & Johnson, G. J. (2012). Opportunities to Learn Reasoning and Proof in High School Mathematics Textbooks. *Journal for Research in Mathematics Education, 43*(3), 253-295.	A study about the nature and extent of reasoning and proof in high school mathematics textbooks (topics: exponents, logarithms, and polynomials).	/	20 contemporary high school mathematics textbooks	US	About 50% of the identified properties in the three topic areas were justified, with about 30% of the addressed properties justified with a general argument and about 20% justified with an argument about a specific case. Less than 6% of the exercises in the homework sets involved proof-related reasoning, with developing an argument and investigating a conjecture as the most frequently occurring types of proof-related reasoning.	/
Charalambous, C. Y., Delaney, S., Hsu, H. Y., & Mesa, V. (2010). A Comparative Analysis of the Addition and Subtraction of Fractions in Textbooks from Three Countries. *Mathematical Thinking and Learning, 12*(2), 117-151.	The study relates to a comparison of the treatment of addition and subtraction of fractions in primary mathematics textbooks used in Cyprus, Ireland, and Taiwan.	Document analysis, cross-cultural study	4^{th} grade textbooks in Cyprus and Taiwan, 5^{th} grade textbooks in Ireland	Cyprus, Ireland, Taiwan	Several similarities and differences are found among the textbooks regarding the topics included and their sequencing, the constructs of fractions, the worked examples, the cognitive demands of the tasks, and the types of responses required of students. The researchers suggest the existence of recognizable "textbook signature" per country.	The findings emphasized the need to examine textbooks in order to understand differences in instruction and achievement across countries.
Choppin, J. (2011). The Impact of Professional Noticing on Teachers' Adaptations of Challenging Tasks. *Mathematical Thinking and Learning, 13*(3), 175-197.	This study investigates how teacher attention to student thinking informs adaptations of challenging tasks.	Classroom observations & interviews	Five teachers with at least three years' experience with the curriculum program	US	The level of attention to student thinking influenced development of conjectures of how that thinking developed across instructional sequences which in turn informed teachers' adaptations.	The results suggest that forming communities of inquiry around the use of challenging curriculum materials is important for providing

Reference	Object of study	Method	n	Context	Results/Conclusions	Reported relevance
						opportunities for students to learn with understanding.
Jansen, A., Herbel-Eisenmann, B., & Smith, J. P. (2012). Detecting Students' Experiences of Discontinuities Between Middle School and High School Mathematics Programs: Learning During Boundary Crossing. *Mathematical Thinking and Learning, 14*(4), 285-309.	This study reports on how students experience differences between curriculum programs as thy transition from middle school to high school.	Observational study, interview data	2 students	US	An analytic process for detecting discontinuities between settings from participants' perspectives is presented. Students experiencing a discontinuity concurrently reported a change in attitude. These students' experiences illustrate two opportunities to learn during boundary-crossing experiences: identification and reflection.	/
Otten, S., Gilbertson, N. J., Males, L. M., & Clark, D. L. (2014). The Mathematical Nature of Reasoning-and-Proving Opportunities in Geometry Textbooks. *Mathematical Thinking and Learning, 16*(1), 51-79.	The study characterizes the justifications given in the exposition and the reasoning-and-proving activities expected of students in the exercises.	Document analysis	six secondary-level geometry textbooks	US	The majority of expository mathematical statements were general, whereas reasoning-and-proving exercises tended to involve particular mathematical statements. Although reasoning-and-proving opportunities were relatively numerous, it remained rare for the reasoning-and-proving process itself to be an explicit object of reflection.	Relationships between these findings and the necessity principle of pedagogy are discussed.
Stylianides, G. J. (2009). Reasoning-and-Proving in School Mathematics Textbooks. *Mathematical Thinking and Learning, 11*(4), 258-288.	This study relates to an analysis of the opportunities designed in mathematics textbooks for students to engage in reasoning-and-proving	Document analysis	1 mathematics textbook	US	An analytic/methodological approach for the examination of the opportunities designed in mathematics textbooks for students to engage in reasoning-and-proving is presented and exemplified.	The findings are used as a context to discuss issues of textbook design in the domain of reasoning-and-proving that pertain to any textbook series.

Tabell 11. Coding of selected research articles for category "Curriculum research journals"

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
Choppin, J. M. (2009). Curriculum-Context Knowledge: Teacher Learning From Successive Enactments of a Standards-Based Mathematics Curriculum. *Curriculum Inquiry, 39*(2), 287-320.	The study conceptualizes and documents the formation of curriculum-context knowledge.	Combination of video-stimulated and semi structured interviews to analyze the ways the teachers adapted task over time.	3 teachers in Grades 6 and 8 who are experienced users of a standards-based curriculum program.	US	The teachers made noticeable adaptations over the course of three or four enactments that demonstrated learning. Their learning informed the instructional practices and was intertwined with their discussion of content and how best to teach it.	The results point to the larger need to account for the knowledge necessary to use Standards-based curricula and to relate the development and existence of well-elaborated knowledge components to evaluations of curricula.
Charalambous, C. Y. H., H. C. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship. *Journal of Curriculum Studies, 44*(4), 443-466.	An analysis of the individually and joint contribution of teacher knowledge and curriculum programs to instructional quality (introductory paper to special issue).	/	/	US	/	/
Charalambous, C. Y., Hill, H. C., & Mitchell, R. N. (2012). Two negatives don't always make a positive: Exploring how limitations in teacher knowledge and the curriculum contribute to instructional quality. *Journal of Curriculum Studies, 44*(4), 489-513.	An analysis of the contribution of mathematical knowledge for teaching (MKT) and curriculum programs to the implementation of lessons on integer subtraction.	Videotaped lesson observations, Post-lesson interviews, general interviews, paper & pencil MKT surveys.	Three teachers with differing MKT levels using two editions of the same set of curriculum materials that provided different levels of support.	US	Results suggest that MKT and the support provided by the curriculum programs individually contribute to certain aspects of instructional quality. The findings also suggest a joint contribution, for instance, supportive programs can help low-MKT teachers provide adequate instruction.	/
Hill, H. C., & Charalambous, C. Y. (2012). Teaching (un)Connected Mathematics: Two teachers' enactment of the Pizza problem. *Journal of Curriculum Studies, 44*(4), 467-487.	Documenting the ways (MKT) and curriculum materials appear to contribute to the enactment of a 7th grade Connected Mathematics Project lesson on comparing ratios.	Videotaped lesson observations, Post-lesson interviews, general interviews, paper & pencil MKT surveys.	2 7th grade teachers with widely differing MKT scores	US	Results suggest that MKT and the curriculum contribute to a number of aspects but that the relation between MKT and whether and how students participate in mathematical reasoning is less straightforward. The results also show that this ambitious curriculum poses challenges to the teacher – challenges that only the high-	/

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
					MKT teachers met successfully.	
Hill, H. C., & Charalambous, C. Y. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Lessons learned and open issues. *Journal of Curriculum Studies, 44*(4), 559-576.	An analysis of the unique and joint contribution of MKT and curriculum programs to instructional quality (conclusion paper of special issue).	Cross-case analysis	Four cases, including a total of nine teachers teaching in 6^{th}, 7^{th}, 8^{th} grades	US	Differentiated contribution of MKT and curriculum programs to instructional quality. Development of three tentative hypotheses regarding the joint contribution of MKT and curriculum programs.	Reference to policy implications and issues for further research.
Lewis, J. M., & Blunk, M. L. (2012). Reading between the lines: Teaching linear algebra. *Journal of Curriculum Studies, 44*(4), 515-536.	An analysis of lessons on linear equations from the same curriculum programs taught by two teachers of different levels of MKT.	Videotaped lesson observations, Post-lesson interviews, general interviews, paper & pencil MKT surveys.	Two teachers using the same curriculum materials	US	The results indicate that the mathematical quality of instruction in the two classrooms appears to be a function of differences in MKT. The results also indicate that the curriculum programs might have moderated the discrepancies in quality of instruction.	/
Parks, A. N. (2010). Metaphors of hierarchy in mathematics education discourse: the narrow path. *Journal of Curriculum Studies, 42*(1), 79-97.	Study analyzes language about children in the discourse of mathematics education.	Adoption of a rhetorical perspective to analyze texts in mathematics education (conversations in an elementary classroom, a university mathematics methods classroom, mathematics textbooks, and standards documents)	/	US	The results reveal that all the texts support talking and thinking about children in hierarchical ways.	Implication for teacher educators concerned with equity issues: examining their own language and practices for hierarchical language.
Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies, 41*(4), 467-500.	Introduction of curriculum strategy framework to characterize teacher's interactions with curriculum materials.	/	10 elementary school teachers, using a reform based mathematics curriculum for the first time.	US	Identification of patterns in curriculum strategies.	Reference to implications for further research and curriculum designers.
Sleep, L., & Eskelson, S. L. (2012). MKT and curriculum materials are only part of the story: Insights from a lesson on fractions. *Journal of Curriculum Studies, 44*(4), 537-558.	A study of the contribution of MKT and curriculum programs to instructional	Videotaped lesson observations, Post-lesson interviews, general	2 teachers with differing MKT.	US	In addition to MKT and curriculum programs, teachers' orientations toward mathematics and mathematics	/

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
	quality related to a fractions problem.	interviews, paper & pencil MKT surveys.			teaching contributed to instructional quality. Study argues that aspects of developing orientations and goals are related to MKT.	
Van Steenbrugge, H.; Valcke, M.; Desoete, A. (2013). Teachers' views of mathematics textbook series in Flanders: Does it (not) matter which mathematics textbook series schools choose? *Journal of Curriculum Studies, 45*(3), 322-353.	Teacher's views of the mathematics textbooks used in class.	Survey study conducted with teachers, analysis of test results of students from these teachers, document analysis of curriculum programs	814 elementary school teachers, 1,579 elementary school students, five curriculum programs	Flanders (Belgium)	Teachers' views of mathematics textbooks differed based on the textbook used in class, which is an indication that teacher's view of the curriculum can be considered as an important mediating variable in the process of curriculum enactment.	/

Tabell 12. Coding of selected research articles for category "Teacher education journals"

