Vetenskapsrådet

Delrapport från SKOLFORSK-projektet

FORMATIV BEDÖMNING PÅ 2000-TALET

– en översikt av svensk och internationell forskning

VETENSKAPSRÅDETS RAPPORTER 2015

FORMATIV BEDÖMNING PÅ 2000-TALET – EN ÖVERSIKT AV SVENSK OCH INTERNATIONELL FORSKNING

VETENSKAPSRÅDET
Box 1035
SE-101 38 Stockholm, SWEDEN

© Swedish Research Council
ISBN 978-91-7307-269-4

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.

FORMATIV BEDÖMNING PÅ 2000-TALET

Åsa Hirsh och Viveca Lindberg

INNEHÅLL

FÖRORD 3
SAMMANFATTNING 4
ENGLISH SUMMARY 5
FÖRFATTARPRESENTATION 6
1 INLEDNING OCH RAM 7
1.1 Bakgrund 7
1.2 Teoretisk ram/bedömningstraditioner 9
1.3 Studiens syfte och frågeställningar 12
2 ÖVERSIKTENS DESIGN OCH METODISKA ÖVERVÄGANDEN 13
2.1 Design: val av söktjänster och sökstrategier 13
Procedurer för sökning och urval av internationellt publicerad forskning 13
Procedurer för sökning och urval av forskning publicerad på svenska 15
Nordiska utblickar 16
2.2 Bearbetning och analys 17
3 RESULTAT 18
3.1 Formativ bedömning i internationell forskning 18
Generell överblick 18
Teoretisk förståelse av formativ bedömning/begreppsutveckling 20
Policy 26
Verktyg 30
Implementering/Professionellt lärande 41
Elevers uppfattningar 45
3.2 Formativ bedömning i svensk forskning 48
Dokumentation som grund för formativ bedömning 54
Formativ bedömning i klassrumskommunikationen 59
Formativ bedömning och skolans styrning 65
Sammanfattning av de svenska studierna 67
3.3 Nordiska utblickar 67
4 SLUTSATSER OCH DISKUSSION 70
4.1 Formativ bedömning – teori, begrepp och metoder 71
4.2 Formativ bedömning – hinder och möjligheter för ett framgångsrikt arbete 72
4.3 Värdet av befintlig forskning om formativ bedömning i relation till svenska sammanhang 74
4.4 Kartläggningsmodellens starka och svaga sidor 76
4.5 Rekommendationer 77
REFERENSLISTA 79

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

SAMMANFATTNING

I denna översikt kartläggs och redovisas forskningsläget i Sverige och internationellt med avseende på 2000-talets forskning kring formativ bedömning i grundskolan eller motsvarande. Vår analys visar att teoretiska studier i allt högre utsträckning positionerar formativ bedömning som en kontextbunden social praktik, snarare än en uppsättning generella metoder. Samtidigt har ett fåtal stora metastudier som presenterar anmärkningsvärda effektstorlekar på ökat lärande hos elever lett till att man på policynivå fattat beslut om storskalig implementering i många länder i världen. Forskare tycks överens om att klassrumsimplementering av formativa praktiker behöver stödjas av processer där professionellt lärande ges utrymme, men i realiteten saknas ofta sådana processer. Man har kunnat se att avsaknaden av kollegialt lärande bland lärare och skolledare många gånger innebär att pseudo-formativa praktiker utvecklas, där förståelsen av formativ bedömning tenderar att bli instrumentell och kopplad till ett mer rituellt arbete med ett generellt metodpaket. Vi har också sett att IKT-relaterade verktyg för formativ bedömning är på stark frammarsch. Då formativ bedömning betraktas som resurskrävande ur lärarsynpunkt anses IKT-verktygen vara ett sätt att effektivisera undervisningen. Påtagligt i vissa av dessa studier är att man tenderar att betrakta datorn/den mobila enheten som en *egen aktör*, som är den som analyserar elevens ”lärande” (svar på frågor) och ger feedback. Som följd av detta förutspår vi att frågor som blir viktiga att studera framöver exempelvis gäller vilken typ av feedback som kan ges av datorer, och med vilken kvalitet den kan ges. Konsekvensen av digitala prov och digital feedback för lärares professionalism blir naturligtvis en annan viktig fråga, då formativ bedömning ska tjäna syftet att utveckla inte bara eleven, utan även lärarens förståelse av hur undervisningsprocesser kan anpassas för att möta elevers behov. En tredje intressant fråga, kopplad till den förra, är naturligtvis också vem som utvecklar de program som används i detta syfte.

På ett generellt plan kan konstateras att empiriska studier genomförda i grundskolemiljö är få, och att man ofta stödjer sig på studier genomförda i högre utbildning då man uttalar sig om gynnsamma effekter av formativ bedömning på elevers lärande. Metastudier som uttalar sig om generella effekter av formativ bedömning problematiseras, bland annat för att paraplytermen formativ bedömning innefattar så många och disparata företeelser att det är närmast omöjligt att uttala sig om en samlad effekt.

Det svenska forskningsfältet är mycket knapphändigt, i synnerhet med tanke på hur stort genomslag formativ bedömning som en uppsättning metoder/arbetssätt fått i den svenska skolan. Ändå visar resultaten från de olika studierna på en viss samstämmighet. Studier av dokumentation, främst genom individuella utvecklingsplaner, dominerar forskningsfältet tillsammans med klassrumsstudier av lärares formativa bedömningsarbete. En typ av studier fokuserar också den nationella – och i viss mån den kommunala - styrningen av skolan och vilka konsekvenser olika typer av styrning kan få för lärares bedömningsarbete. Behovet av olika typer av studier som innefattar flera nivåer av aktörer är påtagligt: en samverkan mellan forskning baserad på kvalitativa respektive kvantitativa data behövs. Interventionsstudier som mäter effekten av olika aspekter av formativa praktiker behöver exempelvis kompletteras med studier av hur lärare, skolledare och huvudmän analyserar och använder bedömningsinformation för att forma den undervisning som i förlängningen ska leda till ökat lärande/ högre måluppfyllelse hos elever. Dessutom behövs studier av svenska lärarutbildningar, mot bakgrund av att internationell forskning visar att alldeles för få personer inom lärarutbildningar har den fördjupade kunskap som krävs för att utbilda studenter i bedömningsfrågor. Vår översikt visar också att en påtagligt låg andel studier tar elevernas perspektiv, och vi vet därför mycket lite om hur de upplever och påverkas av olika arbetsmetoder som klassas som formativ bedömning.

ENGLISH SUMMARY

This systematic review maps and reports on 21st century Swedish and international research into the field of formative assessment in compulsory school. Our analysis shows that theoretical studies are increasingly positioning formative assessment as a context-bound social practice, rather than a set of general methods. Meanwhile a few large meta-studies, featuring remarkable effect sizes of increased learning on behalf of students, have contributed to policy decisions advocating large-scale implementation of formative assessment practices in many countries around the world. Researchers seem to agree that classroom implementation of formative assessment practices needs to be supported by processes in which professional learning is given space, but in reality such processes are often lacking. It has been seen that the lack of peer learning among teachers and school leaders often means that pseudo-formative practices evolve, where the understanding of formative assessment tends to be instrumental and linked to ritual work with a general method package. We have also seen that ICT-related tools for formative assessment are gaining ground. When formative assessment is regarded as time consuming from the teacher point of view, ICT tools are seen as one way to streamline instructional processes. It is evident in some of these studies that computers/mobile devices tend to be understood as independent actors, who in themselves are regarded as formative, i. e. as being the ones analysing students' "learning" (answers to questions) and providing feedback. As a result, we predict that important issues for future research will concern, for instance, which type of feedback that can be provided by computers, and with what quality it can be provided. The consequence of digital tests and digital feedback for teachers' professionalism is of course another important issue, since formative assessment serves to develop not only the students but also the teachers' understanding of how learning processes can be customized to meet students' needs. A third interesting question, linked to the former, concerns *who* develops the software used for this purpose.

Generally, it can be concluded that empirical studies conducted at compulsory school level are few, and that we often rely on studies conducted in higher education when commenting on beneficial effects of formative assessment on student learning. Meta studies dealing with general effects of formative assessment are problematized, in part because the umbrella term formative assessment involves so many and disparate phenomena that it is problematic to speak of *one* overall effect.

Swedish research into formative assessment is rather scarce, especially considering how large impact formative assessment as a set of methods/approaches seems to have in Swedish schools. Nevertheless, the results from the various Swedish studies point to a certain coherence. Studies of documentation, mainly through individual development plans, dominate the research field, along with a few classroom studies of teachers' formative assessment work. There are also studies that focus on national - and to some extent local/municipal - governance of the school, and the consequences of various types of control for teachers' assessment work. The need for different types of studies that include the perspectives of multiple levels of actors, is evident: a collaboration between research based on qualitative and quantitative data are needed. Intervention studies that measure the impact of different aspects of formative assessment need to be complemented by studies of how teachers, principals and local/municipal policy levels analyse and use assessment information to shape the instructional processes that will ultimately lead to increased learning/higher goal attainment among students. In addition, surveys and studies of Swedish teacher training would be desirable, given that international research shows that too few educators within teacher education programs have the in-depth knowledge required for adequately making teacher students assessment literate. Our review also shows that a very small percentage of studies take the students' perspective, wherefore we know very little about how they perceive and are affected by various methods which are classified as formative assessment.

FÖRFATTARPRESENTATION

Viveca Lindberg är docent i didaktik vid institutionen för pedagogik och didaktik vid Stockholms universitet och gästforskare vid institutionen för pedagogiska studier vid Karlstads universitet. Hon var koordinator för den av Vetenskapsrådet finansierade Nationella forskarskolan i pedagogisk bedömning 2007-2012 och för Nätverket Pedagogisk bedömning 2011-2014.

Åsa Hirsh är lektor i pedagogik vid Högskolan för lärande och kommunikation i Jönköping. Hon har skrivit sin avhandling inom bedömningsfältet och sedan fortsatt bedriva bedömningsrelaterad forskning. Även Åsa var knuten till forskarskolan i bedömning och är med i Nätverket Pedagogisk Bedömning.

1 INLEDNING OCH RAM

Formativ bedömning är ett begrepp som uppfattas som relativt nytt och kopplas ofta samman med en forskningsöversikt av Paul Black och Dylan Wiliam från 1998. Begreppet är dock inte särskilt nytt utan har funnits med åtminstone sedan 1950-talet. Innebörden i det har emellertid förändrats över tid. Inledningsvis beskrivs bakgrunden till och några specifika aspekter av begreppet, såsom de forskare som tidigt arbetade med det introducerade och betonade det. Därpå följer en teoretisk bakgrund till formativ bedömning, där vi dels skissar på olika innebörder av begreppet beroende på vilket teoretiskt perspektiv man väljer som utgångspunkt, dels tydliggör våra egna ställningstaganden och hur de påverkat vilka frågor vi ställt till de artiklar, kapitel och böcker som ingår i översikten.

I kapitel 2 presenterar vi hur vi designat översikten, det vill säga vilka databaser vi valt för att komma åt forskning om formativ bedömning, vilka nyckelord vi använt för att söka i databaserna och vilka övriga källor vi använt oss av för översikten. Vi presenterar också hur vi bearbetat och analyserat texterna. I kapitel 3 presenteras översiktens huvudsakliga resultat: kapitlet inleds med fem områden/kategorier som kan relateras till internationell forskning, varpå resultatet av svensk forskning följer. Det fjärde och avslutande kapitlet innehåller en diskussion/konklusion, samt rekommendationer till Skolforskningsinstitutet.

1.1 Bakgrund

Formativ bedömning som fenomen har under slutet av 1990-talet fått stor spridning i stora delar av världen, och framför allt har en forskningsöversikt av Paul Black och Dylan Wiliam (1998) varit betydelsefull för det nyvaknade intresset för denna aspekt av bedömningsforskning. Begreppet är dock äldre än så och har även tidigare varit aktuellt också i Norden, framför allt på 1970-talet. Enligt Bertil Roos och David Hamilton (2005) kan den formativa bedömningens historiska och teoretiska rötter spåras tillbaka till 1930- och 1940-talen. En vanlig uppfattning är att uppdelningen i formativ och summativ bedömning handlar om bedömningars olika syften, men Roos och Hamilton påpekar att det även handlar om att de bygger på skilda teoretiska traditioner. Medan summativa bedömningar framför allt kan relateras till psykometri (psykopedagogisk mätningslära) och, menar författarna, behavioristiska grundantaganden om kunskap och lärande, vilar den formativa traditionen på kognitivistiska och konstruktivistiska grundantaganden (mer om detta i kapitel 3).

Även om begreppets utveckling kan spåras så långt tillbaka som Roos och Hamilton funnit, är det framför allt följande texter som lyfts i relation till tidig utveckling av innebörden i formativ bedömning:

- The methodology of evaluation av Michael Scriven1 (1967),
- Handbook on formative and summative evaluation av Benjamin S. Bloom, J. Thomas Hastings och George F. Madaus (1971) och
- Formative assessment and the design of instructional systems av Royce Sadler (1989).