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
Davis, E. A., Beyer, C., Forbes, C. T., & Stevens, S. (2011). Understanding pedagogical design capacity through teachers' narratives. *Teaching and Teacher Education, 27*(4), 797-810.	The study describes and attempts to understand how teachers adjust the curriculum.	The teachers were asked to write narratives about their use of and changes to reform-oriented science lesson plans.	2 elementary school teachers	US	The study describes on which resources the teachers drew when adapting the curriculum and how that related to particular adaptations.	The paper discusses implications for teacher education, professional development, and educative curriculum materials.
Marz, V., & Kelchtermans, G. (2013). Sense-making and structure in teachers' reception of educational reform. A case study on statistics in the mathematics curriculum. *Teaching and Teacher Education, 29*, 13-24.	The study aims to capture the complexity of implementing educational innovations by focusing on how individual and collective processes of sense-making, the actual structural factors and processes in the school mediate implementation practices.	Exploratory study, Semi-structured interviews	/	Flanders, Belgium	/	/
Stylianides, A. J., & Stylianides, G. J. (2008). Studying the classroom implementation of tasks: High-level mathematical tasks embedded in 'real-life' contexts. *Teaching and Teacher Education, 24*(4), 859-875.	The study proposes an analytic framework for describing and explaining the classroom implementation of different kinds of tasks, and uses the framework to analyze a classroom episode.	/	1 secondary school teacher	/	An analytic framework for describing and explaining the classroom implementation of tasks.	Implications for research are discussed.
Haser, C. (2010). Learning to teach in the national curriculum context. *European Journal of Teacher Education, 33*(3), 293-307.	The study investigates the difficulties beginning middle school mathematics teachers faced in the Turkish national curriculum context.	Participants were interviewed after their first year teaching and after the fourth year or during the fifth year about factors in the national curriculum, school, and classroom contexts that affected their learning to teach.	6 beginning middle school mathematics teachers	Turkey	Findings revealed that the national curriculum context interfered with the school and classroom contexts, often by presenting many additional difficulties. The teacher education period seemed to be ineffective in preparing beginning teachers for the difficulties arising from the national curriculum context.	Teacher education programs in centralized systems should provide pre-service teachers with knowledge of possible difficulties specific to these contexts that they might face in their first years.

Tabell 13. Coding of selected research articles for category "General education"

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
Cobb, P., & Jackson, K. (2008). The Consequences of Experimentalism in Formulating Recommendations for Policy and Practice in Mathematics Education. *Educational Researcher, 37*(9), 573-581.	This is a response to *Foundations for Success: The Final Report of the National Mathematics Advisory Panel (2008)* in which the authors question the limitations and consequences of the Panel's assumption that only experimental research studies can produce scientific evidence.	A discussion of the theoretical underpinnings, potential contributions, and limitations of experimental studies. A focus on three issues central to improving mathematics learning and teaching: equity, nature and content of textbooks, graduate education.	/	US	The authors illustrate the limitations of developing implications for policy and practice by relying exclusively on research conducted using a single methodology.	/
Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using Educative Curriculum Materials to Support the Development of Prospective Teachers' Knowledge. *Educational Researcher, 43*(3), 154-162.	The authors focus on how prospective teachers can learn to read and use educative curriculum materials in ways that support them in acquiring the knowledge needed for teaching.	Review of literature and based on own experiences, two conceptual examples are presented to illustrate how educative curriculum materials might be supportive toward prospective teachers' development of knowledge for teaching.	/	US	Presentation of a set of empirically based design principles.	Implications related to the use of educative curriculum materials in teacher education.
Porter, A., Polikoff, M. S., Barghaus, K. M., & Yang, R. (2013). Constructing Aligned Assessments Using Automated Test Construction. *Educational Researcher, 42*(8), 415-423.	The authors describe an innovative automated test construction algorithm for building aligned achievement tests, based on Surveys of Enacted Curriculum (SEC).	/	/	US	A much more valid tests than result from current test construction practices – target domain could be state content standards, a particular curriculum, or an intervention.	/
Harwell, M., Post, T. R., Cutler, A., Maeda, Y., Anderson, E., Norman, K. W., & Medhanie, A. (2009). The Preparation of Students From National Science Foundation-Funded and Commercially Developed High School Mathematics Curricula for Their First University Mathematics Course. *American Educational Research Journal, 46*(1), 203-231	The study analyzed the impact of participation in either a reform-oriented or traditional curriculum program on difficulty level of the first university	/	/	/	The results provide evidence that reform-oriented curricula do not prepare students to initially enroll in more difficult university mathematics courses as well as traditional curricula, but once enrolled students	Implications relate to selection of high school mathematics curricula and to future research in this area.

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
	mathematics course a student enrolled in and the grade earned in that course.				earn similar grades.	
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, 47*(4), 833-878.	The article reports about three studies designed to evaluate the impact of replacement units (SimCalc approach) targeting student learning of advanced middle school mathematics. The SimCalc approach integrates an interactive representational technology, paper curriculum, and teacher professional development	Each study addressed both replicability of findings and robustness across various settings, with varied teacher characteristics and student characteristics.	/	US	Analyses revealed statistically significant main effects, with student-level effect sizes of .63, .50, and .56.	SimCalc is effective in enabling a wide variety of teachers in a diversity of settings to extend student learning to more advanced mathematics.
Sleep, L. (2012). The Work of Steering Instruction Toward the Mathematical Point: A Decomposition of Teaching Practice. *American Educational Research Journal, 49*(5), 935-970.	The study focuses on what it takes to identify the goals of instruction and use those goals to manage the work.	Data comes from preservice teachers' mathematics lessons.	/	US	The study identified and illustrated seven central tasks of "steering instruction toward the mathematical point": attending to and managing multiple purposes, spending instructional time on mathematical work, spending instructional time on the intended mathematics, making sure students are doing the mathematical work, developing and maintaining a mathematical storyline, opening up and emphasizing key mathematical ideas, and keeping a focus on meaning.	Findings have implications for teacher education, research on teaching, and the design of curriculum materials.
Stein, M. K., & Kaufman, J. H. (2010). Selecting and Supporting the Use of Mathematics Curricula at Scale. *American Educational Research Journal, 47*(3), 663-693.	The study focusses on how teacher capacity (their level of education, experience, and knowledge) and their use of curriculum	The data include interviews and surveys with teachers, as well as observations of instruction, over a two-year period;	48 teachers implementing two standards-based mathematics curricula-Everyday Mathematics and Investigations-	US	Findings indicate that teachers' implementation of Investigations was considerably better than teachers' implementation of Everyday Mathematics in terms of maintaining	/

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
	influence instruction.	qualitative analysis of the curriculum programs.	in two school districts.		high levels of cognitive demand, attention to student thinking, and mathematical reasoning. The implementation measures were not correlated to measures of teacher capacity across school districts but they were correlated with teachers' lesson preparation that took into account the big mathematical ideas within curriculum. Further qualitative analysis indicated that the Investigations curriculum provided more support to teachers for locating and understanding the big mathematical ideas within lessons compared to Everyday Mathematics.	
Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban school districts. *American Journal of Education, 114*(4), 583-626.	The article aims at understanding how districts can create organizational environments that foster teachers' opportunities to learn the new ideas and practices required to carry out ambitious reforms.	Applying communities of practice theory, data relates to a longitudinal study of the implementation of ambitious mathematics curricula in two urban districts	2 school districts	US	The article shows how the district reform effort in one district led to significant opportunities for teacher learning and alignment with reform goals while efforts in the other district coordinated action but failed to spur meaningful opportunities for teacher learning.	Implications related to policy and practice.
Davis, E. A., Palincsar, A. S., Arias, A. M., Bismack, A. S., Marulis, L. M., & Iwashyna, S. K. (2014). Designing Educative Curriculum Materials: A Theoretically and Empirically Driven Process. *Harvard Educational Review, 84*(1), 24-52.	The authors describe a design process in the development of educative curriculum materials that is theoretically and empirically driven.	A design-based research approach.	/	US	A description of a design process for incorporating educative features intended to promote teacher learning into existing, high-quality curriculum materials. The authors argue that testing and refining processes for developing curricular supports for teachers is of paramount importance.	Implications for the design of educative curriculum materials.

Tabell 14. Coding of selected research articles for category "Economics journals"

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review, 34*, 107-121.	The study evaluates curricular effectiveness in elementary mathematics. Importantly, the study allows for curriculum quality to be non-uniform across various mathematics subtopics.	A large panel dataset of schools and districts	/	Florida, US	The authors find evidence of variability in curricular effectiveness across different subtopics within the same curriculum.	The findings suggest that educational administrators should consider the topical performance of their various curricular alternatives when making adoption decisions.
Leme, M. C., Louzano, P., Ponczek, V., & Souza, A. P. (2012). The impact of structured teaching methods on the quality of education in Brazil. *Economics of Education Review, 31*(5), 850-860.	The study estimates the impact of the use of structured methods (structuring of curriculum content, development of teacher and student textbooks, and the training and supervision of teachers and instructors) on the quality of education for students in primary public school in Brazil.	A difference-in-differences estimation strategy.	/	Public schools (4^{th} and 8^{th} grade students) in different municipalities in Sao Paulo	The 4th- and 8th-grade students in the municipalities with structured methods performed better in Portuguese and mathematics than did students in municipalities not exposed to these methods. No finding of differences in passing rates.	/

Tabell 15. Coding of selected research article for category "Sociology journals"

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
Coburn, C. E., Mata, W. S., & Choi, L. (2013). The embeddedness of teachers' social networks: evidence from a study of mathematics reform. *Sociology of Education, 86*(4), 311-342.	As teachers' social networks can be influential in teacher learning and organizational change, the study focuses on why some teachers have networks that are likely to support individual and organizational change, while others have not.	A longitudinal, qualitative study of implementation of a mathematics curriculum in four schools.	4 schools	US	The district policy (1) shaped the tie formation process, influencing the structure of networks; (2) mobilized resources that teachers subsequently accessed via their networks, influencing the benefits accrued through network exchanges; and (3) introduced interaction routines that interrupted the conventional ways teachers talked together.	/

Tabell 16. Table 16. Coding of selected research articles for category "Technological resources journals"

Reference	Object of study	Method	N	Context	Results/Conclusions	Reported relevance
Kim, P., & Olaciregui, C. (2008). The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere. *British Journal of Educational Technology, 39*(4), 700-714.	An electronic portfolio system, designed to serve as a resource-based learning space, is tested in a fifth-grade science class.	Multiple regression modelling	No specific information, probably one 5th grade science class	/	The group of students that worked with the electronic portfolio system did better on an information-processing performance test and on a 3-day delayed memory retention test.	/
Voogt, J. (2010). Teacher factors associated with innovative curriculum goals and pedagogical practices: differences between extensive and non-extensive ICT-using science teachers. *Journal of Computer Assisted Learning, 26*(6), 453-464	Building on previous findings concerning math and science teachers' use of ICT, this study relates to a secondary analysis and explores differences between extensive and non-extensive ICT-using science teachers with respect to pedagogical orientation, ICT competencies and professional engagement.	Secondary analysis, use of t-tests and effect sizes	/	Data is collected in 22 countries	The findings showed that both extensive and non-extensive ICT-using science teachers had a pedagogical orientation that reflected traditionally important as well as lifelong learning curriculum goals and practices, but extensive ICT-using science teachers, much more than their non-extensive ICT-using colleagues pursued curriculum goals and practices that are oriented toward lifelong learning. In addition, extensive ICT-using science teachers appeared more confident about their ICT competencies and felt more professionally engaged.	/

APPENDIX F. DESCRIPTION OF RESEARCH ARTICLES BASED ON A COMPLETE READING

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. Journal for Research in Mathematics Education, 39(3), 247-280.