En av de centrala aspekterna i Scrivens kapitel om summativ och formativ *evaluation* (utvärdering), som han skriver om i relation till högre utbildning, berör frågan om formativ utvärdering som en del av curriculum development. Det engelska begreppet curriculum kan översättas till svenska med tre sinsemellan relaterade men delvis olika innebörder: a) curriculum som läroplan/utbildningsplan, b) curriculum som kursplan eller c) curriculum som lärarens plan för undervisningen. Scrivens text har huvudsakligen koppling till den utveckling

[1] Trots att Michael Scrivens artikel har kommit att bli den som fått störst genomslag, finns det också en tidigare artikel av Lee Cronbach (1963), som pekade på behovet av kursutveckling genom utvärdering som inte byggde på jämförelser av normrelaterade studieresultat utan istället på analys på uppgiftsnivå av vad studenterna hade svårt med. Hans artikel fick dock endast begränsad spridning, istället kom det att bli Scrivens artikel som fick genomslag (Stufflebeam & Shinkfield 2014).

av kursplanen som blir möjlig genom implementering.[2] Bloom m.fl. tar sin utgångspunkt i Scriven (1967) och vidareutvecklar formative *evaluation*: ”We regard formative evaluation as useful not only for curriculum construction but also for instruction and student learning” (s. 117). De skriver vidare att ”Formative evaluation is for us the use of systematic evaluation in the process of curriculum construction, teaching, and learning for the purpose of improving any of these three processes” (s. 117). När det gäller deras vidare utveckling av texten är den baserad på dels en behavioristisk syn på lärande och dels en hierarkisk/taxonomisk syn på kunskapsutveckling. Dessa utgångspunkter präglar deras syn på såväl undervisning som på vad läraren bör uppmärksamma för att identifiera vad i undervisningen som behöver justeras, liksom deras syn på vad som utgör tecken på elevers kunskapsutveckling. Inom pedagogisk forskning kom formativ bedömning under den här tidsperioden främst att fokusera utformning av diagnostiska prov å den ena sidan (Black & Wiliam 2003) och elevanalyser för att kartlägga elevernas kunskapsmässiga positioner som utgångspunkt för lärarens undervisningsplanering å den andra[3].

Medan såväl Scriven som Bloom m.fl. använde begreppet *evaluation,* blev *assessment* det begrepp Sadler (1989) använde. Med de närmare 20 år som gått mellan de förstnämnda texterna och Sadlers artikel hade även feedback blivit ett etablerat begrepp inom bedömningsforskning. I sin artikel pekar Sadler specifikt på vilken funktion feedback har i förhållande till formativ bedömning:

> Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student’s competence by short-circuiting the randomness and inefficiency of trial and- error learning. […] Feedback is a key element in formative assessment, and is usually defined in terms of information about how successfully something has been or is being done. […] Broadly speaking, feedback provides for two main audiences, the teacher and the student. Teachers use feedback to make programmatic decisions with respect to readiness, diagnosis and remediation. Students use it to monitor the strengths and weaknesses of their performances, so that aspects associated with success or high quality can be recognized and reinforced, and unsatisfactory aspects modified or improved (a.a., s. 120-121).

Den följande artikeln som vi inkluderat i bakgrundskapitlet är Backs och Wiliams (1998) ovan nämnda forskningsöversikt *Assessment and classroom learning*. Detta är den vanligast förekommande artikeln som andra forskare hänvisar till och citerar i såväl nationell som internationell forskning inkluderad i denna översikt. Redan i inledningen konstaterar de att det saknas en entydig definition av innebörden i formativ bedömning som vunnit brett stöd i forskarvärlden. Den vagare avgränsning de nöjer sig med för sin översikt relaterar till ”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” (s. 7-8). Det som Sadler pekar på, att formativ bedömning används för att informera såväl lärarens undervisning som elevernas lärande, återfinns alltså även i Black och Wiliams utgångspunkter för översikten. De avslutar sin omfattande översikt med en förteckning över elva centrala uppdrag till forskare att studera (s. 58-59, *vår översättning*):

- Vilka antaganden ligger till grund för läroplanen och för lärares didaktiska arbete?
- Vilken/vilka rationalitet/-er ligger till grund för hur det elever förväntas lära sig designas och presenteras?
- Vad karakteriserar det som tas som intäkt för kunnande i elevers svar i relation till olika typer av bedömning?
- Vilka tolkningsramar använder sig lärare och elever av i relation till elevernas svar?
- Vad karakteriserar det lärande som kommer till uttryck i relation till de uttolkningar som kan identifieras?
- Hur fördelas ansvaret mellan lärare och elever i dessa processer?
- Vilka uppfattningar och föreställningar har elever om sig själva som lärande personer, om sitt eget lärande, om syftet för sina studier och om sina studiemetoder?

---

[2] Förutom behaviorismen hade även undervisningsteknologin stort inflytande över 1960-talets forskning, Undervisningsteknologin påverkade såväl utformningen av utbildningssystem (utbildningsplanering) som undervisningsplanering. Det samtida intresset för utveckling av undervisningsmål, undervisning, lärande och målrelaterad bedömning representerades förutom av Benjamin Bloom (1956) även av bl.a., Ausubel (1972), Gagné (1965), Glaser (1963) – jfr Lindberg (2009).

[3] För exempel på hur detta tog form i svensk pedagogisk litteratur jfr Sellergren (1975).

- Vilka uppfattningar och föreställningar har lärare om lärande, om förmågor, om sina elevers framtid och om sin egen roll som bedömare?
- Vad karakteriserar de sociala sammanhangen i klassrum, såsom de utvecklats av elever och lärare och av de yttre ramar dessa föreställer sig att skolsystemet skapar?
- Frågor relaterade till etnicitet, samhällsklass och genus – dessa är föga uppmärksammade i forskning om formativ bedömning
- I vilken mån är kontexten för någon studie överhuvudtaget artificiell och vilka möjliga effekterna har det för resultatens generaliserbarhet?

Författarna kommenterar dessa punkter med att konstatera att frågorna inte låter sig varken besvaras eller kontrolleras genom studier med enbart kvantitativa ansatser, snarare behövs en kombination av studier som även gör det möjligt att komma åt andra och rikare processer och interaktionen inom klassrum. Med utgångspunkter i ovan nämnda texter av Scriven, Bloom m.fl. och Sadler kan man se Black och Wiliams uppmaningar till forskarvärlden som ett slags ingång till översikten: vad har hänt inom forskning om formativ bedömning under 2000-talet?

1.2 Teoretisk ram/bedömningstraditioner

Principiellt baseras olika bedömningstraditioner på skilda teoretiska perspektiv[4] på framför allt kunskap, lärande och undervisning. I relation till dessa får också synen på vad formativ bedömning kan/skulle kunna innebära olika uttolkningar – uttolkningen kan inte stå i logisk konflikt med centrala grundantaganden inom en teori. Här håller vi oss till tre huvudsakliga bedömningstraditioner, psykometriskt grundad normrelaterad bedömning, målrelaterad bedömning i ljuset av kognitivistiska och individkonstruktivistiska perspektiv och målrelaterad bedömning i ljuset av socialkonstruktivistiska/sociokulturella perspektiv.

I Sverige vann den psykometriska traditionen i form av normrelaterad bedömning stort stöd under 1940 talet. Detta måste förstås i förhållande till psykologins intresse för intelligensmätningar under tidigt 1900-tal där intelligenskvoten ses som ett mått på individers medfödda kognitiva förmågor[5]. Testteori och psykometrisk provkonstruktion utgår från snarlika tanketraditioner (Wedman, 2003). På 1960-talet, i samband med implementeringen av grundskolan, skedde den formella implementeringen av norm- /grupprelaterad bedömning i Sverige, men den typen av bedömningar hade i praktiken varit i bruk sedan 1940-talet. Huvudprincipen för den här typen av kunskapsmätningar är att varje individs prestationer jämförs med gruppens prestationer som helhet. Elevernas individuella betyg är således alltid beroende av hur gruppen presterar som helhet, inte av individens prestation som sådan. Betygen fördelas enligt en normalfördelningskurva, vilket gör att samma prestation (i poäng) kan resultera i skilda betyg under olika år. Provuppgifterna konstrueras och prövas ut för att normalfördelningskurvan ska uppstå[6]. Om gruppen som helhet presterar bättre än förväntat höjs kraven (normeringen av provresultat till betyg), och det krävs fler poäng för höga betyg. Om gruppen som helhet presterar sämre än förväntat sänks kraven, d.v.s. det krävs färre poäng för högt betyg. Då grunden för denna typ av bedömningar och betygssättning är baserat på att sortera individer, fyller formativ bedömning ingen funktion.

Mål-/kriterierelaterad[7] bedömning bygger på andra grundantaganden om kunskap och lärande, som påverkat synen på undervisning och bedömning. Det finns dock två dominerande – och sinsemellan konkurrerande –

---

[4] Den typen av grundantaganden som det handlar om, går att spåra tillbaka till skillnader som kom till uttryck redan hos de grekiska filosoferna och utgör en del av vårt västerländska kulturarv (Liedman, 2002).

[5] Detta antagande påverkar såväl konstruktionen av provet, utformningen av mätsituationen och i slutändan även normeringen av testresultatet.

[6] Vissa uppgifter ska kunna lösas av de flesta, medan andra uppgifter är utformade så att endast en liten andel av eleverna förväntas kunna lösa dem.

[7] I Sverige - liksom inom finlandssvensk terminologi benämns detta målrelaterad bedömning (alternativt utvärdering), medan engelsktalande länder istället använder begreppet *criterion referenced assessmen,* d.v.s. kriterierelaterad bedömning.

uttolkningar av den här typen av bedömning. De tidiga uttolkningarna kan ses som representanter för teknisk rationalitet/undervisningsteknologi och kognitivistiska perspektiv, där intresset riktas mot individens tänkande och uttryck för detta (Black, 2001; Shepard, 2001). Dominerande föreställningar till grund för de tidiga versionerna var att kunskap låter sig brytas ned i allt mindre delar och att kunskaper ses som hierarkiskt uppbyggda från kunskaper på låg kognitiv nivå (t.ex. fakta) till kunskaper på högre nivåer (t.ex. analys och slutsatser). Människor uppfattas lära sig genom att först lära sig de kunskapsformer som ligger på hierarkiskt lägre kognitiva nivåer för att sedan lära sig kunskapsformer på hierarkiskt högre nivåer. De första versionerna av kriterierelaterad bedömning kom som alternativ till normrelaterad bedömning och ett intresse för kognitiva kunskapsnivåer vilka sågs som uttryck för individens utveckling. Blooms och Krathwohls analyser av olika typer av kunskap och uppbyggnaden av en taxonomi för den kognitiva kunskapsdomänen på basis av analysen (Bloom 1956) är en av de mest spridda taxonomierna i västvärlden[8].

Den diskussion som fördes om formativ bedömning på 1960- och 1970-talen kom i huvudsak att fokusera diagnostiska prov som underlag för att bestämma var undervisningen kunde börja – på elevens nivå (Black 2001). Även frågor om nivågruppering kan kopplas till kognitivistiska synsätt under denna period och ses som uttryck för uttolkningar av formativ bedömning.

I Sverige infördes aldrig denna första uttolkning av målrelaterad bedömning, trots att en sådan reformering av betygen förbereddes inom projektet Mål och utvärdering i skolan (MUT-projektet) under 1970-talet (Wedman, 20013; Wallin, 2005). Enligt Wedman berodde detta på att ” det så småningom stod klart att detta arbete skulle resultera i en rad besvärande pedagogiska omständigheter, t.ex. att lärarens egen frihet att skapa gynnsamma omständigheter för inlärningen mer eller mindre skulle gå förlorad” (a.a., s. 311).

Sedan 1970-talet har det internationella bedömningsfältet genomgått rätt stora förändringar. Gipps (1994) benämner detta ett paradigmatiskt skifte från en test- och provkultur till en bredare bedömningskultur (*assessment culture*)[9]. Det anses inte längre endast vara standardiserade prov (i Sverige standardprov/centralprov/nationella prov) som ger viktiga indikationer om elevernas kunskapsnivåer, utan även lärarnas klassrumsbedömningar (lärarkonstruerade prov och formativ bedömning) utgör en central aspekt av lärares arbete.