Journal category: Mathematics education

Criteria

- Criterion 4.1: Describe effectiveness of curriculum materials and provide data on degree of implementation.
- Criterion 4.2: Describe what and how students learn from interacting with particular curriculum materials, rather than describing how much they learn.

Description

Tarr et al. (2008) report on a two-year study investigating links between written, enacted and learned curriculum with a primary focus on the following research questions: (1) Does the implementation of district-adopted textbooks differ by curriculum type? (2) To what extent does curriculum type predict student achievement? (3) Does the relationship between curriculum type and student achievement differ based on variations in how the learning environment aligns to the NCTM *Standards*?

The study uses a mix of qualitative and quantitative data. The data are analyzed statistically with hierarchical linear modeling to take into account the multi-level nature of the data (student, teacher, and school). Participants in the study, all from the US, are either using one of three textbooks funded by the National Science Foundation (NSF) or one of seven publisher-developed textbooks. In total the data sample includes 2,533 students in Grades 6-8, from ten different schools (rural, small community, suburban and urban), and their 33 mathematics teachers.

The methodology includes detailed descriptions of how the researchers have designed measures for degree of textbook implementation and for instructional alignment with the NCTM *Standards*. For the degree of implementation of district-adopted textbooks the authors have developed a *composite implementation index* (in this summary referred to as the CI-index), which is a linear combination of six different variables: (1) relative frequency of textbook use by teacher during instruction, (2) relative frequency of textbook use by students during instruction, (3) relative frequency of textbook use for homework assignment, (4) influence of textbook on lesson content, (5) influence of textbook on lesson presentation, and (6) textbook coverage. The first three are measured as percentages of instructional days as reported by teachers in textbook-use diaries during three 10-days intervals. The fourth and fifth variable are based on observational data from three lessons and measured on a 4-point scale. The sixth variable is the percentage of available textbook lessons taught during the school year, as reported by the teacher. To measure the extent to which the learning environment could be considered as consistent with the NCTM *Standards*, the authors designed a *standards-based learning environment index*, the SBLE-index. Observers indicated on a scale from 1-3 the extent to which lessons (1) provided opportunities for students to make conjectures, (2) fostered development of conceptual understanding, (3) let students explain their responses and solution strategies, (4) encouraged multiple perspectives, and (5) had discussions, aiming at a shared understanding, that built on students' statements about mathematics. Based on observations of three lessons per teacher an SBLE-index in the range 0-10 was obtained for every teacher.

Students' prior achievement was reported by the schools in terms of test scores on standardized tests (performed the spring before the study took place) converted to normal curve equivalents. In the spring of Years 1 and 2 of the study, student achievement was measured by use of the TerraNova Survey (TNS) and the Balanced Assessment in Mathematics (BAM). TNS is a norm-referenced multiple-choice test while BAM is criterion-referenced test with fill-in-the-blank questions and contextualized problems where the student should

explain his/her work. TNS professes to measure performance in all content areas with focus both on skills and concepts. BAM professes to measure reasoning, problem solving and communication, as well as proficiency in skills and procedures.

The results show no significant differences in the CI-index between teachers using NSF-funded and teachers using publisher-developed textbooks, i.e., both groups of teachers utilized their textbooks at approximately the same level. The same holds for the underlying variables with one exception: teachers using NSF-funded textbooks reported a significantly higher frequency of use of textbooks by students during instruction than did teachers using publisher-developed textbooks ($p < .05$). The CI-index was also found not to be a significant predictor of student achievement, neither on TNS nor on BAM.

Teachers using NSF-funded textbooks were more likely to have high SBLE scores, but even among this group of teachers about 30% had low SBLE scores (<3). But when student background and teacher variables were taken into account, neither SBLE nor curriculum type (NSF-funded or publisher-developed) were significant predictors of student achievement. However, the combination of use of NSF-funded textbooks and moderate or high SBLE scores had a significant positive impact on student achievement on the BAM tests ($p < .05$), while the use of NSF-textbooks in combination with low levels of SBLE manifested in lower achievement. Similar differences were not found in students' performance on the TNS-test. The authors have two possible explanations for this: (1) the multi-choice items of TNS might make it too blunt a tool for this kind of measurement and (2) it is known that TNS-results are highly correlated with results on the kind of tests that was used for measuring students' prior knowledge.

The authors point out that the CI-index essentially measures frequency of use of curricular materials and content coverage but that it is not sensitive enough to capture other differences in how the curriculum is enacted and how teachers supplement the curriculum, which might explain why no significant impact on student achievement was found. The authors call for future research aiming at identifying essential factors for textbook implementation and a more accurate metric for detecting student achievement in relation to particular uses of curricular materials.

The fact that positive effects on student learning only were found when NSF-funded textbooks where used in highly standards-based learning environments shows that "consistency between curriculum and instruction is needed in order to actualize student learning" (Tarr et al., 2008, p. 275). The findings that the use of standards-based curricula does not automatically translate into high levels of SBLE add to earlier findings about standards-based curricula being challenging to enact well. It suggests that professional development aimed at yielding a strong implementation of such curricula is needed; otherwise there might even be negative effects on student learning. Research is needed to identify factors that influence successful implementation.

Brown, S. A., Pitvorec, K., Ditto, C., & Kelso, C. R. (2009). Reconceiving Fidelity of implementation: An Investigation of Elementary Whole-Number Lessons. Journal for Research in Mathematics Education, 40(4), 363-395.

Journal category: Mathematics education

Criteria

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.
- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.
- Criterion 4.1: Describe effectiveness of curriculum materials and provide data on degree of implementation.

Description

This study elaborates on the concept of fidelity of curriculum implementation in order to explore how curricula support and hinder teachers as they engage students in opportunities to learn mathematics, and how teachers' instructional moves and choices impact the enactment of curricula. They review how the research field has conceptualized, interpreted and measured fidelity and present a view of fidelity as having two forms. The first is *fidelity to the literal lesson*, by which they mean alignment between an observed lesson and the written instructional materials. The second form is *fidelity to the (curriculum) authors' intended lesson*, i.e., alignment between authors' intended opportunities to learn and the opportunities to learn observed in an enacted lesson.

To investigate these aspects of fidelity the researchers have undertaken a qualitative study of lessons and curricular materials. The data sample consists of 33 videorecorded classroom lessons in Grades 1 and 2, with 11 teachers at five US schools, all using the standards-based *Math Trailblazer* curriculum. The schools varied in terms of demographics and length of use of this curriculum. The observed lessons addressed one of the three content areas addition/subtraction, multiplication, and place value. All parts of the *Math Trailblazer* curriculum program, including textbooks,teacher implementation guide, student guide books and unit resource guides were analyzed.

The enacted literal lessons are described in terms of its lesson steps, i.e., the explicit instructional steps, recommendations and suggestions in the curriculum material. For every observed enacted lesson each lesson step was coded as *implemented*, *partially implemented* or *not implemented*. Based on this coding each lesson was then given an overall rating as having a *low*, *moderate* or *high* level of fidelity (LOF) to the literal lesson. Here it was also taken into account if the implemented/not implemented steps were significant for the whole lesson structure or not.

The curriculum materials were analyzed in order to identify statements about the ways in which students were expected to engage in the activities and content. This led to nine *opportunities to learn* (OTL) codes of two different categories: opportunities to *reason* about mathematics and opportunities to *communicate* about mathematics respectively. The transcribed lessons were segmented, each segment starting and ending with a shift in activity, and each segment was coded with respect to OTL codes. In addition, the extent of engagement in the opportunity to learn was rated as *arouse*, *limited* or *missed*. The information was gathered in a lesson observation protocol and the trends in the ratings of the OTL codes were compared to the researchers' model of the curriculum authors intended lesson. This model consisted of descriptions of students potential forms of engagement with the content, infered by the researchers from the curricular material or by feedback from its authors. The comparison was used to rate every enacted lesson as having a *low*, *moderate* or *high* LOF to the authors' intended lesson.

The results show that the LOF to the literal lesson does not predict the LOF to the authors' intended lesson or vice versa. Looking at individual teachers there is some consitency between lessons, i.e. most of the teachers either had low/moderate or moderate/high LOF to the authors' intended lesson in all observed lessons. This

“suggests consistencies in the ways that teachers’ instructional moves and choices impact their enactments of lessons” (Brown et al., 2009, p. 383).

Looking at individual lessons, two types emerge: one where the LOF to the authors’ intended lesson vary by teacher and one where it (essentially) does not but is low/moderate or moderate/high for all teachers. The first type means that enacted lessons might vary greatly in students’ opportunities to learn even among teachers that closely follow the literal lesson steps. Since there are many statements in the *Math Trailblazers* aiming at developing teachers understanings about conditions necessary for students’ engagements in opportunities to learn, this raises questions about what sense teachers make of those statements. This highlights the need for research on how curricula might support teacher learning. The second type of lessons indicate that for some lessons the LOF to the authors’ intended lesson is tied to the lesson itself. There were examples of lessons where teachers who generally had high LOF to the authors’ intended lesson only had moderate or even low LOF. The researchers hypothise that this might be a result of inconsistenicies between the instructional steps of the particular lesson and the authors’ general intentions. Several examples of such inconsistencies are given in the article, for instance when teachers are instructed to tell students to use a particular representation or strategy even though there are general intentions about letting students select representations and generate strategies on their own. They contend that analyses of teachers’ use of lesson materials are likely to miss crucial aspects like this one and put the burden of implementation solely on the teacher, and they call for “further research on *curricular coherence*” (Brown et al., 2009, p. 389, italics in original).

Cai, J. F.; Ding, M. X.; Wang, T. (2014). How do exemplary Chinese and US mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265-280.