Då det blev aktuellt med en reformering av läroplanen och betygssystemet på 1990-talet blev det en annan uttolkning av målrelaterad bedömning som kom att dominera i Sverige – åtminstone i retoriken. Den svenska uttolkningen kan relateras till den internationella kritik som fördes fram mot de undervisningsteknologiska lösningarna: den vägledande principen för såväl mål som betygskriterier är att de måste utformas som öppna och tolkningsbara för att de inte ska inte låsa lärarens undervisning. Riktningen ska vara tydlig, men ge förutsättningar för olika uttolkningar beroende på såväl lärarens specifika kunskaper som skolans lokala resurser (Popham, 1978; Carlgren, 2002). För de två senaste betygssystemen i Sverige gäller dessutom att betygskriterierna (Lpo-94, Lpf-94) respektive kunskapskraven (Lgr-11, GY-11) inte kan läsas fristående, de utgör inte avprickningsbara listor. De ska istället läsas *i relation till* målen och förmågorna i respektive kursplan. Kursplanens olika delar ska således läsas både i förhållande till varandra och till det specifika innehåll som undervisningen och elevernas arbete fokuserat. Gemensamt för de båda uttolkningarna av målrelaterad bedömning är att elevens prestationer jämförs med målen och kriterierna och betygssätts utifrån hur väl prestationen svarar mot målen. Detta har medfört en central skillnad jämfört med normrelaterad bedömning – inom målrelaterad bedömning tillbakavisas idén om normalfördelning av betyg, och istället förespråkas jämförelser av individens kunnande med samhälleliga förväntningar uttryckta i mål och kriterier. Såväl

---

[8] Man talar sällan om att det inom denna tolkningstradition finns olika syn på vad som ska utgöra grunden för högre respektive lägre betyg. Medan en grupp ser skillnader i kunnande som en fråga om något elever hade mer eller mindre av (andelen avprickade kriterier – ibland omnämnd som en kvantiativ kunskapssyn) så anser andra grupper att skillnaderna bör vara av kvalitativ karaktär. Vidare kan man skilja på dem som ser de kvalitativa skillnaderna som tecken på en hierarkisk förändring av kunnande (jfr Bloom & Krathwohl 1956) medan andra (t.ex. Hirst 1974) menar att skillnaden handlar att kunna urskilja kvalitativt olika nyanser av något (Carlgren, Forsberg & Lindberg 2009).

[9] Notera att ett paradigmatiskt skifte inte innebär att alla under en viss period tänker lika/accepterar samma grundantaganden. Ett paradigmatiskt skifte handlar snarare om att tillräckligt många i ett samhälle enats om rimligheten i argumentationen så att tankefiguren får genomslag i samhället.

kognitivistiska som (individ)konstruktivistiska perspektiv på lärande kan sägas ha vissa gemensamma grunder för en uttolkning av elevprestationer som relaterade till elevens utvecklingsnivå/hur långt eleven kommit i sin kunskapskonstruktion. En central skillnad mellan dessa perspektiv är relaterad till frågan om kunskap: medan kognitivistiska perspektiv har ett grundläggande antagande om att kunskapen ligger utanför människor och så att säga ska hämtas in, tar konstruktivistiska perspektiv avstånd från en sådan föreställning. Istället finns det en skiljelinje inom de konstruktivistiska perspektiven som är relaterad till huruvida kunskap ses som individuellt konstruerad (individkonstruktivism) eller socialt konstruerad (socialkonstruktivism), d.v.s. som ett resultat av gemensamma förhandlingar och därmed något som tar form mellan människor, i relation till den/de miljöer och sammanhang där kunskaperna brukas och till de redskap som finns.

I artikeln *Socio-cultural aspects of assessment* av Caroline Gipps (1999) fokuserar hon specifikt på kunskapsbedömning som social praktik, det vill säga att kunskapsbedömning handlar om relationen mellan den som bedömer och den/de som bedöms, men bedömning är också en fråga om politiska beslut som påverkar lärares bedömningsarbete. Hur kunskapsbedömningar genomförs – vad som bedöms och hur bedömningen går till - är relaterat till tidsmässiga, sociala och kulturella sammanhang (Moss m.fl., 2008; Murphy, 2008). Gipps lyfter också fram att bedömning handlar om makten att definiera vad som räknas som kunskap, vem som har tillgång till vad som räknas som kunskap och vad som karakteriserar de procedurer som ligger till grund för bedömningen:

> The didactic relationship between teacher and student is traditionally a hierarchical one and the assessment relationship one of judgment or surveillance. If we are serious about taking an interpretive approach and bringing the student into some ownership of the assessment process (and hence into self-evaluation), teachers must share power with students rather than exerting power over them (a.a. s. 386).

I huvudsak pekar Gipps på det orimliga i att endast den ena parten (läraren) är medveten om vad elevens prestationer jämförs med – vad som ligger till utgångspunkt för bedömning. Vidare pekar hon på det didaktiska uppdrag som ligger i sociokulturella perspektiv på undervisning, lärande och bedömning – att läraren har ansvaret för att eleverna under sin skoltid utvecklar en rimlig förståelse för innebörden i mål och betygskriterier inom de olika skolämnena. Detta ser hon som en grundläggande förutsättning för att elever ska få möjlighet att utveckla den kunskap de förväntas bli delaktiga i under sin skoltid. Hon avslutar artikeln med följande ord (s. 387):

> We need to bring out into the open the nature of the power relationship in teaching and assessment and point out the possibility of reconstructing this relationship. Perhaps most important, we need to encourage teachers to bring pupils into the process of assessment, in order to recognize their social and cultural background, and into self-assessment, in order to develop their evaluative and metacognitive skills. All of these acts are, on the basis of this review, both possible and necessary if assessment is to be more equitable and fulfil its promise to aid and support high-quality learning.

Med den här korta introduktionen till några teoretiska perspektiv på formativ bedömning avslutar vi med att de grundantaganden som vi bedömt som rimligast är de som förenar socialkonstruktivistiska och sociokulturella perspektiv på kunskap, lärande och bedömning. Av forskare krävs dock att man kan bedöma vetenskapliga arbeten på deras premisser, vilket kräver att kvalitet i forskning inte nödvändigtvis är relaterad till ett specifikt perspektiv. Istället är vårt uppdrag att granska studierna utifrån de grundantaganden som gjorts – förutsatt att de deklarerat för sina utgångspunkter. Enkelt uttryckt så finns det principiellt forskning inom varje perspektiv som håller hög kvalitet, liksom det finns forskning av sämre kvalitet inom alla perspektiv. Perspektivet avgör inte kvaliteten, det är istället en fråga om systematik i det metodiska arbetet och stringensen i argumentationen, liksom konsistensen i arbetet som helhet.

1.3 Studiens syfte och frågeställningar

Då vi påbörjade arbetet med översikten ombads vi formulera såväl ett övergripande syfte som några specificerade forskningsfrågor. Frågorna har på ett övergripande plan utformats i relation till de teoretiska grundantaganden vi delar. Den första frågan, att bedöma värdet av befintlig forskning i relation till elevers måluppfyllelse i svenska sammanhang, bygger på den kunskap som finns om svårigheterna med att flytta över modeller från ett sammanhang (t.ex. ett land) till ett annat (jfr Wermke, 2013). På liknande sätt kommer våra gemensamma teoretiska antaganden till uttryck i den del av den andra frågan som poängterar att vi även i arbetet granskar förutsättningarna för modellernas användning. Utifrån våra respektive tidigare arbeten formuleras även den tredje frågan – vi har i tidigare forskning konstaterat att det finns olika uppfattningar om innebörden i formativ bedömning och att uppfattningarna får betydelse för hur studierna utformas, vad man efterfrågar, fokuserar och hur man arbetar metodiskt liksom för vilka slutsatser som blir möjliga att dra. Utformningen av frågorna är således bunden till grundantaganden om att kunskap om formativ bedömning utvecklas i varierande forskningsmiljöer där det är vanligt att miljöer som samarbetar med varandra över tid utvecklar samsyn kring vilka frågor som är viktiga att studera och hur dessa kan studeras.

Syfte och frågeställningar presenteras nedan. Under arbetets gång och då resultatet vuxit fram, har vi dock kunnat se att varken internationell eller svensk forskning ger tillräckliga förutsättningar för att i tillräcklig utsträckning besvara samtliga frågor.

Det övergripande syftet är att systematiskt kartlägga och redovisa forskningsläget i Sverige och internationellt vad gäller formativ bedömning i grundskolan eller motsvarande. I kartläggningen ingår att inventera vilka teorier som ligger till grund för forskning om formativ bedömning samt vilka metoder som utprövats.

Ytterligare syften är att

- bedöma värdet av befintlig forskning om formativ bedömning i relation till elevers måluppfyllelse i svenska sammanhang, samt
- sammanställa vad som karakteriserar modeller för formativ bedömning – men även förutsättningarna för deras användning – för vilka det finns indikationer på att måluppfyllelsen ökar

Våra forskningsfrågor för översikten är följande:

- På vilka sätt har formativ bedömning uppfattats och använts i forskningssammanhang?
- Vilka skillnader och likheter förekommer i forskning om formativ bedömning a) mellan länder b) mellan olika skolämnen?
- Vad karakteriserar de modeller för formativ bedömning som förefaller leda till ökad måluppfyllelse? Under vilka villkor används dessa modeller och under vilka villkor är de framgångsrika?

2 ÖVERSIKTENS DESIGN OCH METODISKA ÖVERVÄGANDEN

Kapitlet inleds med en introduktion av översiktens design, nämligen vilka databaser och andra resurser vi använt oss av, men också vilka sökord vi använde och vad dessa strategier för litteratursökning resulterat i. Då materialet i ett första skede var väldigt omfattade, handlar följande aspekt av designprocessen om urval och avgränsning av texter för inkludering. Detta utgjorde det första skedet i designen av litteraturstudien. Utifrån det som utgjorde resultatet av det första skedet designades det andra skedet genom en värdering av vad vi fått i relation till dels våra kunskaper om fältet, dels texternas representativitet i relation till uppdraget att presentera internationell och nationell forskning. Slutet av kapitlet ägnas åt en översiktlig beskrivning av hur vi bearbetat och analyserat materialet.

2.1 Design: val av söktjänster och sökstrategier

För internationellt publicerad forskning har vi först valt internationella databaser och sedan sökt i var och en av dessa med hjälp av samma nyckelord och avgränsningar. Vidare har vi valt internationellt publicerade handböcker eller böcker av handbokskaraktär. För svensk forskning har vi använt oss av svenska söktjänster, där vi avgränsat oss till avhandlingar och licentiatuppsatser, samt tre svenska tidskrifter med peer review. Nedan redogör vi för hur vi gått tillväga vad gäller sökning, urval och avgränsning, bearbetning och analys av resultatet av våra sökningar.

Procedurer för sökning och urval av internationellt publicerad forskning

Tre databaser för internationellt publicerad forskning valdes på basis av våra respektive erfarenheter av att använda dem: *Academic Search Elite, ERIC* och *Pro Quest Social Science*. Vår erfarenhet är att det tar tid att utveckla förtrogenhet med hur en databas bäst används för att ge produktiva resultat. Även om likheterna i resultat är relativt stora förekommer det ändå skillnader som gör det intressant att pröva samma litteratursökning i flera databaser än en. Konstruktionen av databaserna – hur sökvägarna byggts upp, vilka tidskrifter som ingår och vilka årgångar som är representerade i var och en av dem, samt vilka databaser som är länkade till varandra – allt detta bidrar till vad man hittar i vilken databas, liksom vilka databaser som är tillgängliga för vilket lärosäte.[10] En första avgränsning gjordes i relation till tid. De artiklar som inkluderades i detta skede har publicerats under perioden 2000-01-01 – 2014-06-30.

De sökord vi använt är ”formative assessment” respektive ”assessment for learning” och vi specificerade sökningen till artiklarnas titlar, abstract och/eller nyckelord. Detta motiveras av att antalet träffar med något av sökorden var som helst i texten är ohanterligt stort. En översiktlig granskning visar att det som är objektet för litteraturöversikten kan vara perifert eller till och med endast förekomma i referenslistan. Genom att avgränsa placeringen till titel, abstract och/eller nyckelord säkerställer vi att fenomenet är centralt för artikeln. Ett

[10] Som exempel kan nämnas att Högskolan för lärande och kommunikation i Jönköping tillhandahåller versionen Academic Search Elite (ASE) medan Academic Search Premier (ASP) erbjuds vid Stockholms universitet. EBSCO host äger båda. Enligt hemsidan erbjuder båda ” Over 13,780 indexed and abstracted journals” och vad gäller ” peer-reviewed, indexed and abstracted journals” är skillnaderna marginella – drygt 12.000 för båda. De principiella skillnaderna finns i ” Full text for over 4,770 journals” (ASP) ”More than 2,360 full text journals” (ASE) och i relation till antalet ”peer-reviewed, full-text journals”: drygt 4000 för ASP medan ASE erbjuder drygt 1800. För ASE finns ytterligare informationen att det finns ”PDF-content dating as far back as1985”. Vi har gjort en översiktlig granskning av skillnaderna men de förefaller vara marginella. Det största skillnaderna uppstår i relation till vilka tidskrifter som ingår eller utelämnas och från vilket årtal olika tidskrifter har inkluderats i databaserna (citaten från http://www.ebscohost.com/academic.

alternativ hade varit att ytterligare snäva in sökningen till att gälla enbart titeln, men då hade vi missat flera centrala artiklar.

Ytterligare en avgränsning gjordes genom att i sökfälten markera att artiklar innehållande ”higher education” inte skulle inkluderas i resultatet. Denna avgränsning fungerade dock inte bra, då det resultat vi fick fortfarande omfattade ett flertal artiklar som trots det berörde högre utbildning. Dessa har därför tagits bort manuellt, med ett undantag. Under arbetets gång visade det sig att en central aspekt av det som efterfrågats innebär att även beakta hur olika lärarutbildningar hanterar formativ bedömning som kunskapsinnehåll för blivande lärare. Dessa har därför lyfts in tillbaka till de artiklar som ingår i resultatet och vi har läst artiklarna översiktligt. Vidare har vi i de internationella databaserna valt artiklar på engelska. Vi har även gjort några manuella kontrollsökningar direkt på några tidskrifters hemsidor för att bedöma rimligheten i de träffar vi fått fram med hjälp av sökmotorerna[11]. För valet av dessa tidskrifter hade vi följande urvalskriterier: de ska huvudsakligen adressera frågor om bedömning (*assessment*) och utvärdering (*evaluation*) eller undervisning. Utvärdering inkluderades dels för att tidskrifter som har fokus på utvärdering även publicerar artiklar om bedömning, dels för att de engelska begreppen *assessment* och *evaluation* ibland används synonymt även om det är vanligare att *evaluation* står för utvärdering och *assessment* för bedömning. Undervisning inkluderades av tidigare nämnd relation mellan undervisning och formativ bedömning. Tidskrifternas hemsidor administreras av det förlag som publicerar respektive tidskrift, exempelvis Sage eller Taylor and Francis Group (www.tandfonline.com). På varje tidskrifts hemsida kan man fylla i sökord och sedan kryssa för något av alternativen *Within current journal* eller *Entire site*. I och med att vi ville veta vad som fanns i en specifik tidskrift i relation till våra sökord valde vi det förstnämnda alternativet och de manuella sökningarna har följaktligen gjorts i samtliga artiklar under den aktuella perioden. Följande tidskrifter inkluderades i kontrollen:

- Assessment in Education – Principles, Policy, and Practice
- Applied Measurement in Education
- Educational Assessment
- Journal of Psychoeducational Assessment
- Studies in Educational Evaluation
- Teachers and Teaching, Theory and Practice

Fem av dessa ingår i tabell 2 över de tidskrifter i vilka vi funnit flest publicerade artiklar av relevans för vår översikt. Den som inte ingår är Journal of Psychoeducational Assessment.

Databaserna för internationellt publicerad forskning kompletterades med ett tre handböcker, *Handbook of formative assessment* (Andrade & Cizek, 2010), *Formative Assessment: Improving Learning in Secondary Classrooms* (CERI, 2005) och *Handbook of Research on Teaching* (4. uppl., Richardson 2001). Andrades och Cizeks handbok är explicit riktad mot formativ bedömning, och den gör anspråk på att representera internationell forskning även om kapitlen huvudsakligen är skrivna av forskare från USA. Rapporten av CERI består av två delar: en första del som har karaktär av handbok med Black och Wiliam som författare och en andra del som består av lägesöversikter avseende formativ bedömning i åtta länder (Australien/Queensland, Danmark, England, Finland, Italien, Kanada, Nya Zeeland och Skottland). Den tredje har ett bredare fokus och har inkluderats specifikt för att den berör forskning om undervisning där formativ bedömning vanligen anses utgöra en integrerad del av undervisningen, eller åtminstone ha en nära koppling till undervisning (jfr Gipps, 1999; Black & Wiliam, 2009;). Handboken ges ut av den amerikanska forskningsorganisationen AERA och uppges ha ambitionen att belysa samtida forskningstrender, perspektiv, frågor och metoder. Den fjärde

[11] Syftet med den kontrollen var att säkerställa att vi fått med centrala artiklar och författare via litteratursökningarna i databaserna, vilket motiveras av de erfarenheter Forsberg och Lindberg (2010) gjorde i en tidigare svensk översikt. De fann skillnader i det resultat som sökningarna via söktjänsterna gav jämfört med tidskrifternas hemsidor. Den mest markanta skillnaden gällde *Scandinavian Journal of Educational Research*, där de fann 54 träffar på tidskriftens hemsida som inte förekom i resultaten från söktjänsterna. En detaljgranskning av informationen i söktjänsterna visade att skillnaderna främst kunde förklaras av vilka årtal tidskriften inkluderats i respektive söktjänst.

upplagan, som publicerats inom det tidsspann denna översikt gäller, består av åtta delar. I den sista delen, som samlat artiklar under temat *Instruction* finns ett kapitel, *The Role of Classroom Assessment in Teaching and Learning* (Shepard 2001[12]), som är direkt relaterat till bedömning för lärande.

Utöver det har vi som referensram använt några böcker av internationellt erkända forskare inom området, dock huvudsakligen skrivna på engelska och därmed främst representerande Australien, England, USA och Nya Zeeland[13].

Procedurer för sökning och urval av forskning publicerad på svenska

Forskning publicerad på svenska, har sökts via söktjänsterna LIBRIS, SwePub, samt Skolporten. Utöver det har vi sökt manuellt via hemsidorna för tre svenska tidskrifter med peer review. Det finns egentligen ingen riktigt bra söktjänst för svensk forskning, vilket är skälet till att vi använt oss av flera söktjänster för att få en rimlig överblick. I den översikt som Forsberg och Lindberg (2010) genomförde avseende svensk forskning om bedömning och betyg konstaterade de att det var nödvändigt att känna till forskningsfältet för att veta vad som måste finnas i resultatet av litteratursökningarna för att översikten skulle kunna vara representativ. Vi har i det här arbetet gjort motsvarande erfarenheter. Den ovan nämnda översikten utgjorde en grund för vårt arbete, vilket avgränsade det till en komplettering av det som publicerats från och med den senare delen av 2009. Vi har under arbetets gång konstaterat att sökresultaten inte är samstämmiga och framför allt att vi trots tre söktjänster[14] inte kommit åt all den forskning vi genom vår förtrogenhet med forskningsfältet vet finns publicerad. Att utelämna det som inte kommer fram genom litteratursökningarna har vi inte sett som ett alternativ – vårt uppdrag är ju att ge en överblick av forskningsläget. Av tabell 3/kapitel 3, framgår vilka studier som ingår i översikten av svensk forskning. Vi återkommer till detta i kapitel 4 med en metoddiskussion. I det följande beskriver vi hur vi gått tillväga. Vi inleder med en beskrivning av söktjänsterna, sedan beskrivs urvalet av tidskrifter och hur litteratursökningarna på deras hemsidor genomförts.

Sökvägarna i de tre använda söktjänsterna skiljer sig från de internationella genom att de alternativ som erbjuds inte är desamma, framför allt går det inte att söka i abstract. Gemensamt för LIBRIS och SwePub är att man kan välja mellan enkel och utökad sökning. För Skolporten finns bara ett alternativ. I samtliga fall har vi använt sökord på svenska som motsvarade de engelska för internationellt publicerad forskning, nämligen ”formativ bedömning” och ”bedömning för lärande”.

LIBRIS är en nationell söktjänst, gemensam för de svenska universiteten och högskolorna (www.libris.kb.se). Vi använde oss av alternativet utökad sökning och valde *fritextsökning* i kombination med *materialvalet* avhandling.[15] Valet av fritextsökning uppges hitta alla poster där sökordet/sökorden förekommer i något av de sökfält som finns tillgängliga. För vår del var det följande fält som var av intresse: titel (titlar om ett ämne), ämnesord, klassifikation och ämneskategori.

SwePub upprätthålls av Kungliga biblioteket och innehåller referenser till de forskningspublikationer (artiklar, avhandlingar, konferensbidrag och populärvetenskapliga texter[16]) som finns registrerade i ett trettiotal svenska lärosätens publiceringsdatabaser – dock varierar den tidsperiod för vilken forskningspublikationerna ingår, så att de publikationer från exempelvis Uppsala universitet (inkl. Gotlands högskola) som tills vidare införts endast är de som publicerats från och med 2008, från Karlstad universitet från och med 2006, och från

---

[12] Kapitlet är i huvudsak detsamma som den tidigare nämnda artikeln av Shepard (2000).

[13] Dessa böcker är Assessment Reform in education: Policy and Practice (Berry & Adamson, 2011), Assessment and Learning, 2nd edition (Gardner m fl, 2012) och Assessment for Education Standards, Judgement and Moderation (Klenowski & Wyatt-Smith 2013).

[14] Vi har även gjort en kontrollsökning i söktjänsten Digitala Vetenskapliga Arkivet (DiVA), en gemensam söktjänst för forskningspublikationer och studentuppsatser från 35 (eller 36 – båda uppges på hemsidan) lärosäten och forskningsinstitutioner (www.diva-portal.org). Inte heller denna söktjänst ger hela det resultat som denna översikt omfattar.

[15] Tills vidare fångar detta materialval i Libris såväl doktorsavhandlingar som licentiatuppsatser. Det tidigare använda begreppet licentiat*avhandling* benämns numera enligt Högskoleförordningen (SFS 1993/100, Bilaga 2) vetenskaplig uppsats (i Sverige), medan begreppet avhandling är förbehållet doktorsavhandlingar.

[16] Endast avhandlingar och artiklar har inkluderats.

Göteborgs universitet från och med 2004 (www.swepub.se). Den här databasen har använts för artiklar, avhandlingar och licentiatuppsatser. Söktekniken är uppbyggd av LIBRIS och följer deras kategoriseringar (ovan) men inte fullt ut deras sökalternativ. Även här valde vi utökad sökning som utgångspunkt med fritextsökning men utan att specificera publikationstyp. Istället gallrade vi i resultatet.

Skolporten AB är ett företag som, bland annat, bedriver informationsförmedling via internet. En del av informationstjänsterna berör forsknings- och utvecklingsarbete (FoU). Det finns till exempel en söktjänst avseende avhandlingar som berör skolan (www.skolporten.se). På basis av de erfarenheter Forsberg och Lindberg (2010) gjorde – att denna söktjänst gav ett bättre resultat i termer av fler träffar av relevans för deras översikt – har vi valt att inkludera Skolportens söktjänst även för denna översikt.

Litteratursökningen i de svenska söktjänsterna skiljer sig från dem i de internationella genom att sökalternativen är konstruerade utifrån andra kriterier. Framför allt går det inte att söka direkt i abstract via de svenska söktjänsterna. Alternativet fritextsökning i LIBRIS och SwePub ger enligt uppgifter om dessa söktjänster möjlighet att hitta samtliga poster i vissa fördefinierade fält, t.ex. ämnesord och genrebeteckningar. Resultaten gällande avhandlingar skiljer sig dock åt, och dessutom konstaterades att vi inte ens då vi slog ihop resultatet av sökningarna från de tre söktjänsterna kom åt den forskning vi hade kännedom om (jfr tabell 3). Vad gäller de tre tidskrifter vi sökt i kan vi konstatera att de flesta svenska artiklarna vi funnit inte ingick i SwePub (vilket kunde ha varit fallet) och bland dem som publicerats på engelska fanns några, men inte alla, med i det internationella sökresultatet.

Tre svenska tidskrifter, Pedagogisk forskning i Sverige (PFiS), Utbildning och demokrati. Tidskrift för didaktik och utbildningspolitik (U&D), samt Educare - Vetenskapliga skrifter har inkluderats. De tre tidskrifterna representerar en bredd i svensk forskning med potentiellt olika ingångar till frågor om lärares bedömningsarbete. Den förstnämnda, PFiS, är en pedagogisk tidskrift som grundades 1996. Numera drivs tidskriften av föreningen Svensk Förening för Pedagogisk Forskning SFPF/SWERA (http://pedagogiskforskning.se/om/). Utbildning och demokrati är en allmändidaktisk tidskrift som funnits sedan 1992 och ges ut vid Örebro universitet (http://www.utbildningochdemokrati.se/ud/). Educare – Vetenskapliga skrifter ges ut vid Fakulteten för lärande och samhälle på Malmö högskola och har som ambition att fungera som nationellt och nordiskt forum för forskning inom utbildningsvetenskap. Skriftserien är sakkunniggranskad och har getts ut sedan hösten 2005 (http://dspace.mah.se/handle/2043/8142). I detta fall har vi sökt manuellt på tidskrifternas hemsidor (PfiS och U&D) med sökorden ”formativ bedömning” och ”bedömning för lärande”. För Educare gäller att vi har sökt först på hemsidan utifrån artiklarnas titlar och sedan i respektive dokument (artiklarna är åtkomliga i elektronisk form).

Nordiska utblickar

Av nordisk forskning relaterad till formativ bedömning är främst norska och svenska studier publicerade i internationella sammanhang och utgör därmed en del av det internationella. Men liksom svensk forskning publiceras på svenska så publiceras även nordiska studier på danska, finska, isländska och norska. När det gäller dansk forskning om bedömning är den dels volymmässigt mindre än vad som är fallet med svensk och finsk forskning (jfr Forsberg & Lindberg, 2010), dels har den främst varit relaterad till andra delar av utbildningssystemet (yrkesutbildning, högre utbildning) än grundskolan. Det finns dock en dansk avhandling av Kristine Kousholt (2009) som handlar om den danska grundskolans (*folkeskolen*) bedömningspraktiker, som inkluderas i översikten till de delar den berör formativa aspekter av bedömningspraktikerna. Den finska forskningen representeras genom en studie om bedömningens etiska aspekter av Päivi Atjonen (2007). Merparten av norsk forskning om bedömning gäller högre utbildning och flera av studierna handlar om portfolio som redskap för bedömning. Kari Smith har dock, med olika medförfattare, skrivit fyra artiklar publicerade i internationella tidskrifter, två av dem ingår i resultatet från litteratursökningen och ytterligare ett kapitel i en antologi har inkluderats i översikten.