Journal category: Mathematics education

Criterion

- Criterion 2: Describe how teachers prepare for teaching.

Description

Instructional coherence has been identified as an important factor for effective instruction and student learning. Cai et al. (2014) report a study aiming at answering the questions: (a) How do exemplary teachers view the meaning of instructional coherence? and (b) How do exemplary mathematics teachers view possible ways to achieve instructional coherence? To answer these questions 20 Chinese elementary teachers and 16 US middle and high school teachers were chosen, all of which had been awarded for exemplary teaching and also had a specialization in mathematics teaching. Through an open-ended survey the teachers were asked about the meaning of coherent teaching, how to achieve it, and its correlation to student learning.

Both Chinese and US teachers emphasized connections between activities, lessons and/or topics, on a macro level (between lessons) as well as on a micro level (within lessons). However, "Chinese teachers' discussions considered more of the coherent nature of the mathematical knowledge system itself" (Cai et al., 2014, p. 271). Concerning the planning stage the importance of clear objectives and of studying the students and their prior knowledge was highlighted in both groups. However, while the connecting of new knowledge to prior knowledge was referred to by the US teachers as a result of coherent teaching, the Chinese teachers emphasized incorporation of information about students' prior knowledge in the planning of a lesson. Some 55% of the Chinese teachers also stressed the importance of studying the textbooks to understanding the curriculum authors' intentions and how they connected to prior and later pieces of knowledge. This was not at all mentioned by the US teachers.

Most teachers in both groups (87.5-90%) pointed to the importance of a lesson structure, often framed by a BME-model (beginning, middle, end). Some US teachers, but no Chinese ones, explicitly emphasized students' physical engagement. But the Chinese teachers to a higher degree went beyond the BME structure and emphasized the beforehand design of teaching sequences, teaching language, and teacher questions. Nine Chinese teachers (45%) and three US teachers (20%) also expressed that instructional coherence should be achieved through challenging student mathematical thinking. The Chinese teachers distinguished between surface coherence and real coherence: "Surface coherence is indicated by a smooth flow enacting the planned lesson without challenging student thinking, whereas real coherence is marked by a smooth flow that challenges student thinking" (Cai et al., 2014, p. 275). Nine Chinese (45%) teachers, but only one USteacher (6%), argued that in order to achieve coherence, a teacher should deal well with emerging classroom events (e.g., unexpected student responses, questions, and difficulties) and that how a teacher treats this will yield different kinds of coherence.

The majority of US teachers (87.5%) expressed unreserved agreement that a coherent lesson fosters student learning while a majority of the Chinese teachers only partially agreed and expressed that surface coherence need not be helpful.

According to the authors, their findings challenge the field to rethink the essence of instructional coherence and they mean that the Chinese teachers' differentiation between surface and real coherence potentially makes a unique contribution. They also mean that the lack of pre-design and study of textbooks in US teachers' lesson planning is a cause for concern and that "these findings challenge textbook designers to support teachers' lesson planning, teaching, and learning as they strive for instructional coherence" (p. 278). They suggest future research that explores strategies to help teachers improve their pedagogical design capability and flexibility to handle emerging events in the classroom.

When it comes to implications for classroom instruction and teacher professional development the authors point out that “the study of textbooks appears to offer a starting point to improve teachers’ pedagogical design abilities and their coherent knowledge base in order to reach real instructional coherence. It may also be helpful to focus on cultivating teachers’ beliefs in the significance of lesson planning” (Cai et al., 2014, p. 279).

Choppin, J. (2011). The Impact of Professional Noticing on Teachers' Adaptations of Challenging Tasks. Mathematical Thinking and Learning, 13(3), 175-197.

Journal category: Mathematics education

Criteria

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.
- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.

Description:

Choppin (2011) is a qualitative case study, with five US teachers in Grades 6-8, focusing on the impact of teacher noticing on the use of challenging tasks, especially adaptation and enactment of such tasks: (1) What do teachers notice as they observe and reflect on enactments of challenging tasks? (2) How do teachers' adaptations of challenging tasks enhance students' opportunities to engage with tasks at a high demand and provide opportunities for sense making? (3) How do teachers use evidence from past enactments to inform their task adaptations?

All participating teachers used the Connected Mathematics Program (CMP) curriculum and were observered teaching at least six (usally consequitive) classes from at least one instructional unit. All teachers had taught the observed units at least twice before and they all expressed support for the CMP curriculum. The instructional units were videotaped and supplemented with a pre-unit interview as well as a video-stimulated (with material from the observed classes) post-unit interview. The data analysis for this study focused primarly on the teacher interviews. Data was analyzed with respect to three main categories: noticing of student thinking, connecting noticing to adaptation, and adapting challanging tasks. In relation to the first and third category two groups of teachers emerged. In the first group (consisting of three teachers), almost all (73 of 79) references to student thinking were coded as evaluation, typically in reference to students being unable to do something, and without explanations to why students were having difficulties. They adapted tasks in ways that reduced their complexity and their primary reason for adaptation was that in prior enactments of the tasks students had needed extensive support to complete them. In the other group (two teachers) only one reference out of 85 references to student thinking was coded as evaluation. Instead they described specific features of students' solutions and often interpreted them in terms of how students made sense of mathematical ideas or made connections. They adapted tasks in ways that maintained or enhanced their complexity, for instance by facilitating attention to multiple strategies. Their main reason for adaptations was to provide opportunities for students to make sense of key mathematical ideas. (The paper gives several examples, with excerpts from teacher interviews, of how the two groups of teachers evaluated/interpreted student thinking, adapted tasks and justified their adaptations.)

According to the author the most notable outcome of the study was the extent to which the second group of teachers "developed conjectures about the development of students' thinking over an instructional sequence and the ways these conjectures influenced the ways these teachers adapted tasks" (Choppin, 2011, p. 192), which the author describes as local instruction theory-building. Here instructional sequence refers to sequences of tasks related to particular mathematical topics. The author argues that the comprehensive inclusion of challenging tasks in the curriculum materials were crucial for this teacher learning to take place since "these tasks helped teachers make sense of the ways students were thinking about mathematical concepts by eliciting a variety of student responses" (Choppin, 2011, p. 193). At the same time, the study shows that use of challenging curriculum material "does not inherently lead to teacher learning, especially if teachers' selective attention is focused on evaluating the success or failure of their practices rather than inquiring into the details of student thinking" (Choppin, 2011, p. 194), as exemplified by the first group of teachers. The author points to the importance of teacher colaboration in comunities of inquiry for successful implementaion of standards-

based curricula. Futher, it is suggested that teachers need opportunities to enact the same instructional units year after year to be able to generate the interpretations and conjectures about student thinking that support the development of teachers' ability to teach for understanding.

Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban school districts. *American Journal of Education, 114*(4).

Journal category: General education

Criteria

- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.
- Criterion 3.2: Combine inclusion of schools and districts as settings for curriculum implementation.

Description

Stein and Coburn (2008) use social practice theory to understand how districts can create organizational environments (*district learning environments*) for teacher learning required for implementation of new curricula. The purpose is to investigate mechanisms by which connections between communities mediate teachers' opportunities to learn in response to district policy.

Four elementary schools from two urban were chosen for the study. Both districts had adopted new curricula (*Investigations in data, number, and space* and *Everyday mathematics*, respectively) and data for the study were collected during the first two school years of implementation. On the district level data were collected in the form of 13 interviews (45-90 min each) with people on different organizational levels (from deputy chancellor to teacher leaders) and observation of district-led meetings and professional development (24 full days). In addition district documents related to mathematics instruction were collected. On school level six focal teachers from each school were selected, representing different attitudes to the curriculum and the range of school years present in the school. During both years of the study every teacher was interviewed five times and observed in classroom six times. Additional interviews were conducted with the mathematics coaches, the principal and six other teachers at each school. On 3-5 occasions in every school observations were made when teachers interacted in professional development, grade-level meetings, coaching sessions etc. Parts of the data collection were designed to identify the participants' social networks and communities of practise. Other parts of the data collection could then focus on the nature of the interaction in these communities.

Data were analyzed to determine the formally designated roles and structures for supporting mathematics reform and how those actually played out, that is, the interactions that characterized their enactments. This is referred to as the *designated organization*. Of particular interest were the ways in which *boundary objects*, *boundary practices* and *brokers* were used to forge connections between different segments of the district. Using social network analysis, teachers' social networks were identified. These were analyzed in terms of *shared domains*, *regular interaction* and *shared practice*, which are the criteria for a community of practice. These communities are referred to as *lived local communities*.

In Region Z, a new organization, outside the administrative line, was created for the implementation. This organization included one full-time mathematics coach per school, regional instructional specialists (RIS) and a Central mathematics leadership team. Together with the curriculum itself they were the key link between the district and the teachers. The border practices included monthly meetings between the leadership team and RIS and biweekly coach training involving RIS's and coaches (every RIS being responsible for overseeing approximately 60 coaches). Teachers' primary link to the system was through the coach assigned to their school and the professional development sessions that the coach was responsible for. Despite intentions to the contrary, most connections between communities were unidirectional. On all levels professional development focused on how to manage and navigate through the curriculum material. Aside from the curriculum itself, one of the most important boundary objects was a detailed, district-developed, pacing guide. Teachers consistently received the message that they must not skip any lessons and that they need to keep up with the schedule in the pacing guide.

In Greene, as in Region Z, mathematics coaches were the primary link between the district leaders and the teachers. But instead of full-time coaches Greene had two part-time coaches in every school who also taught

part-time. Instead of RISs, the school principals were involved in border practices with coaches and district leaders. In general, there was a richer flora of boundary practices in Greene than in Region Z, and the links between communities being bidirectional. On all levels professional development involved doing mathematics together and discussing the nature of the involved mathematics. Similar to Region Z, a pacing guide was a key boundary object. The curriculum itself was viewed more as a tool to meet grade-level objectives than as the curriculum itself. The pacing guide also placed primacy on the objectives and which curriculum modules to use to achieve them. Coaches were encouraged to use grade-level meetings to focus on ways to use the curriculum to meet the state objectives, without losing sight of the mathematics.

In summary, Region Z had a more complex organization with the extra level of the RISs, the interactions were more unidirectional and exclusive for math experts. Greene provided more of bidirectional interactions and substantially included the administrative line. Both attended to issues of coordination/management, but Greene balanced this with a strong focus on mathematics and student learning.