2.2 Bearbetning och analys

Vad gäller internationell forskning började vi med att läsa igenom abstracten för de ca 2000 artiklar som ingick i resultatet av den litteratursökningen. I relation till denna genomgång exkluderade vi alla de artiklar som gällde bedömning i högre utbildning, gymnasieskola och vuxenutbildning. Det visade sig att det som återstod var 340 artiklar. Av dem som exkluderades gällde merparten högre utbildning. Följande steg i proceduren handlade om att se vad det var för slags resultat vi hade fått i förhållande till de internationella tidskrifter som rankas högt med avseende på *impact factor* inom ISI Web of Science (WoS) (Social Sciences) och konstaterade att om vi enbart hade inkluderat artiklar enligt detta rankingsystem så hade artiklar publicerade i tidskriften *Computers and Education* framstått som de som har störst inflytande på bedömningsområdet[17], medan en tidskrift som har mycket högt anseende inom forskningsfältet, *Assessment in Education, Principles, Policy, and Practice* (utgiven av organisationen International Association for Educational Assessment[18], IAEA) inte hade kommit med överhuvudtaget, då den inte listas i WoS. Då artiklarna inom tidskriften Computers and Education innehållsligt domineras av dels tekniska lösningar, dels av en instrumentell förståelse av fenomenet formativ bedömning/bedömning för lärande gjorde vi bedömningen att WoS som kvalitetskriterium för vår översikt förfelar sitt syfte.

Som inkluderingskriterier har vi istället nöjt oss med att artiklarna ska vara skrivna på engelska, de ska beröra grundskolekontext eller motsvarande och lärarutbildning. Våra språkliga begränsningar gör att forskning som publicerats på andra språk (t.ex. spanska och franska) inte ingår, däremot finns det en del studier som representerar delar av spansk- och fransktalande länder. Av tabell 1 i kapitel 3 framgår att studier från dessa länder är svagt representerade.

Det främsta inkluderingskriteriet för svensk forskning är att studierna ska gälla grundskolan och materialet består av avhandlingar, licentiatuppsatser och artiklar i de tidigare nämnda tidskrifterna.

Bearbetningen för såväl internationell som svensk forskning har följt samma principer – texterna har tematiserats utifrån sitt huvudsakliga forskningsobjekt och syfte. För ett fåtal av de svenska studierna har vi frångått principen i de fall avhandlingen har bedömts ge ett substantiellt bidrag till förståelsen av fenomenet formativ bedömning/bedömning för lärande trots att forskningsobjektet är ett annat. I dessa fall är det empiriska materialet relaterat till formativ bedömning, vilket är skälet till att studien inkluderats. Vidare förekommer det i några fall att en studie bidrar med kunskap till två teman. I de fall vi sett det som nödvändigt återfinns studien i båda de teman som är aktuella.

[17] Givet att man följer logiken att de artiklar som citeras mest i tidskrifter listade i WoS är de som i praktiken också har störst impact på ett område.

[18] IAEA grundades 1975 och har till syfte att stödja utbildningsorganisationer i utvecklingen och tillämpningen av pedagogiska bedömningstekniker för att främja utbildningens kvalitet (http://www.iaea.info/).

3 RESULTAT

I det här kapitlet presenterar vi resultaten av de artiklar, avhandlingar och licentiatuppsatser som ligger till grund för översikten. I vissa fall stödjer vi oss också på texter från de handböcker och antologier av handbokskaraktär som tidigare nämnts. Vi har valt att inleda med de internationella studierna för att resultatet av den översikten ska finnas som en bakgrund för läsaren vid ingången till de svenska studierna. Innan kapitlet avslutas gör vi en kort utblick i de nordiska studierna.

3.1 Formativ bedömning i internationell forskning

Vid genomgång av de 340 artiklar som svarade mot våra inkluderingskriterier gjordes en initial kategorisering i fem huvudsakliga kategorier med avseende på artiklarnas innehåll. Dessa fem kategorier är *teori/begreppsutveckling*, *policy*, *verktyg*, *implementering/professionellt lärande* och *elevers uppfattningar/upplevelser.* Var och en av kategorierna är innehållsmässigt bred, och kommer nedan att redovisas under olika underrubriker. I viss mån är kategorierna också överlappande, exempelvis då aspekter som behandlas under teori/begreppsutveckling också är viktiga aspekter av policy, då aspekter av implementering/professionellt lärande utgör en central fråga för policy, eller när elevers uppfattningar rör specifika verktyg.

Resultatdelen inleds med den generella överblick över fältet som vi fick när vi gjorde sökningar och tillämpade våra inkluderingskriterier. Detta kan betraktas som ett indirekt resultat, som också säger något om forskningsfältet formativ bedömning. Resultatkapitlet övergå därefter till den kartläggning som har gjorts utifrån inkluderade artiklar, och presenteras kategori för kategori. Det svenska forskningsläget presenteras som en separat del i slutet av resultatkapitlet.

Generell överblick

I den databassökning som initialt gav drygt 2000 träffar motsvarade 340 våra inkluderingskriterier. Studierna kommer från 29 olika länder världen över och fördelar sig mellan världsdelar och länder enligt tabell 1:

Tabell 1. Fördelning av inkluderade internationella artiklar över världsdelar och länder.

Europa	**Nordamerika**	**Oceanien**	**Asien**	**Afrika**
168	102	35	34	1
Storbritannien, 110	USA, 88	Nya Zeeland, 25	Hongkong, 12	Egypten, 1
Nederländerna, 22	Canada, 14	Australien, 10	Taiwan, 8	
Norge, 8			Israel, 4	
Sverige, 4			Singapore, 4	
Portugal, 4			Jordanien, 2	
Belgien, 3			Kina, 1	
Irland, 3			Malaysia, 1	
Tyskland, 3				
Cypern, 2				
Finland, 2				
Spanien, 2				
Grekland, 1				
Makedonien, 1				
Schweiz, 1				
Tjeckien, 1				
Österrike, 1				

Bilden ger vid handen att studier från Storbritannien och USA är de som dominerar, följt av Nya Zeeland/Australien. I sammanhanget är det viktigt att tänka på att vi endast har inkluderat artiklar som publicerats på engelska. Det är rimligt att anta att det i varje nationell kontext finns ytterligare forskning som publicerats på det nationella språket. Noterbart i tabell 1 (ovan) är att vi inte har några artiklar alls från Sydamerika, vilket rimligen kan antas bero på att man där ofta publicerar sig på spanska. Då vi efter de initiala sökningarna tillämpade våra inkluderingskriterier kan konstateras att de artiklar vi rensade bort på grund av att de var på annat språk än engelska i hög utsträckning var just på spanska.

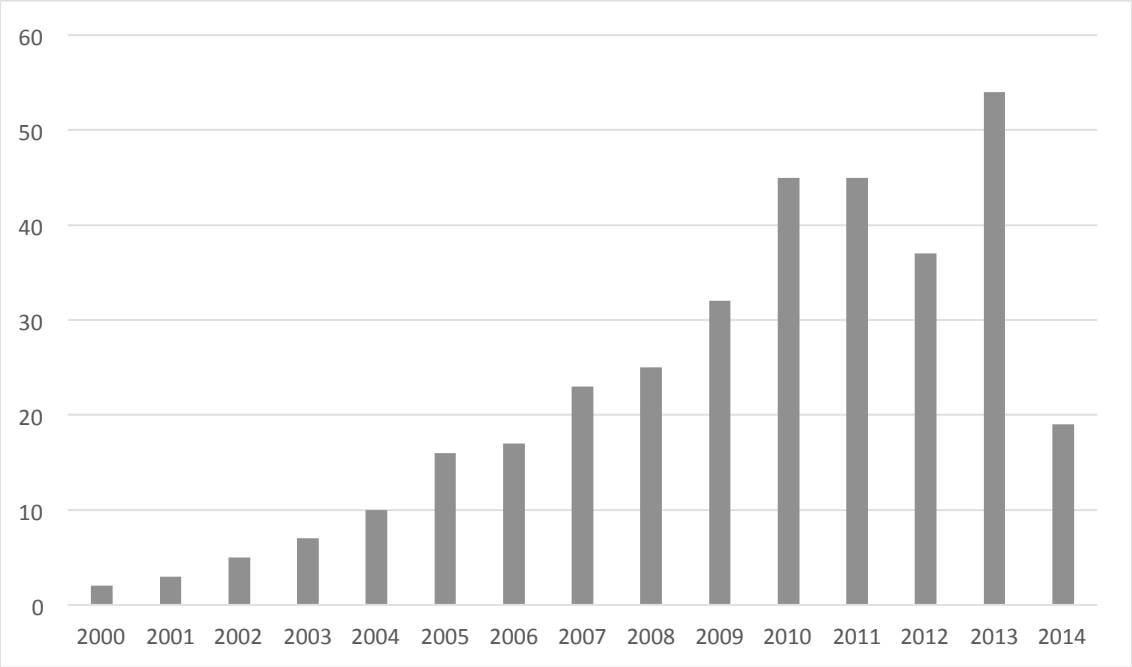
En annan bild av fältet får man av tabell 2, där det framgår i vilka tidskrifter artiklarna publicerats (av utrymmesskäl utelämnas här de tidskrifter där antalet artiklar som publicerats understiger sex):

Tabell 2. Antal artiklar i tidskrifter

Tidskrift	**Antal artiklar**
Assessment in Education: Principles, Policy & Practice	62
Studies in Educational Evaluation	28
The Curriculum Journal	22
Computers and Education	18
Theory into Practice	12
Teaching and Teacher Education	11
Applied Measurement in Education	10
Educational Measurement	10
Educational Research	9
British Educational Research Journal	8
Procedia: Social and Behavioural Sciences	7
Research papers in Education	7
Teachers and Teaching	7
International Journal of Science Education	6
Learning and Instruction	6

De tre största tidskrifterna är alltså *Assessment in Education: Principles, Policy & Practice*, *Studies in Educational Evaluation* och *The Curriculum Journal*. Ingen av dessa tre listas i Web of Science, men noterbart är att de namn som vi (utifrån vår förkunskap kring fältet) vet är stora namn inom bedömningsforskning publicerar sig i hög grad i just dessa tidskrifter. Den tidskrift som har högst impact factor av de ovan listade är Computers in Education. Påtagligt i de artiklar kring formativ bedömning som publicerats i Computers in Education är att de i hög grad handlar om tekniska lösningar för formativt arbete, och att författarna till dessa artiklar sällan förekommer i andra bedömningsrelaterade artiklar, antologier eller handböcker. Nämnas bör också att de tidskrifter där antalet publicerade artiklar i vår sökning understiger sex i hög grad är ämnesdidaktiska.

Våra sökningar har gällt åren 2000-2014, och som framgår av diagrammet nedan har antalet publicerade artiklar rörande formativ bedömning/bedömning för lärande successivt ökat:



Figur 1. Antal publicerade artiklar per år

Efter denna korta generella överblick övergår resultatpresentationen nu till våra huvudsakliga kategorier. Som tidigare nämnts är det i viss mån svårt att dra absoluta gränser mellan dem, då de delvis går in i varandra och handlar om aspekter som skulle kunna passa in på flera ställen. Indelningen är gjord efter det vi uppfattar vara respektive artikels *huvudsakliga syfte*.

Teoretisk förståelse av formativ bedömning/begreppsutveckling

Denna kategori tar sin huvudsakliga utgångspunkt i den teoretiska förståelsen av begreppet *formativ bedömning*, hur dess innebörd har definierats, samt hur detta har format empirisk forskning.

Ett stort antal artiklar har som sitt huvudsakliga syfte att teoretiskt positionera/utveckla begreppet formativ bedömning. Dessa artiklar kommer i huvudsak från den brittiska eller australiensiska/nya zeeländska kontexten, där man generellt har en längre tradition av såväl empiriska studier som teoribildning i relation till formativ bedömning.

Att närma sig forskningsfältet formativ bedömning utan att nämna Paul Black och Dylan Wiliam är omöjligt. Genomgående under den tidsperiod som ligger till grund för vår översikt kan man konstatera att de alltsedan de skrev översikten *Assessment and Classroom Learning* (1998) varit stora, tongivande och vida citerade namn internationellt. Av de 340 artiklar vi översiktligt eller mer djupgående har gått igenom är det ytterst få som inte på ett eller annat sätt förhåller sig till denna översikt och/eller någon av alla de artiklar som i olika former – men på samma tema – publicerats av samma författare sedan dess. Även i Sverige har deras definitioner och modeller i hög utsträckning kommit att dominera förståelsen av formativ bedömning, inte minst bland verksamma lärare och skolledare, då bland annat Skolverket kommit ut med skrifter baserade på dessa i samband med implementeringen av senaste läroplanen Lgr 11 (exempelvis Kunskapsbedömning i skolan, 2011; Kunskapsbedömning – vad, hur och varför, 2011).

Som tidigare nämnts listas inte vår översikts tre mest frekvent förekommande tidskrifter i Web of Science. Det är dock relevant att särskilt lyfta fram innehållet i vissa av de artiklar som är listade och mest citerade i WoS, då de utifrån logiken om impact factor *kan* antas ha stor påverkan på fältet. Högst i antalet citeringar ligger just en teoriutvecklande artikel: *Developing the theory of formative assessment* (Black & Wiliam, 2009)

som publicerats i *Educational Assessment, Evaluation and Accountability*. Artikeln är i första hand ett försök att teoretiskt definiera vad formativ bedömning är, men också ett försök att ge grunder utifrån teorier om lärande. Författarna hävdar att deras översikt från 1998 i första hand syftade till att dra samman forskningsresultat med relevans för begreppet formativ bedömning. Syftet med artikeln 2009 är att utveckla en teori för formativ bedömning bortom det som tidigare gjorts. Utgångspunkterna för teoretiserandet tas i praktiknära klassrumsforskning som har lett fram till definitionen av fem huvudsakliga aktivitetstyper som kan sägas vara effektiva för lärande (Black et al, 2003; Wiliam & Leahy, 2007):

- Dela framgångskriterier med eleverna
- Klassrumsfrågor
- Kommentarer på elevprestationer
- Kamrat- och självbedömning
- Formativ användning av summativa prov

I en annan vida citerad artikel definierar Wiliam och Thompson (2007) undervisning och lärande utifrån Ramaprasads (1983) tre nyckelprocesser:

- Fastställandet av var den lärande befinner sig just nu
- Fastställandet av vart den lärande är på väg
- Fastställandet av hur gapet mellan nuläge och mål ska överbryggas.

Ett grundantagande är att läraren inte är den enda aktören inblandad i dessa nyckelprocesser; eleverna är/bör vara delaktiga. Nyckelprocesserna kan därmed betraktas som involverande tre olika nivåer av aktörer (lärare, kamrat, elev), och därmed kan formativ bedömning enligt Black och Wiliam (2009) konceptualiseras utifrån figur 2:

<table>
<tr><th>Aktörer</th><th>Vart den lärande är på väg</th><th>Var den lärande befinner sig just nu</th><th>Hur gapet mellan nuläge och mål ska överbryggas</th></tr>
<tr><td>Lärare</td><td rowspan="3">Förstå och dela mål och framgångskriterier för lärande
1</td><td>Skapa effektiva klassrumsdiskussioner och andra situationer som ger bevis för lärande
2</td><td>Ge återkoppling som tar den lärande framåt
3</td></tr>
<tr><td>Kamrat</td><td colspan="2">Aktivera elever som resurser för varandra
4</td></tr>
<tr><td>Elev</td><td colspan="2">Aktivera elever som ägare av sitt eget lärande
5</td></tr>
</table>

Figur 2. Nyckelprocesser i formativ bedömning (efter Black & Wiliam 2009, s. 8 – vår översättning)

De aktivitetstyper som tidigare nämnts kan enligt Black och Wiliam (2009) betraktas som medel för att arbeta med de fem nyckelstrategierna i figur 2 ovan, där exempelvis klassrumsfrågor är *ett* sätt att arbeta med strategi 2 och kommentarer på elevprestationer är *ett* sätt att arbeta med strategi 3.

Även om Black och Wiliam erkänner att kontextuella aspekter (såsom skolämnesmässiga eller åldersmässiga aspekter) spelar roll för användandet av de formativa nyckelstrategierna menar de samtidigt att de fem strategierna är tillräckligt generella för att kunna vara den lins genom vilken praktiker såväl som forskare betraktar vad som innefattas i termen formativ bedömning. Att nyckelstrategierna blivit en sådan lins bekräftas i vår genomgång av empiriska studier, vilket vi återvänder till när vi diskuterar relationen mellan formativ och summativ bedömning nedan. Under rubriken *Verktyg* återvänder vi också till frågan kring huruvida man kan dra generella slutsatser oavsett kontextuella villkor. Helt klart är i alla fall att kontextberoende kontra generellt är en fråga som genomgående följer teoretiserandet kring formativ bedömning under den tidsperiod vi har studerat.

Wiliam (2009) ser stor potential, inte minst för praktiker, i den tydlighet som modellen över nyckelstrategier förmedlar, då man utifrån modellen kan förse lärare med forskningsgrundade och handfasta metoder/tekniker för att arbeta formativt i klassrummet. Han menar, vidare, att lärare utifrån sådana metoder/tekniker kan agera sig fram till en teoretiskt fördjupad förståelse av varför de gör som de gör, enligt logiken ”it is generally easier to get people to act their way into a new way of thinking than it is to get people to think their way into a new way of acting” (s 28-29). Han hävdar också vikten av att processen med att agera sig till ett nytt sätt att tänka behöver stödjas av kollegialt lärande i så kallade *Teacher Learning Communities* (TLC). Detta återvänder vi till under rubriken *Implementering/Professionellt lärande*

Pryor och Crossouard (2008) hävdar på basis av mångårig empirisk forskning på olika nivåer i utbildningssystemet att formativ bedömning ofta har skrivits fram på ett teoretisk förenklat sätt, och att fenomenet måste förstås som en diskursiv social praktik som involverar dialektiska och ibland konfliktfyllda processer, där de resultat som blir alltid är beroende av den sociala kontext i vilken de produceras. De menar att undervisning, oavsett skolnivå, fortfarande i hög grad utgår från en utbildningsmodell som bygger på kunskapsöverföring och kunskapsförvärv, där formativ bedömning i hög utsträckning förstås som ett instrumentellt komplement eller en ’quick-fix’ för att lösa vissa utbildningsproblem. Pryor och Crossouard föreslår därför en teoretisk förståelse av formativa bedömningsprocesser utifrån verksamhetsteoretiska perspektiv, där den kontextuella basen av Engeströms (2001) triangulära verksamhetssystem – regler, gemenskap och arbetsdelning – positionerar formativa processer som situerade.

I andra upplagan av *Assessment and Learning* (Gardner, 2012) skiljer sig de två bidrag i antologin som står för teoretiseringen av formativ bedömning (*Assessment in Harmony with our Understanding of Learning* av Mary James samt *Developing a Theory of Formative Assessment* av Paul Black & Dylan Wiliam) från samma författares texter i upplagan från 2006 så till vida att båda har orienterat sig mer mot socio-kulturella perspektiv på lärande. James menar att behavioristisk teoribildning ligger till grund för en första generation bedömningsförståelse (1G), kognitivistisk/konstruktivistisk för en andra generation (2G) och socio-kulturell/situerad för en tredje generation (3G). Hon hävdar att teoretiseringen av formativ bedömning till stor del har legat i 2G, men att den under senare år i allt högre utsträckning rört sig i riktning mot 3G. Hon menar vidare att 3G-bedömning innefattar mycket av 2G-bedömning, men att förståelsen av bedömning och synen på sammanhang i vilka man kan göra bedömningar av någons kunskap är vidare i 3G. Elwood (2006), liksom Crossouard (2012) för ett liknande resonemang, men menar att det är *nödvändigt* att förstå bedömning (som ständig följeslagare till lärande) som en kulturell aktivitet. Med en sådan förståelse erkänner man att både lärare och elever har med sig såväl kulturella/historiska erfarenheter som normativa förväntningar in i klassrummet, och att detta påverkar de bedömningar som görs i summativa såväl som formativa syften. Båda är i sammanhanget också kritiska till att forskningen kring formativa bedömning i så låg grad problematiserar aspekter som genus, etnicitet och social klass (jfr också Kirton m fl, 2007).

James (2012) problematiserar också den linjära steg-för-steg-förståelse som ofta förenklat kommer till uttryck när vi talar om formativ bedömning, och menar att progression i stället behöver förstås i analogi med Vygotskijs begrepp ZPD (den proximala utvecklingszonen). Själva ordet zon indikerar det icke-linjära i barns utveckling: i ZPD kan barnets förståelse till synes flytta framåt eller bakåt, åt vänster eller åt höger samtidigt som den expanderar och fördjupas (jfr Grigorenko & Sternberg, 1998). Inom så kallad *dynamisk bedömning* (som kan betraktas som antingen släkt med eller del av begreppet formativ bedömning) praktiseras bedömning av ’assisted performance’, det vill säga det som kan åstadkommas med hjälp av andra. Dynamisk bedömning praktiseras oftast i en-till-en-situationer med elever i svårigheter, eller genom uppgifter där så kallade

ledtrådsstrukturer används. James menar vidare att eftersom 3G-bedömning förstås som situerad bör det rimligen få till följd att den görs av dem som är nära eleverna snarare än av externa bedömare, att den lägger lika stor vikt vid grupplärande som individuellt lärande, att den dokumenteras med hjälp av exempelvis olika former av inspelningar och/eller portfolio och att den görs utifrån ett holistiskt perspektiv. Det främsta problemet med detta, enligt James, är att sådan bedömning inte genererar de absoluta mått på lärande och prestationer som det omgivande samhället vill ha. Även om sociokulturella metoder i flera avseenden kan göra anspråk på större validitet, har de långt kvar när det gäller att tillhandahålla övertygande garantier om reliabilitet i bedömningsresultat. Klenowski (t.ex. 2011) menar att frågan om balansen mellan storskaliga tester och autentiska lärarbedömningar är mer aktuell än någonsin. Det handlar om likvärdighetsfrågor som inte enbart gäller test eller bedömning i teknisk mening, utan snarare ifall testsystemeten som sådana tar hänsyn till sociokulturella representationer av prestation. Med utgångspunkt i sina studier hävdar James att socio-kulturella perspektiv på lärande och bedömning tilltalar många lärare, och i formativa syften - särskilt i lägre åldrar där standardiserade tester eller betyg inte finns– är det lättare att tankegångarna får fäste och genomsyrar praktiken.

Birenbaum, Kimron och Shilton (2011) har provat att empiriskt undersöka relationen mellan formativt klassrumsarbete och den kontext som arbetet är inbäddad i på såväl klassrumsnivå som skolnivå. Studien involverade i första steget en kvantitativ del, där 122 lärare på skolans alla stadier fick svara på en rad olika frågor kring hur man definierar vad som innefattas i begreppet formativ bedömning, samt hur man arbetar med formativ bedömning i klassrummet. Sex lärare deltog sedan i den kvalitativa studien, och de utgör de huvudsakliga informanterna. Med dessa sex lärare genomfördes materialbaserade intervjuer, där medtagna artefakter som planeringsdokument, tester, bedömningsmatriser, exempel på elevarbeten, exempel på skriven feedback från lärare till elev och från elev till elev, var en viktig utgångspunkt i intervjuerna. En kollega till var och en av de huvudsakliga informanterna intervjuades också, i syfte att få ytterligare en bild av skolans bedömningskultur. Dessutom fokusgruppintervjuades totalt 22 elever från de sex lärarnas klasser (i grupper om 2-6 deltagare) i syfte att få elevperspektiv på bedömningskulturen i de olika klassrummen. Utifrån analys av alla data sammantaget kunde de sex huvudinformanterna sorteras in på ett kontinuum (graderat 1-5) med avseende på åtta olika aspekter: formativ bedömningspraktik (låg kvalitet – hög kvalitet), epistemologi (positivistisk – konstruktivistisk), socialt klassrumsklimat (negativt – positivt), elevmotivation (prestationsinriktad – lärandeinriktad), struktur (centraliserad – decentraliserad), reglering (extern – intern), lärarmotivation (låg – hög) och professionellt lärande (ytligt – djupt). Två ”extremfall” kunde konstateras bland de sex lärarna, där en genomgående hamnade på kontinuumets första steg (1) och en genomgående hamnade på kontinuumets två sista steg (4-5). Dessa ”extremfall” diskuteras som exempel på hur lärares kunskapssyn, yttre strukturer och reglering samt former för professionellt lärande tycks sammanfalla med kvaliteten på den formativa bedömningspraktiken. Det sociala klassrumsklimatet samt vilken motivation som driver eleverna förefaller dessutom vara faktorer som samspelar med de övriga. I ett flertal andra studier som gåtts igenom diskuteras relationen mellan yttre strukturer och reglering å ena sidan och professionellt lärande å andra sidan. Detta återkommer vi till längre fram.