The article gives several examples of contrasts found in the lived communities in the two regions. In general, the communities of teachers, coaches and school leaders in Region Z were smaller, less likely to span the boundaries of the school and more likely to be incongruent with the reform. Also, "the shared practices in Region Z tended to be focused on superficial discussion and sharing activities, while those in Greene were much more likely to be focused on in-depth discussion of student learning and the nature of mathematics" (Stein & Coburn, 2008, p. 605) .

In Greene the interactions within teacher communities were more likely to stretch over grade levels and to go beyond management of the curriculum material to discussions about instructional strategies, student learning etc. In the lived encounters between teachers and coaches there was also a mix between coordination issues and mathematics teaching and learning, whereas in Region Z those interactions focused on management of the material. In Greene coaches often consulted each other, outside of formal meetings, about instructional matters. No evidence of this was found in Region Z. In both regions principals and teachers seldom interacted around mathematics. Administrators mainly engaged in the reform through interactions with the coaches. However, in Greene these interactions took on issues of how to best carry on the reform while in Region Z it was more about handing over the responsibility to the coaches. Also, principals and coaches in Greene more often acted as brokers and also reached out to the district leadership. The authors write that "in so doing, the Greene coaches and principals can be seen as having the opportunity to shape district direction as opposed to being recipients of district mandates" (Stein & Coburn, 2008, p. 613).

The authors emphasize two structural differences between the districts: That the architecture was bidirectional in Greene and unidirectional in Region Z and that it was inclusive in Greene (with regard to the administrative staff) and exclusive in Region Z. On the other hand, the author's admit that most teacher interactions in both districts were congruent with the districts' guidelines, which could be interpreted as if direction and inclusivity do not matter when it comes to attaining congruence with the district program. However, the authors argue that the analysis of engagement in boundary practices reveals qualitatively different opportunities for teacher learning in the two districts, differences that can be explained in terms of how the districts integrated *reification* (of boundary objects) and *participation* (in boundary practices).

In both districts the key boundary objects were the adopted curriculum and the district-developed pacing guide. The Investigations curriculum (adopted by Greene) provides open-ended activities, information about mathematical ideas at play and helps the teacher to anticipate student responses but without specifying a teaching route to follow. It thus provides ample openings for individuals to negotiate meaning about design of teaching and learning and allows teachers to adjust their practice when confronted with particular scenarios. This would explain why participation in Greene was richer and focused on mathematics concepts and how students learn them.

The Everyday Mathematics curriculum (adopted by Region Z), on the other hand, is divided into daily lessons with fairly narrow objectives and its tasks tend to channel students toward a particular route. There are few details about anticipated student responses but very detailed guidance on how to organize classroom work. Hence, there is little room for learning as an emergent phenomenon. Thus this curriculum provides few openings for RISs, principals, coaches and teachers to negotiate the meaning of the reform for their practices.

This would explain why the participation in and between the communities in Region Z often focused on management and procedure.

The authors conclude by stating that "Good designs […] are those that strike a balance between reification and participation—a balance that allows for negotiation of meaning surrounding features of practice about which participants care and are knowledgeable while still keeping the entire organization moving toward shared, recognizable goals" (Stein & Coburn, 2008, p. 617).

Gavin, M. K., Casa, T. M., Adelson, J. L., & Firmender, J. M. (2013). The Impact of Challenging Geometry and Measurement Units on the Achievement of Grade 2 Students. Journal for Research in Mathematics Education, 44(3), 478-509.

Journal category: Mathematics education

Criteria

- Criterion 4.1: Describe effectiveness of curriculum materials and provide data on degree of implementation.
- Criterion 4.2: Describe what and how students learn from interacting with particular curriculum materials, rather than describing how much they learn.

Description

Gavin et al. (2013) investigate student achievement in Grade 2 when they are taught units on geometry and measurement developed according to principles originally intended for promising students. The overall approach in development of these units was to encourage students to take on the role of a practicing mathematician. Key characteristics of the units included focus on important mathematical ideas, attention to depth of understanding and complexity, differentiation, and communication to promote reasoning and problem solving. Several samples from the curriculum materials are presented in the article together with a description of the development process.

This study is part of a broader one with participants from 11 public schools from urban, suburban, and rural districts in four US states. At each school, teachers were randomly assigned to an experimental or a comparison group. For the Grade 2 experiment reported here there were 12 teachers assigned to each group (with a total of 191 and 189 second-grade students, respectively). Teachers in the experimental group implemented the program for a total of 12 weeks – six on geometry in mid-fall and six on measurement in mid-spring – as a replacement of the regular grade-level curricula. Prior to the implementation, teachers in the experiment group attended a four-day summer institute about the philosophy, strategies and content of the units, and during the school year another two days of professional development. Every teacher was also visited once a week during the implementation by members of the research team for fidelity of implementation checks and for collaboration on lesson planning. If teachers were not following the prescribed teaching strategies the team member modeled the intended approach in the classroom. Teachers in the comparison group received no professional development but a majority of them were visited by professional development staff members.

Student learning outcomes were measured with two different tests. Students in the experimental and comparison groups completed both tests at the beginning as well as at the end of the second school year. The first test was the Math Concepts subscale of the Iowa Tests of Basic Skills (ITBS), a norm-referenced standardized multiple-choice test covering all content areas of second-grade mathematics. The second was an open-response test on geometry and measurement, designed by the researchers to be challenging to second-grade students, and influenced by the NCTM Standards as well as by state, national, and international assessments. The development of this test is described in the article.

Data were analyzed statistically with a 2-level, hierarchical linear model described in the paper. After accounting for state and pre-test scores there was no significant difference on the ITBS post-test between students in the experiment group and the comparison group. However, students in the experiment group outperformed those in the comparison group on the open-response assessment. On average they scored 4.43 points higher (possible range was 0 to 23 points), the effect size being as large as 0.89.

The study does not involve an analysis of which features of the implemented curriculum that are responsible for influencing the outcome but the authors conjecture two such to be salient: complex tasks that require reasoning at high levels and instructional strategies that encourage students to explain their reasoning both verbally and in writing. One possible explanation could also be that more time was spent on geometry and measurement in the experiment group than in the comparison group. If that was the case, than students in the experiment group spent less time on other topics, but still performed as well as the comparison group on the

ITBS-test. Hence their overall understanding was not sacrificed for the greater understanding on geometry and measurement.

The findings of this study are in line with findings on use of gifted students' curriculum for all students in topics like reading and writing.

Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies, 41*(4), 467-500.

Journal category: Curriculum research

Criteria

- Criterion 2: Study describes how teachers prepare for teaching.
- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.

Description

Arguing that greater understanding of the interactions between teachers and curriculum materials is needed, Sherin and Drake propose a curriculum strategy framework (CSF). This framework is developed and used to investigate patterns of curriculum use among teachers and focuses on three interpretive activities: reading, evaluating and adapting the curriculum and does so before, during and after instruction. As such, the CSF is defined by a 3 x 3 matrix capturing all of these interpretive activities at differing moments with regard to instruction. A teacher's curriculum strategy relates to a particular way in which these three interpretive activities are manifested before, during and after instruction.

Ten teachers, teaching mathematics at different elementary schools (Grades 1-3) in the US participated in the study. All these teachers piloted a new reform-based mathematics curriculum. Characterization of a teacher's curriculum strategy was based on an analysis of several sources throughout an entire school year: observations, interviews, curriculum materials. For each lesson, a table that listed evidence of the three interpretive activities was composed. This table served as a basis to abstract a curriculum strategy matrix per teacher.

The study revealed patterns in teachers' curriculum strategy, or in when and how teachers read, evaluate and adjust the curriculum. With regard to reading, one group of teachers read prior to instruction to get a sense of the broad overview of the lesson. A second group read prior to instruction to know more about the details of the lesson. Yet another group read both before and during instruction and did so before instruction to get a broad overview of the lesson while adding this with more detailed information read during the lesson. There was evidence that teachers do not read all information, and that there were differences both in nature of information read by the teachers and when they read this information.

When it comes to evaluating and adapting, the study pointed out that these activities occurred usually at the same point in the instructional process. Each teacher typically had one specific audience in mind and that audience generally changed from one instructional moment to the other. For instance, the study revealed that teachers who evaluated what they were expected to do with the curriculum prior to instruction and then shifted the focus of their evaluation to students during instruction usually adapted the curriculum by creating new activities rather than omitting or replacing activities. Sherin and Drake also found that adaptation prior to instruction related to organizational aspects while only during instruction, adaptations also concerned content related aspects. It is argued that the CSF is valid as the models were consistent across time and data type, and given that the CSF did a good job for all of the teachers.

Suggestions for future research describe that an analysis of how teachers read, evaluate and adapt the curriculum might provide information about practices that constrain or enable instruction; CSF might serve as a tool to come to a better understanding between teacher identity and instructional practices; CSF might serve as a useful tool for examining how teachers use curriculum materials over time – and what they learn from them; expanding the CSF to other subject areas might be worth as well.

Implications for the design of curricula and for professional development are mentioned as well. Considering the former, Sherin and Drake describe that the development of curricula must go hand in hand with research on how teachers use curriculum materials. With regard to the latter, it is mentioned that it might be instructive for teachers to learn about the different approaches to curriculum use.

Hill, H. C., & Charalambous, C. Y. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Lessons learned and open issues. Journal of Curriculum Studies, 44(4), 559-576.

Journal category: Curriculum research

Criteria

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.
- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.

Description

This research article describes the outcomes of a cross-case analysis investigating the unique and joint contribution of teachers' mathematical knowledge for teaching (MKT) and curriculum programs to the mathematical quality of instruction. Nine teachers teaching 6^{th}-, 7^{th}, or 8^{th}-grade students are examined in four case studies, each case study describing how two or three teachers with varying levels of MKT teach the same lessons or lessons on very similar content. All teachers are using the Connected Mathematics project curriculum, a *Standards*-based curriculum program funded by the US National Science Foundation. Data sources include videotaped lessons, post-lesson interviews, general interviews, and paper-and-pencil MKT surveys.

With regard to MKT's unique contribution, the four case studies provided information on how a strong MKT might be beneficial for certain aspects of the mathematical quality of instruction. For instance, the most consistently observed affordance of MKT was related to teacher's use of a dense and precise mathematical language. Mixed evidence, however, was found of the association between teachers' MKT and teachers' eliciting, understanding, and using student productions in instruction. The analysis also suggests that the association between MKT and student participation in meaning-making and reasoning might be weak.