Relationen mellan formativ och summativ bedömning

En central fråga i genomgångna artiklar handlar om huruvida det är relevant att upprätthålla distinktionen mellan summativ och formativ bedömning. Harlen (2012) liksom Cowie och Bell (1999) hävdar att distinktionen behöver vidmakthållas när det gäller förståelsen av bedömningars olika syften, men att relationen snarare behöver förstås som ett kontinuum än en dikotomi enligt figur 3:

Formativ ----------> Summativ				
	Informell formativ	**Formell formativ**	**Informell summativ**	**Formell summativ**
Huvudsakligt fokus	Vilket är nästa steg i lärandet?		Vad har åstadkommits så här långt?	
Syfte	Att se/förstå nästa steg i lärandet	Att se/förstå nästa steg i lärandet	Att övervaka framsteg gentemot mål	Att fastställa individers nivå
Hur bevis samlas in	Som ständigt pågående del av vanligt klassrumsarbete	Vid vissa tillfällen och kopplat till vanligt klassrumsarbete	Vid vissa tillfällen och kopplat till vanligt klassrumsarbete	Separata tester
Grund för bedömning	Individ- och kriterierelaterat	Individ- och kriterierelaterat	Kriterierelaterat	Kriterierelaterat
Bedöms av	Elev och lärare	Lärare	Lärare	Lärare eller extern bedömare
Vidtagna åtgärder	Återkoppling till elev och lärare	Återkoppling till lärarens planering / undervisningen	Återkoppling till lärarens planering / undervisningen	Rapportering till elev, föräldrar, egen lärare, andra lärare, skolledning etc
Benämning	Bedömning för lärande	Matchning	Mätsticka	Bedömning av lärande

Figur 3. Relationen formativ – summativ bedömning (efter Harlen 2012, s. 98, vår översättning)

Harlen (2012) menar vidare att vi, snarare än att dikotomisera, behöver arbeta för synergi mellan bedömningsformers olika syften. Samma bedömningsprocess kan användas i båda syftena, vilket hon exemplifierar med studier från Queensland, där man exempelvis använder portfoliometoden både i syfte att ge elever återkoppling underhand, men också i syfte att bedöma övergripande prestationer vid vissa givna tillfällen. För att öka tillförlitligheten i lärares summativa bedömningar krävs moderering (bedömningssamarbete), samt träning för lärare i att planera för bedömning och utveckla bedömningskriterier. Queensland, Australien är ett exempel som ofta lyfts fram då det gäller att systematisera lärares gemensamma bedömningsplanering och moderering (Klenowski, 2011; Klenowski & Wyatt-Smith, 2014).

Crooks (2011) menar att distinktionen mellan bedömning i formativa och summativa syften är viktig, men att det är omöjligt att tala om en sådan distinktion i absoluta termer. Bedömning i formativt syfte är individrelaterad – särskilt den som i Harlens figur ovan kallas *informell formativ*, och som Cowie (2000) kallar *interaktiv formativ,* som försiggår i klassrumskontext kontinuerligt – medan summativ bedömning är mål-/kriterierelaterad. Man kan dock aldrig med säkerhet hävda att ens den individrelaterade interaktiva klassrumsbedömningen som sker i varje stund tjänar enbart formativa syften. Hill (1999; 2000) talar om termen *head-noting* som en process där lärare gör ’mentala anteckningar’ om elevers prestationer i kontinuerlig interaktion, som senare mycket väl kan användas i summativt syfte. Det finns idag inte tillräckligt med forskning för att vi ska kunna säga något om i vilken utsträckning sådan bedömning i dubbla syften kan anses ha hög validitet i båda hänseendena, menar Crooks (2011). Bennett (2011) är också en av dem som menar att det är problematiskt att göra en absolut distinktion mellan summativ och formativ bedömning. Ett noggrant utformat summativt bedömningstillfälle kan på ett framgångsrikt sätt ligga till grund för formativt arbete. Kunskap i att designa den typ av summerande tillfälle, som inte bara summerar utan ger tillräckligt rik information för att ha potential att fungera formativt i relation till uppställda mål/kunskapskrav är därför vad vi

behöver sträva efter. Att vidmakthålla distinktionen mellan formativ och summativ bedömning som svarandes mot olika syften framhålls dock generellt i de genomgångna studierna som viktigt.

Den formativa bedömningens plats i summativa system

I ett flertal artiklar som problematiserar den formativa bedömningens plats i summativa system uttrycks en oro för att begreppet formativ bedömning successivt, under de senaste fem åren, kommit att får en ny innebörd. Även i vår genomgång av empiriska studier kan skönjas en tendens till en mer instrumentell förståelse av begreppet formativ bedömning, framför allt i studier som fokuserar verktyg för formativ bedömning där IKT kommit att bli ett allt vanligare inslag. Chappuis (2005) (jfr även Chappuis & Chappuis, 2008) definierar problemet enligt följande:

> Currently being reshaped /…/ is the term formative assessment. It now is at risk of being understood merely as testing that is done often. In some extremes, it is little more than frequent summative assessment: testing that doesn't originate in the classroom, that creates another mark for the gradebook or a set of data to be analysed, and that, in theory, tracks individual and/or group progress toward the ultimate summative test – the high-stakes test that quantifies the school's adequate yearly progress.
>
> And that's the whole point, isn't it? When this is the way formative assessment is defined and used, clearly what we are trying to "form" is higher test scores. The need to raise test scores is of course real, and unlikely to go away any time soon. Still, we may want to watch out for what could be sacrificed unintentionally on our way to adequate yearly progress. (s. 38)

I artikeln *Assessment for Learning revisited: an Asia-pacific perspective* diskuterar Klenowski (2009) definitionen av begreppet *assessment for learning* (bedömning för lärande) med utgångspunkt i den tredje internationella konferensen kring bedömning för lärande som hölls på Nya Zealand 2009. Ett viktigt resultat av konferensen var ett positionerande paper där man beskriver hur tidigare definitioner av begreppet succesivt kommit att misstolkas och hur förståelsen av begreppet därmed har börjat ändra innebörd. Som exempel nämns att den definition som säger att bedömning för lärande handlar om "deciding where the learners are in their learning, where they need to go, and how best to get there" (Klenowski, 2009, s 263) i allt högre grad har kommit att (miss)tolkas som en uppmaning till lärare att summativt testa sina elever ofta för att bedöma uppnådda nivåer i enlighet med föreskrivna nationella mål/kunskapskrav. I ett sådant scenario blir resultatet/poängen – som är tänkta att vara en indikator på lärande – målet i sig. Torrance (2007) har studerat bedömningssituationer i högre utbildning och beskriver problematiken tydligt: "The practice of assessment has moved from assessment *of* learning, through assessment *for* learning, to assessment *as* learning, with assessment procedures coming completely to dominate the learning experience, and 'criteria compliance' replacing 'learning'." (s 281).

Hume och Coll (2009) ser samma tendens i fallstudier de har genomfört inom ramen för NO-undervisning på motsvarade högstadiet i Nya Zealand. De beskriver den smala och instrumentella förståelsen av formativ bedömning som en ny variant av 'teaching to the test', och menar att assessment as learning är en framväxande internationell trend. Senare års forskning (t.ex. James m fl, 2007) har visat på tydliga tendenser till att formativa klassrumspraktiker blir mekaniska och ritualiserade metoder som läggs ovanpå lärares ordinarie undervisningsmetoder och används på ett oreflekterat sätt, där tid och strukturer för gemensamt kollegialt arbete kring såväl teoretisk förståelse som praktisk tillämpning av formativa praktiker inte finns på plats. Bennett (2011) menar att det är olyckligt och grovt förenklat att tala om begreppet formativ bedömning som varandes *antingen* instrument/metod *eller* process. Det är en förenkling att definiera formativ bedömning som ett instrument, då även det mest noggrant och vetenskapligt konstruerade instrumentet sannolikt inte kommer fungera effektivt om de processer i vilka instrumentet används är felaktiga. Likaså är det en förenkling definiera formativ bedömning som enbart en process, eftersom även de mest omsorgsfullt konstruerade processer sannolikt inte kommer fungera om instrument/metodik inte är väl lämpade för det avsedda ändamålet. Process kan på intet sätt rädda olämplig instrumentering och vice versa. Konceptualisering av begreppet formativ bedömning måste därför, enligt Bennett, handla om hur instrument och process fungerar tillsammans i domänspecifika sammanhang. Detta återvänder vi till under kategorin *Verktyg*.

Sammanfattning – Teoretisk förståelse av formativ bedömning/begreppsutveckling

Teoribildning kring formativ bedömning har under 2000-talet främst tagit sitt ursprung i brittisk och australiensisk/nya zeeländsk kontext. I stort sett samtliga artiklar och bokkapitel som gåtts igenom för denna översikts räkning tar avstamp i Black och Wiliams översikt *Assessment and Classroom Learning* från 1998. De nyckelstrategier för formativ bedömning som nämnda forskare tillsammans med en rad andra brittiska forskare har utvecklat sedan dess har i hög grad definierat de empiriska studier som finns.

Nyckelstrategierna handlar om att gentemot eleverna tydliggöra lärandemål och skapa klassrumssituationer som på ett effektivt sätt visar var eleverna befinner sig i relation till dessa mål. Genom feedback – från lärare till elev, från elev till elev och genom att den enskilde eleven bedömer sin egen väg mot målen - ska lärandet sedan utvecklas.

På senare år har allt mer kritik framförts mot att formativ bedömning skrivs fram på ett teoretiskt förenklat sätt, där hänsyn inte tas till de kontextuella faktorer som ramar in undervisning/lärande. Frågan om huruvida formativa arbetssätt/metoder och de resultat man har kunnat se av sådant arbete kan generaliseras oavsett kontext diskuteras i allt högre grad. Under senare år är det alltfler som menar att det är nödvändigt att förstå (formativ) bedömning som en kontextbunden social praktik, där man inte kan förvänta sig samma resultat oavsett kontextuella villkor.

Det förs också en diskussion kring hur relationen mellan formativ och summativ bedömning ska förstås. Forskare tycks tämligen överens om att distinktionen fyller en funktion, då det handlar om att tydliggöra att bedömningar görs i olika syften. Alltfler menar dock att den dikotomisering som har blivit är olycklig och att man behöver arbeta för att harmoniera syftena. Det största hotet mot formativt klassrumsarbete tycks vara ökade externa krav på kvantifierbara kunskapsmått. I skolsystem där elevers kunskap frekvent ska testas, summeras och redovisas tycks det vara svårt för de formativa praktikerna att utvecklas på ett fördjupat sätt. Det finns en tendens att det formativa arbetet riktas mot att höja resultat på externa tester snarare än att forma och utveckla lärande.

I kapitlet *Developing a theory of formative assessment* gjorde Black och Wiliam (2006) ett försök att utveckla formativ bedömning inom ramen för verksamhetsteori med formativ bedömning som ett verksamhetssystem (jfr t.ex. Engeström, 1987; Engeström & Miettinen, 1999). I ljuset av den artikeln kan vi konstatera att övriga artiklar inom detta tema huvudsakligen arbetat med begreppsutveckling och försök att ringa in fenomenet i relation till forskningsfältet bedömning som helhet. Black och Wiliam har inte gått vidare på det verksamhetsteoretiska spåret[19] utan lämnat det därhän. Däremot har de fortsatt sitt arbete med ställningstagandet att formativ bedömning bör ses som ett situerat fenomen, d.v.s. formativ bedömning handlar om det ömsesidiga i processen undervisning–lärande. Wiliams bok *Embedded formative assessment* från 2011 kan ses som ytterligare ett bidrag till detta ställningstagande.

Policy

De artiklar som sorterats in under kategorin *policy* är på flera sätt nära knutna till sådant som presenterats i föregående kategori, då policy påverkas av förståelsen av begreppet formativ bedömning i enlighet med resonemanget som förts ovan. Artiklar som adresserar policyfrågor handlar i hög utsträckning om den formativa bedömningens plats i ett summativt system, och spänningar mellan bedömningar som görs i summativa och formativa syften.

Formativ bedömning i summativa system

Bedömning kan förstås från olika infallsvinklar (exempelvis politiska, sociala, pedagogiska) och fylla en rad olika funktioner, så som sortering, urval, diagnos, vägledning och prediktion (Morris & Adamson, 2010). Mångfalden av funktioner har gett upphov till olika former av bedömning; om målet är att utforma

[19] Vi är medvetna om att det finns mer att säga om detta, men vi lämnar den fortsatta diskussionen om detta till ett annat sammanhang.

bedömningsverktyg för selektiva ändamål eller för att påvisa och kunna jämföra resultat på aggregerad nivå används oftast standardiserade modeller där eleverna, i rättvisans namn, utför samma uppgifter på samma sätt samtidigt. Om syftet med bedömning däremot är diagnostiskt kan tillvägagångssättet vara helt annorlunda än vid en selektiv process. Då finns inte samma behov av vare sig standardisering eller formella arrangemang. Sådan bedömning benämns ipsativ eller individrelaterad, och innebär att individens nuvarande prestationer relateras till individens tidigare prestationer, snarare än till yttre normer eller kriterier. När vi talar om bedömning i summativa och formativa syften har vi i de flesta skolsystem i världen att göra med två olika system för bedömning som på något sätt måste fungera i relation till varandra. Även om lärare praktiserar kontinuerlig individrelaterad bedömning (formativ) måste individens prestationer vid givna tillfällen jämföras mot fastställda nationella mål/kriterier. Betyg sätts inte utifrån individens framsteg jämfört med sig själv, de sätts utifrån nationellt fastställda kunskapskrav. Att förena dessa båda funktioner har i viss utsträckning visat sig vara problematiskt och kontroversiellt (Berry, 2011a). Det problematiska ligger främst i det faktum att individrelaterad formativ bedömning, som i en stor mängd studier visat sig vara högst gynnsam för elevers lärande och motivation, är svår att hålla fast vid i system som fäster mycket stor vikt vid extern kontroll och ansvarsskyldighet som redovisas och jämförs via skolors resultat på externa mätningar. Visserligen har den forskning kring olika aspekter av formativ bedömning som redovisar resultat i termer av effektstorlekar – och kanske särskilt den översikt av Black och Wiliam (1998) som talar om effektstorlekar som är bland de högsta någonsin för utbildningsinterventioner – påverkat beslutsfattare så till vida att man världen över betonar vikten av formativt arbete i skolan. Frågan är hur relationen mellan individrelaterad formativ bedömning och (storskalig) summativ bedömning ser ut idag.