Given the case studies' design to vary MKT and hold the lesson topic and curriculum program constant, more speculative findings are reported about the unique contribution of curriculum programs to the mathematical quality of instruction. The case studies suggest that the supports provided by the curriculum program matter, but also that the *Standards*-based curriculum program also put a specific demand on the teacher that at times resulted in poor lesson implementation. The authors further mention that curriculum programs are ultimately limited in the support that they can provide, as they cannot anticipate everything that will occur during lesson enactment.

Considering the joint contribution of MKT and curriculum programs, the study suggests that supportive curriculum programs can lead to high-quality instruction, even for low-MKT teachers when following the curriculum program closely. It is also suggested that unsupportive curriculum programs might be problematic, in particular for low-MKT teachers. A final suggestion describes that high-MKT teachers might be able to compensate for some limitations in the curriculum program, enabling a high-quality instruction.

The study further stressed the importance of teachers' orientation toward curriculum programs and instruction, in addition to MKT and curriculum materials as important 'contributors' to the mathematical quality of the lesson. The study also revealed that the relationship between these contributors was more complex than straightforward for middle-rank-MKT teachers – the biggest group of teachers.

Finally, the study describes areas that merit further research. For instance, the authors point at the need for further research that analyzes how variations in the support provided by curriculum programs matter when teachers with varying levels of MKT teach the same topic; the need for more studies that examine how teachers use curriculum programs to plan instruction; and the need for studies that vary the educational and cultural settings. Other resulting avenues for further research relate to a wider study of how low-MKT teacher implement demanding curriculum programs, a study of how teachers might be informed by the use of educative

curriculum programs, and whether and how teachers' orientations evolve through use of the curriculum program.

Davis, E. A., Beyer, C., Forbes, C. T., & Stevens, S. (2011). Understanding pedagogical design capacity through teachers' narratives. Teaching and Teacher Education, 27(4), 797-810.

Journal category: Teacher education

Criteria

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.
- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.

Description

Acknowledging that teachers need to adapt – even high quality – curriculum materials to better support the students in class, the study assumes that teachers themselves are curriculum designers. Such teachers' engagement to adapt the curriculum materials is considered as an aspect of teachers' pedagogical design capacity (PDC) – a notion described by Brown (2009) – which relates to teachers' ability to draw on their personal resources as well as on the resources in the curriculum materials to fruitfully mobilize these materials. The authors describe the importance of understanding how teachers adapt curriculum materials and how teachers can be supported in adapting curriculum materials to suit their students' needs. As teaching is so grounded in practice, the authors argue that such supports within specific lessons through teacher accounts about the lesson may be especially useful. Such accounts could take the form of narratives (or vignettes), in which a teacher describes what adaptations to the curriculum were made, and why these adaptations were considered useful.

Thus, to gain insight in teachers' PDC, the authors engaged two elementary teachers in writing narratives describing how they used and adapted curriculum materials – with an explicit intent to incorporate these narratives into the educative curriculum materials that were at focus in the study. Two concrete research questions were posed, (1) What kinds of curricular adaptations do these two case study teachers describe? and (2) On what knowledge, experiences, and resources do they draw in making decisions about curricular adaptations?

Two teachers – Maggie with six years of varying teaching experience and Catie with four years of teaching (not so varying) experience – participated in a larger longitudinal study following the experiences of beginning and, over time, more experienced teachers. The technology-mediated learning environment (CASES) with its educative curriculum materials was developed by the researchers with the aim of supporting preservice and new elementary teachers' learning with regard to inquiry-oriented science teaching. Through narratives, the curriculum materials included educative support such as guidance concerning student ideas and explanations of the rationales for inquiry-based teaching practices. The two teachers were engaged in writing narratives through phone interviews and online discussions built into CASES. By means of prompts, the teachers were stimulated to describe themselves as a curriculum user. This served as a central theme that guided them in the adaptations they made for specific lessons. These concrete adaptations were the focus of their self-written narratives.

The whole data pool consisted of four interviews with each teacher, written responses to the questions posed online, a self-generated profile describing themselves as teachers, drafts of their narratives, comments on and response to one another's narratives, a journal entry written at the end of the process responding to questions about the theme the teachers selected, and email exchanged during the process. Two coding schemes, one capturing the adaptations made by the teachers and another to capture the basis of these adaptations – teachers' knowledge and beliefs, teachers' experiences, the resources teachers draw on – were developed.

The results revealed that Maggie's adaptations reflected a sophistical interplay between intentions of the curriculum developers and her understanding of the students. She based her curricular adaptations on her reliance of knowledge of her students, the learning goals that she aimed for aligned with the curriculum's inquiry-oriented focus, her rich set of teaching experiences, a wealth of additional resources within as well as

outside of CASES. Unlike Maggie's adaptations, Catie's adaptations were often not in line with the curriculum's intentions. Catie related her adaptations mainly to her learning goals which differed (in sometimes subtle ways) from the curriculum's learning goals and additionally to her knowledge of students.

Whereas the study, due to its design, cannot be generalized to the entire population of elementary (science) school teachers, the study enables to inform about theoretical and practical implications regarding teachers' use and adaptation of curriculum materials. As a first major contribution, the study suggests that the changes teachers make to the curriculum might be more productive if teachers explicitly consider their knowledge of students and the alignment between their own goals and the curriculum's goals. As such, the study sheds light on how teachers adapt curriculum materials. A second contribution relates to the fact that even a subtle lack of alignment of goals can influence the adaptation and the enactment of the curriculum. As a third contribution, the study suggest that teachers likely need support in making productive changes to curriculum materials, and thus, teacher educators should work to support teachers in this regard. Two implications for teacher education, professional development and educative curriculum materials are mentioned. First, learning opportunities around curriculum materials should support teachers in understanding the learning goals of the curriculum material – this might be helpful for teachers to identify mismatches with their own goals. Second, teachers should be helped to develop strategies for better knowing their students. Future research should explore these ideas further.

Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using Educative Curriculum Materials to Support the Development of Prospective Teachers' Knowledge. Educational Researcher, 43(3).

Journal category: General education

Criterion

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.

Description

Drake et al. (2014) describe that the development of educative curriculum materials – these are materials that support student learning as well as teacher learning –provided an opportunity for changing the prevailing view of good teachers as those who don't use curriculum materials into "those who use educative curriculum materials well" (Drake et al., p. 154). This has consequences for curriculum designers whose focus is shifting to support teachers to customize the curriculum materials to the students in class while still remaining faithful to the curriculum developers' intend. Along this line, the authors suggest that teacher education as well should change their goals to support prospective teachers (PTs) to flexibly use such curriculum materials rather than creating either their own materials or implement curriculum materials with high fidelity.

This study builds directly on work of Ball and Cohen – "What is—or might be—the role of curriculum materials in teacher learning and instructional reform?" (1996, p. 8) – and that of Davis and Krajcik – "How can K-12 curriculum materials be designed to best promote teacher learning?" (2005, p. 3) – and proposes that educative curriculum materials should be utilized in teacher education to support PTs to develop the multifaceted knowledge needed to teach successfully. Drake et al. ground their argument in two research fields. First, they build on Shulman (1986, 1987) and Hill, Ball and colleagues (e.g., Ball, Thames, & Phelps, 2008; Heather C Hill, Sleep, Lewis, & Ball, 2007) to illustrate the multifaceted structure of the knowledge needed to teach mathematics well. Second, they review what is currently known about teacher learning from using educative curriculum materials. They argue that most studies have focused primarily on teacher learning *about* those materials whereas teacher learning *from* those materials conceived to a much lesser extent attention.

In this study, Drake et al. argue that "educative curriculum materials can and should be utilized in teacher preparation to support PTs in developing not only knowledge and practices related to curriculum materials and their use, but also the broader knowledge bases needed for successful novice teaching" (p. 155). To that end, two conceptual examples are presented to illustrate that educative curriculum materials can contribute to the development of PTs multifaceted knowledge base. Based on theory, prior research in mathematics and science education, and on their own empirical work, Drake et al. present the following five design principles to help PTs learn to use curriculum materials in educative ways:

- Principle 1: PTs need opportunities to learn that curriculum materials contain educative features to support teacher learning, as well as student learning, and to orient themselves to reading these materials in ways that attend to the educative features.
- Principle 2: PTs require tools to serve as lenses through which the educative features of curriculum materials can be read, understood, evaluated, and adapted.
- Principle 3: PTs require scaffolds to prompt the development of multiple knowledge bases through the use of curriculum materials.
- Principle 4: PTs need to examine multiple lessons and units in order to identify and understand the development of content over time.
- Principle 5: PTs need experiences reading across a variety of curriculum materials in order to compare and contrast the treatment of content.

The study ends by describing a number of implications for future research such as finding an appropriate balance of focusing in depth on the use of one particular set of curriculum materials and introducing PTs to a variety of curriculum materials, and how the use of digital curricula might influence teachers' use of the educative aspects of those digital curriculum materials.

Stein, M. K., & Kaufman, J. H. (2010). Selecting and Supporting the Use of Mathematics Curricula at Scale. *American Educational Research Journal, 47*(3), 663-693.

Journal category: General education

Criteria

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.
- Criterion 2: Study describes how teachers prepare for teaching.
- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.
- Criterion 3.2: Combine inclusion of schools and districts as settings for curriculum implementation.

Description

This study starts off with a description of a district-wide adoption of ambitious mathematics curricula as an attempt to improve instruction and learning on a large scale. The authors argue that, while it may be relatively easy to disseminate the curriculum programs among a considerable amount of teachers, it is much more difficult to select curriculum programs appropriate for specific schools and districts as well as designing conditions to implement these curriculum programs in line with the authors' intend. The central question, posed in this study, relates to what curriculum programs work best under what conditions?

Two ambitious, standards-based curriculum, programs are studied: Everyday Mathematics (EM) and Investigations (INV). Previous research has shown that INV contains more tasks characterized as *doing mathematics* whereas EM contains more tasks characterized as *procedure with connections.* Although both types of tasks are considered as high demanding for students in terms of cognitive load, *doing mathematics* tasks puts more demand on the teacher than do tasks characterized as *procedures with connections*. It could thus be argued that EM is easier to implement than INV. However, earlier research of the authors found that the learning opportunities in both curriculum programs match the demands posed by the programs on the teachers. It was found that INV, the more demanding curriculum program, also provided more learning opportunities for the teachers than did EM, the lower demanding curriculum program. This complicates the abovementioned line of reasoning in that it suggests that teacher capacity might be stronger related to the quality of instruction in districts using INV. The idea is that high-capacity teachers are expected to benefit more from the learning opportunities than low-capacity teachers. To explore this idea, the authors draw on a large set of lesson observations (n = 511) and teacher surveys (n = 86), following teachers of two districts during two consecutive years. One district uses EM; the other district uses INV.

It is argued that both teacher capacity and teachers' use of the curriculum influence the quality of the implementation of the curriculum programs. Quality of implementation (thus, of the actual observed lessons) centers on maintaining a high cognitive demand, attending to student thinking, and vesting intellectual authority in students' mathematical reasoning (as opposed to text and teacher). Teacher capacity is determined by teachers' mathematical knowledge for teaching, experience, education and professional development. Teachers' use of the curriculum programs is operationalized in their perceptions about the usefulness of the curriculum program, the percentage of time that teachers used the curriculum programs, and what they reviewed and talked about when preparing the lesson (divided into non-mathematical, big mathematical ideas, materials).

The results show that quality of implementation is higher in the district where teachers use INV. This challenges the conventional wisdom that EM is easier to implement than INV. Further, the results illustrate that only total amount of hours professional development is the only determinant of teacher capacity that influences the quality of instruction in a significant, beneficial way – although only in the district that adopted INV. More important than teacher capacity seems, as suggested by the results, *how* teachers use the curriculum program. Whereas perceptions about curriculum usefulness is positively associated with instructional quality in the INV

district, especially reviewing the big mathematical ideas of a lesson while preparing for teaching seems to be of importance, given that this plays out in both districts. Thus, unlike teacher capacity, how teachers use the curriculum seems to shape the quality of the lesson.

A further examination of how the curriculum programs might contribute to teachers attending to the big mathematical ideas – based on teacher surveys – revealed that two aspects might be beneficial. A curricular focus on a limited number of mathematical ideas, and explicit supportive information in the curriculum, toward the teacher are deemed to be fruitful. The findings have implications in that they lay out points of interest for administrators facing the choice for a district-wide curriculum program. First, administrators should consider the affordances provided by the programs to help teachers identify the big mathematical ideas of the lessons. The study suggests that the focus on a single mathematical concept per lesson, and clarity in and support for the teacher to present this concept or idea to the class are particular elements of interest in this regard. Second, the study suggests that the adoption of a curriculum program should be supplemented with a professional development program to help teachers identifying the big mathematical ideas in the curriculum program.

Davis, E. A., Palincsar, A. S., Arias, A. M., Bismack, A. S., Marulis, L. M., & Iwashyna, S. K. (2014). Designing Educative Curriculum Materials: A Theoretically and Empirically Driven Process. Harvard Educational Review, 84(1), 24-52.

Journal category: General education

Criterion

- Criterion 1: Describe design principles of educative curricula, these are curricula that are also educative for teachers, and how teachers respond to or use these features.

Description

In this study, Davis and colleagues describe a design-based research approach to designing educative curriculum materials. These are resources designed to be used by teachers in classroom to guide instruction that also support teacher learning. The actual focus of this study relates to teacher's guides in the context of 4th grade science lessons but it is argued by the authors that the principles and processes applied generalize to the design of educative curriculum materials across subject matter areas.

The authors draw on earlier work to describe a design process that finds its foundations on both empirical and conceptual grounds. This is considered important, given that the development of educative curriculum materials typically seems to be based on intuition or theory and has little origins in empirical observations of teachers' use of curriculum materials during class.

Basically, the design-based research approach consists of three iterative steps: designing materials, testing the materials, and refining the materials. The design process includes an analysis of existing curriculum materials in terms of potential learning opportunities for students, lesson observations to observe actual lesson opportunities, and characterization of students' learning outcomes. This process is guided through five lenses: an analysis of the science content, scientific practices, literacy practices, participation structures, and assessment.

In a first step, existing curriculum materials were screened to identify aspects of learning opportunities in correspondence with the five lenses. This also enabled the researchers to develop an observation protocol closely connected to the written lesson plan. In a second step, actual lessons were observed to identify ways that teachers implemented and adapted each lesson. The authors also drew on related student work outside of the observed lessons. The second step provided information about the actual learning opportunities and enabled a comparison with the learning opportunities as specified in the teacher's guide. In a third step, authors characterized student work in terms of the scientific concepts and scientific practices. This allowed for a comparison with what might been expected, given the learning opportunities in the curriculum materials and during class.

Five educative features are described; central for all features is the provision of a rationale for recommendations and guidance to how these recommendations can be implemented and adapted by the teachers to fit the students in their class. Content support features include unit core ideas maps, concept maps, content story lines, and content boxes. Scientific practices educative features included practice overview pages, why-and-how texts, and reminder boxes. As a third type of educational feature, narratives – or vignettes – were included. Such narratives contain considerable detail about how a fictive teacher implemented the lesson. A fourth educative feature aimed at supporting literacy practices. Such features targeted delineating content ideas while reading texts and providing teachers example questions to elicit these content ideas when discussing the content. A fifth educational feature aimed to support assessment practices – to support teachers in assessing as well as helping teachers to anticipate student thinking. This was done by provision of assessment rubrics and sample student work with suggested feedback.

Informed by the information provided in the three abovementioned steps and drawing on existing literature, the authors made specific design decisions. Two such cases are described to illustrate how the authors determined areas that were in need of additional support, and how decisions were made to provide that support.

Based on a tryout of the enhanced materials (the original curriculum materials infused with educative features) the authors refined their design of educative curriculum materials in three ways: developing more extensive literacy educative features, developing capstone questions to support the teacher in understanding why and how to engage students in sense making, content charts to provide teachers specific language support.

In line with design-based research, four aspects of this study are mentioned. First, the authors sought to address questions about the nature of elementary school science learning in context. Second, the study took place in real-world contexts. Third, the authors drew on a variety of information sources to guide design and refinement processes. Fourth, as the goal of the work was to base the design on process-related data, the authors argue that formative assessment is an intrinsic feature of this study. As argued by the authors, this study adds to the research field by conceptualizing the principled design of educative curriculum materials.

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, 47*(4), 833-878.

Journal category: General education

Criteria

- Criterion 3.2: Combine inclusion of schools and districts as settings for curriculum implementation.
- Criterion 4.1: Describe effectiveness of curriculum materials and provide data on degree of implementation.
- Criterion 4.2: Describe what and how students learn from interacting with particular curriculum materials, rather than describing how much they learn.

Description

This quantitative study evaluates the impact on student learning when "business-as-usual" teaching of units on proportionality and linear functions were replaced by units (covering the same topics) where use of interactive representation technology was integrated with the written curriculum. Though it was an intervention study it was designed to be in the form of a large-scale implementation with what could be considered as an affordable amount of professional development and minimal outside support and involvement from the researchers and the developers of the replacement units.

The replacement units – one on rate and proportionality for Grade 7 and one on linear functions for Grade 8 – follow the *SimCalc* approach and make use of the *MathWorlds* software, which has a strong representational focus. The software produces animations of movement and speed based on algebraic and/or graphical representations provided by the student. The Texas state curriculum standards (Texas is the geographical location of this study) have a *formula* approach to proportionality and linearity. This typically involves finding a number given two or three other numbers, graph reading without interpretation, categorization of functions, working within one representation, etc. The authors refer to this mathematics as M1. The SimCalc approach also includes a *function* approach, and asks students to consider the mapping between domain and range, to connect concepts like rate across multiple representations, etc. This mathematics is denoted M2. For both grades, the SimCalc unit was designed to be used daily over a 2- to 3-week period. The material consisted of student workbooks, a teacher's guide and MathWorlds files.

The study consists of three experiments, referred to in this summary as Exp7, Exp8 and Quasi7:

- Exp7 is an experiment in 7th grade where one group of teachers (referred to as *treatment* teachers) uses the replacement unit on rate and proportionality. Student gains are then compared with students of teachers in a control group (referred to as *control* teachers), teaching the same unit in a "business-as-usual" way. A total of 95 teachers completed the experiment, with 796 students in the treatment group and 825 in the control.
- Exp8 is an analogous experiment in 8th grade, with other schools and teachers, where treatment teachers use the replacement unit on linear functions. Fifty-eight teachers completed the experiment, with 522 students in the treatment group and 303 in the control.
- Quasi7 is a quasi-experiment in 7th grade where teachers in the control group of Exp7 during the next school year used the replacement unit in their new 7th grade classes. Thus it was possible to compare teaching outcome when the same teachers did not and when they did us the replacement unit (these teachers are referred to as *delayed-treatment* teachers). 67 teachers completed the experiment, 30 of them in the delayed-treatment group. The total number of students in the delayed-treatment group was 510 Year 1 and 538 Year 2.

Teachers volunteered to participate in the experiment. Among their schools, teachers were chosen randomly and half of them randomly put in the treatment group, the others in the control group. Most schools had only

one participating teacher but when there were more than one all of them were either in the treatment or in the control group. Schools from seven of 20 regions in Texas participated. The samples were diverse in terms of campus poverty level, school size and campus ethnicity, as well as in teachers' gender, ethnicity, teaching experience and mathematical knowledge.

All teachers in Exp7 (treatment as well as control) got professional development on the mathematical content of the unit including "non-SimCalc" materials that could be used to supplement their units to teach M2 concepts. Treatment teachers also got professional development on the SimCalc material, led by the researchers/developers. Delayed treatment teachers got the same professional development on SimCalc before Year 2. Exp8 treatment teachers only got professional development on the SimCalc, held by a group teachers trained by the researchers/developers according to a "train-the-trainer" concept. Control teachers had a workshop on important content knowledge.

Student learning was assessed using a test designed by the researchers (the design process is described in detail in the article) which covered M1 and M2 topics. Students were given the same test before and after the unit. A number of implementation variables were measured. The study reports on three of them: (1) days spent by students in the computer lab, (2) days spent by teacher on teaching the unit, and (3) the extent of which different subtopics of M1 and M2 were covered (M1 were divided into 12 subtopics, M2 into 5).

Data were analyzed statistically with a 2-level, hierarchical linear model described in the paper. In all three experiments treatment students gains (difference between post- and pre-test) were significantly higher than control students (effected sizes between .50 and .63, $p<.005$). On the M1 subscale the differences were small and not significant but on M2 effect sizes were in the range 0,69-0,89, with $p<.005$. With only a few exceptions population factors (like gender, ethnicity and prior achievement) were not significant predictors of learning gains.

Students in treatment groups spent significantly more days in the computer lab than control students, which is in line with the SimCalc program theory. In Exp8 days in lab was also a significant predictor of learning gain, but not in Exp7. Treatment teachers in Exp7 reported significantly more days spent on teaching the unit than did control teachers (mean 14.9 and 12.0 respectively). In Quasi7 there was small, but not significant, decrease from Year 1 to Year 2, i.e. they spent more days on the unit when teaching as usual than when using the replacement unit. In Exp8 treatment teachers also reported a not significant lower number of days spent on the unit than control teachers did. However, there was no significant correlation between days spent on the unit and student learning on M2 topics. Thus student gains cannot be explained by more to time spent on the unit. The only significant difference between topic coverage was that in Quasi7 more days were spent on basic operation in Year 1 than in Year 2. Treatment teacher generally reported higher coverage of topics that are signature aspects of SimCalc (such as multiple representations). Thus treatment teacher interacted with the material in ways that focused on more advanced topics without neglect of the basics.

In summary the study shows that even with small efforts in terms of professional development it is possible, among a diversity of teachers and students, to significantly increase students' advanced knowledge without losing basic skills. Even though students differed greatly in baseline tests according to their ethnicity, region, socio-economic background, etc., they had very similar learning gains. While other studies have shown small or no effects of use of technology this study shows that with another approach to technology, namely a representational focus with tight integration of technology, curriculum and professional development, robust effects can be achieved. This approach is also the authors' recommendation to the field. They also point out that if they only had measured "basic skills" no difference in learning gains would have been discovered. Thus representational technology is a possible way to reach beyond basics.

Coburn, C. E., Mata, W. S., & Choi, L. (2013). The Embeddedness of Teachers' Social Networks: Evidence from a Study of Mathematics Reform. *Sociology of Education, 86*(4), 311-342.

Journal category: Sociology

Criterion

- Criterion 3.2: Combine inclusion of schools and districts as settings for curriculum implementation.

Description

Coburn et al. (2013) investigate the relationship between policy and social networks, how social policy penetrates the organizational boundaries of schools and influences teachers' social networks and how these networks might be embedded in policy in the context of implementing a new curriculum program. It suggests that policy influence three dimensions of teachers' social networks: (1) tie formation and maintenance, (2) resources, and (3) nature of interaction.

The study draws on data from a three-year longitudinal study in an urban school district in southwestern U.S (the Greene School District). The district had adopted a new curriculum (*Investigations in data, number, and space*) the first year of the study and launched an initiative to support implementation. In Year 1 the district initiated multiple structures that required teachers to interact with new people around mathematics. Each school appointed at least two part-time mathematics coaches who worked with the teachers and taught part-time. Weekly grade-level meetings for planning and biweekly school-based professional development were instituted. A district-level team supported the coaches and professional development was also provided in the summer. In Year 2 the district offered additional professional development for teachers in cross-district settings and on school-level it shifted to cross-grade-level configurations. Professional development for coaches continued and for all (teachers as well as coaches) became more focused on student learning and the nature of mathematics. In Year 3 the district shifted focus from mathematics to education of ELL (English Language Learners) students. Time for mathematics instruction was reduced and resources for professional development were directed to ELL instruction. District funding of mathematics coaches was ended. As a consequence, many structures for interaction around mathematics were dismantled.

For this study, four schools were chosen to represent the four possible combinations of strong/weak professional community and strong/weak teacher expertise. Twelve focal teachers from these schools were interviewed and observed in class several times during each school year of the study. An additional number of six teachers per school were interviewed once per year. The schools' coaches and principals were interviewed at least once a year and at each school, on three to five occasions per year, observations were made when teachers interacted about mathematics instruction. District leaders were also interviewed, professional development sessions with coaches and teachers were observed and relevant district documents were analyzed. Parts of the data collection were designed to identify the focal teachers' egocentric social networks. They were asked about whom they sought out for advice on mathematics instruction, and why, and about frequency and content of these interactions. Other parts of data were collected to investigate in more depth the nature of the focal teachers' networks. (The article has a comprehensive appendix including definitions used in coding, school and focal teacher characteristics, interview questions and further details on sample, the egocentric approach to social networks, social network measures, etc.)

In Year 1 teachers' networks were small (on average including four others) and most ties were to grade-level collages (53%) and coaches at the school (28%). In Year 2 the networks became essentially larger (11 others) with more ties across grades and outside school and less with grade-level colleges and coaches. But in Year 3 the networks contracted somewhat and most ties were once again to grade-level colleges (71%). The reasons for whom to seek out also shifted. In Year 1 proximity and homophily were the main reasons, while in Year 2 the reasons related to reform activities, which seemed to create a shared focus across grades and schools. But when structured opportunities for interaction ceased in Year 3 teachers were unable to maintain these ties. However, even though networks decreased in size and diversity in Year 3, teachers were able to maintain

expertise in their networks. Most teachers increasingly went to others for reasons of expertise across the years of the study. Also, focal teachers' perceptions of expertise moved closer to the authors' assessment of expertise. The authors relate this to the fact that meetings in Year 2 became more focused on mathematics teaching and learning and hence created opportunities for people to learn the location of expertise. While teachers in the beginning mainly sought expertise among the coaches, they also turned to grade-level colleges in Year 2 and to colleges in other grades in Year 3.

Policy influenced how resources in terms of information, materials and expertise flowed or was accessed through interactions in teachers' networks. District information declined over the years while information from other sources increased. District information was most likely accessed from coaches, but people outside school also played an important role in the beginning. Over time, though, teachers relied more on one another (within schools) to learn about district information. District materials were also accessed indirectly via networks, but in contrast to district information, exchange of materials stayed on the same level even when the district withdrew support for the mathematics initiative. This suggests that materials are a more durable resource than information. However, the role of coaches in bringing district material to teachers' networks expanded and contracted as the district initiative did so. Most of the people in teachers' networks with expertise had it because they had participated in district sponsored professional development, and their part increased over the years. By investing in professional development there were more people to choose from when teachers got increasingly interested in, and sought for, expertise. Like materials, but unlike information, expertise was proved a durable resource. Coaches were an especially important source of district-developed expertise, but their role became diminished when the district initiative waned.

The district influenced interaction in networks by designing specific "routines of interaction" that diffused to the school via the coach and turned up in teachers' interactions with each other. Eight distinct routines of interaction that mathematics leaders repeatedly used with coaches were identified, including task analysis, investigation of students' strategies, etc. The presence of these routines in interactions increased from Year 1 to 2 and declined in Year 3. They also moved from being largely confined to coach-teacher interactions in Year 1 to teacher-teacher interactions within grade-level and school in Year 3, where they seemed to remain even after district support for such interaction was withdrawn. Further, interactions that involved district-designed routines more often yielded a high or moderate depth. "When teachers engaged in district-designed routines, their interaction was extended, structured, and focused on mathematics teaching and student learning. Thus, district-designed routines fostered interaction that was distinctly counter-normative" (Coburn et al., 2013, p. 327).

The study provides evidence for how teachers' social networks are amendable to outside influence and how policy can foster conditions in schools for teachers to seek out their colleagues, share information and learn from each other. It also shows how new ties, especially those that stretch beyond functional areas, decay with time when supporting structures are removed. Finally it suggests that to understand the role of teachers' social relations, greater attention to the social networks themselves is required.

Voogt, J. (2010). Teacher factors associated with innovative curriculum goals and pedagogical practices: differences between extensive and non-extensive ICT-using science teachers. Journal of Computer Assisted Learning, 26(6), 453-464. doi: 10.1111/j.1365-2729.2010.00373.x

Journal category: Technological resources

Criterion

- Criterion 3.1: Relate to an integrated analysis of curriculum resources and teacher resources, and how they interact.

Description

Voogt (2010) explores differences in pedagogical orientations, ICT competencies and professional engagement between extensive and non-extensive ICT-using science teachers. The study is based on a second analysis of data from SITES 2006 (Second Information Technology in Education Study), in which schools and Grade 8 math and science teachers from 22 educational systems participated, with 400 schools per system and four teachers per school. For the study reported here, the Grade 8 science teachers who had answered affirmative to the question whether they "used ICT in teaching and learning activities in Grade 8" were chosen and divided into three subgroups depending on whether they used ICT every week, extensively during a limited period, or none of these two. The first subgroup (n=1754) were considered extensive ICT users, the third (n=1273) non-extensive. Only these two subgroups of teachers were used in the analysis.

Based on teachers' answers to questions in SITES 2006 indicators were developed for pedagogical orientation, ICT competencies, and professional engagement. Pedagogical orientation was based on indicators for curriculum goals and pedagogical practice and a distinction was made between a pedagogical orientation reflecting traditionally important goals and practices and a pedagogical orientation reflecting lifelong learning goals and practices. The former complies with needs in the industrial society and aligns with a teacher-centered view on teaching and learning; the latter complies with the needs of the knowledge society and aligns with learner-centered teaching and learning.

Both groups of teachers (extensive and non-extensive ICT users) considered traditional as well as lifelong learning curriculum goals and pedagogical practices important, but extensive ICT users had a stronger orientation to lifelong learning than non-extensive ICT users. Both groups more often used ICT to support traditionally important practices than lifelong learning practices, but extensive ICT users more often applied ICT to lifelong learning practices than non-extensive ICT users did.

Extensive ICT users considered themselves more confident in ICT competencies than non-extensive users, especially when it came to competencies related to pedagogical use of ICT. They were also more professionally engaged: they more often co-taught with other teachers, they more often collaborated with teachers from other schools, and they attended more ICT-related professional development. In particular, extensive ICT users had attended more subject specific ICT-courses.

In summary, there were positive relationships between (self-reported) frequency of ICT use and a pedagogical orientation to lifelong learning, and between ICT use and professional engagement. According to the author the findings underline the often expressed need for pedagogically oriented ICT-related professional development and the findings also support the notion that teachers need technological pedagogical content knowledge. But the author also points at teachers needing time to build routines in the integration of ICT in their educational practice.

Vetenskapsrådet genomförde under 2014 ett projekt, SKOLFORSK, för att kartlägga befintlig utbildningsvetenskaplig forskning. Arbetet skedde på uppdrag av regeringen för att resultera i kartläggningar av svenska och internationella forskningsresultat med relevans för skolväsendet. Syftet var att skapa en plattform av kunskapsunderlag till det nybildade Skolforskningsinstitutet. Slutsatserna i denna delrapport är författarnas egna. Vetenskapsrådets sammanfattande rapport, Forskning och skola i samverkan, med en beskrivning av projektet och med de frågeställningar, resultat och rekommendationer som redovisats inom delprojekten kan liksom de övriga delrapporterna laddas ner från Vetenskapsrådets webbplats.

Västra Järnvägsgatan 3 | Box 1035 | 101 38 Stockholm | Tel 08-546 44 000 | vetenskapsradet@vr.se | www.vr.se