I artikeln *The first ten years of Assessment in Education* (2004) sammanfattar Broadfoot och Black jamestidsperioden 1993-2003 för tidskriften med samma namn. Åren de blickar tillbaka på faller till stor del utanför den tidsperiod vår översikt tar sin utgångspunkt i, men är intressanta så till vida att författarna använder summeringen av perioden som underlag för att se framåt på kommande tioårsperiod. Tidskriften Assessment in Education startades som svar på ett ökat intresse hos forskare för kontextens påverkan på formandet av bedömningspraktiker, och karaktäristiskt för artiklar som publiceras i tidskriften är att bedömning ses som en kontextbunden social praktik, där tänkandet, vanorna, teknologierna och politiken under olika tidsperioder förstås som formande för de bedömningspraktiker som växer fram på skolor och universitet såväl som på arbetsplatser och i mindre formella lärandemiljöer.

Samtidigt som man ser en ökad medvetenhet kring att bedömningspraktiker är kontextberoende ser man naturligtvis också att den ökade globaliseringen i viss mån för med sig att vi brottas med samma problematik länder emellan. Broadfoot och Black konstaterar att forskare världen över under perioden 1993-2003 i allt högre grad har ägnat ökat intresse åt bedömningars och bedömningsmetoders olika syften och spänningen mellan olika syften. Oförenligheten mellan den policy som i många länder har lett till en test-driven utbildningskultur å ena sidan och det explicita främjandet av termen livslång lärande å andra sidan är det mest anmärkningsvärda exemplet på en sådan spänning som studerats och diskuterats flitigt.

> Research findings presented in this journal – see for example, Crooks, et al (1996), Airasian and Gregory (1997) and Firestone (1998), as well as elsewhere – make it clear that this 'empowerment' agenda cannot be achieved alongside the punitive use of high-stakes testing to raise 'standards'. Any possible short-term gains that the more or less extreme instrumentalism of the latter engenders, encourages 'teaching the test' (Morrison & Tang Fun Hei, 2002) as well as anxiety and low self-esteem among the less successful (Harlen & Deakin-Crick, 2003), and is bought at the price of turning many students off formal learning forever. It would seem from these studies, that whilst it may be technically possible to reconcile formative and summative assessment within the same spirit of personal growth and empowerment, there will need to be a substantial shift in the prevailing political and policy priorities of many countries to achieve this. (Broadfoot & Black, 2004, s 11).

Hur denna utveckling har sett ut under den efterföljande tioårsperioden är naturligtvis högintressant för vår översikt. Vi har inom givna ramar inte haft möjlighet att närstudera policy i olika länder, men kan utifrån de artiklar vi gått igenom konstatera att det inte finns något som pekar på att det skulle ha skett en sådan omprioritering som beskrivs i citatet ovan. Användandet av nationellt standardiserade tester och internationella

kunskapsmätningar/jämförelser är fortfarande dominerande, vilket innebär att situationen som beskrivs ovan fortfarande måste betraktas som minst lika påtaglig. Sverige är ett exempel på land som under den senaste femårsperioden i allt högre grad börjat använda nationellt standardiserade tester, och på Nya Zeeland – som ofta framhållits som ett land där lärares bedömningar av eleverna väger tungt och där formativt klassrumsarbete legat i framkant – infördes från 2010 ett nytt nationellt prov-system som fått forskare där att uttrycka oro för att det formativa arbetet kan komma att hamna i skymundan (se exempelvis Crooks, 2011).

Allra tydligast framstår testkulturen i amerikanska studier, där man menar att *No Child Left Behind*-reformen, med sin oerhört starka betoning på frekvent testande för att mäta skolors *annual yearly progress* (AYP) i kombination med påföljder/hot om påföljder för de skolor som inte kan uppvisa sådana årliga framsteg, har lett till att skolornas mål inte längre är elevernas lärande utan just AYP i sig, så som den visar sig i resultat på standardiserade tester i utvalda ämnen/ämnesområden. I en kontext där summativa resultat på standardiserade tester är ett så dominant inslag menar exempelvis Flaitz (2011) att det är en utmaning att få lärare att inse att det går att arbeta formativt. Många menar dessutom att det frekventa testandet leder till andra effekter:

> We are a nation obsessed with the belief that the path to school improvement is paved with better, more frequent, and more intense standardized testing. The problem is that such tests, ostensibly delivered to "leave no student behind", are in fact causing major segments of our student population to be left behind because the tests cause many to give up in hopelessness – just the opposite from what politicians intended. (Stiggins, 2002, s 759).

I en sådan kontext har man ofta kommit att använda sig av så kallade *benchmark assessments* eller *interim assessments*, tester som ofta utvecklas av företag och marknadsförs under epitetet formativ bedömning. Syftet med sådana tester, som man låter eleverna göra med jämna mellanrum under året, är huvudsakligen att de ska fungera som en mekanism för att identifiera elever som riskerar att misslyckas med att uppvisa AYP vid de standardiserade testerna i slutet av året (Flaitz, 2011).

Berry (2011a) diskuterar utvecklingen med avseende på formativ bedömning i Kina, Hongkong och Taiwan, tre länder med en mycket lång och djupt rotad examinationskultur. Hon menar att policynivån i dessa länder under det senaste decenniet svarat på den internationella trenden på sitt eget sätt. De har formulerat policy med intentionen att minska det utbredda testandet och i stället uppmuntrat till formativa praktiker, och initierat en rad pilotprojekt i syfte att implementera bedömning för lärande. Enligt Berry (2011a) har studier i dessa länder dock generellt visat att lärare har saknat tillräcklig kunskap för att omsätta idéer om bedömning för lärande i konkret klassrumspraktik, vilket därför får anses vara en utmaning man har framför sig.

Whitty (2006) menar, utifrån det brittiska perspektivet, att medvetenheten på policynivå kring vikten av formativ bedömning har ökat och att lärares undervisningspraktiker har påverkats – i mångt och mycket som följd av Black och Wiliams artikel från 1998. I England har man inte kunnat se några direkta förändringar av den summativa kontext som ramar in de formativa praktikerna, medan man i Wales, däremot, har använt de resultat som lyfts fram i artikeln som grund för att ta steg bort ifrån nationellt standardiserade tester och publicering av rankinglistor.

Kontextens betydelse

Vi har kunnat konstatera att man på policynivå kommit att erkänna betydelsen av formativ bedömning, mycket på grund av Black och Wiliams översikt från 1998. Den påverkan översikten har haft är i hög grad avhängig dess betoning på statistiskt baserade prövningsmetoder för att fastställa mått på styrkan av effekter (effektmått) på inlärningsvinster hos elever. En annan studie som ofta lyfts fram i relation till påverkan på policy är en av McDonald och Boud (2003) som även den kvantitativt visar på betydande resultatökningar i tester efter att elever har tränats i självbedömning. Den senare är för övrigt ett exempel på hur man på policynivå har en tendens att generalisera resultat utan att i hög grad problematisera de kontexter i vilka studier har gjorts. McDonald och Bouds studie var en interventionsstudie som – enligt författarna själva - genomfördes under osedvanligt gynnsamma omständigheter i high school-miljö, där ett stort antal lärare fick kontinuerlig och professionell utbildning i att träna eleverna i självbedömning. Det finns en tendens till övertro på att de effekter

man har fått kommer bli liknande under andra kontextuella villkor (till exempel med avseende på elevers ålder).

Ett senare – och ännu mer välkänt - exempel på hur effektmått avseende inlärningseffekter kan få kraftigt genomslag för policy världen över är naturligtvis John Hatties *Visible Learning* från 2009. I boken listar Hattie en rad faktorer som påverkar elevers studieresultat, jämför effekterna och konstaterar att lärarens roll är betydande. Hattie talar om läraren som antingen expert, erfaren eller novis, där expertläraren har något som de andra inte har: förmågan att utmana elever på precis rätt nivå och vid rätt tillfälle. Mycket av det som kategoriseras som aspekter av formativ bedömning – såsom utvecklandet av självreglering och metakognitiva strategier hos eleverna samt feedback mellan lärare och elev – betonas av Hattie som viktiga aspekter för ett ökat lärande. Hatties meta-studie har fått stort genomslag på policynivå även i Sverige, men också kritiserats för att exempelvis de mätningar som redovisas inte tar hänsyn till modererande faktorer (se t.ex. Nilholm, 2013). Håkansson (i Sundström, 2011) menar dessutom att den så kallade "expertläraren" inte ska tolkas som en existerande person utan som resultatet av det man sett i mängder av studier. De flesta lärare tillhör förmodligen den erfarna skaran som ibland - vid vissa tillfällen, med en viss elevgrupp eller under vissa andra omständigheter - gör något lysande. Det viktiga är inte att utse experter utan att föra in de här aspekterna i samtalen om undervisning och lärande, menar Håkansson.

Broadfoot och Black (2004) konstaterade redan i sin tioåriga tillbakablick att policy-beslut som baseras på storskaliga jämförande tester ofta är 'kontext-blinda':

> Governments have been quick to seize on the headline findings whether positive or negative, with little regard to the dependability of the data produced or the range of possible explanations for them. All too often, responses have been context blind, the apparent success of other countries being attributed to specifics of curriculum design or pedagogic strategies rather than the more subtle influences of culture and tradition. (Broadfoot & Black, 2004, s 14).

I en artikel från 2005 diskuterar Black och Wiliam hur policy, politik och kultur begränsar och skapar mening i lärares bedömningspraktiker. Bedömningstraditionerna i England, Frankrike, Tyskland och USA jämförs för att undersöka möjligheterna till integration mellan summativa och formativa funktioner av bedömning i den obligatoriska skolan, och bilden som framkommer är något paradoxal. Frankrike står ut som ett land där man har utvecklat effektiva formativa klassrumspraktiker på bred basis, men där nyckelbeslut om elevens fortsatta utbildningsbana görs på grundval av examinationsresultat. I Tyskland är lärare involverade i nyckelbeslut om elevernas framtid, men basen för dessa beslut utgörs av formella tester. I England vägs lärares bedömningar in i nationella bedömningar, men ansvarskraven på lärare har resulterat i dokumentationssystem som inte gynnar elevernas lärande. I USA, slutligen, menar författarna att det finns så många summativa bedömningssystem på olika nivåer att det varken finns tid eller utrymme till att ägna sig åt formativ bedömning.

Ett flertal av de artiklar vi har studerat menar att formativa interventioner *är* kontextberoende, och att ansvaret för ett förändrat arbetssätt därmed inte kan vila enbart på lärarna. I en studie publicerad 2010 beskriver Cooper och Cowie utifrån ett kollaborativt aktionsforskningsprojekt hur de tydligt ser att förändrade bedömningspraktiker är beroende av en dynamisk interaktion mellan personlig nivå (läraren som genomför förändringarna), mikronivå (skolnivå) och makronivå (policynivå). Positiva effekter hos de högstadielärare som deltog i projektet och deras elever tolkas som starkt sammankopplade med de enskilda lärarnas kunskap och engagemang å ena sidan, och formellt stöd och skapande av strukturer från rektor å andra sidan, samt en gemensam förståelse hos lärare och rektor av hur formativa praktiker möter strategiska mål som handlar om att höja resultat på ett sätt som märks i externa mätningar.

Det finns också jämförelser mellan länder som konstaterar likheter oavsett kontext. I boken *Formative assessment: improving learning in secondary classooms* (CERI, 2005), som är skriven inom ramen för ett *What works*-koncept för OECD-länder, samlas exempel på formativt klassrumsarbete från åtta olika länder (Canada, Danmark, England, Finland, Italien, Nya Zealand, Queensland Australien och Skottland). Man konstaterar på ett generellt plan att det är vanligt att lärare använder element av formativ bedömning i sin undervisning, men att det i huvudsak finns två hinder för ett mer utbrett användande; (i) spänningen mellan klassrumsbaserad formativ bedömning å ena sidan och summativa 'high visibility'-test å den andra, och (ii) avsaknaden av

samband mellan klassrumsbedömningar å ena sidan och systematisk skolutvecklingsarbete å den andra. På basis av detta konstaterar man att det inte räcker att ge lärare metoder – tid och pengar måste investeras i utbildning så att lärare kan göra dessa metoder till sina. Dessutom måste lärarutbildningar tänka över de strategier som används för att utbilda lärarstudenter i bedömningsfrågor. Ett liknande resonemang förekommer i flera av de artiklar vi har studerat, och vi återkommer till frågan längre fram, under rubriken *Implementering*/*Professionellt lärande.*

Via forskningslitteratur och fallstudier i de olika länderna som beskrivs i ovan nämnda bok identifierar man följande element som centrala för att formativa klassrumspraktiker ska fungera framgångsrikt:

- Etablerandet av klassrumskulturer som uppmuntrar aktivitet och interaktion
- Tydliga målbeskrivningar och följande av elevernas väg mot målen
- Användandet av varierade undervisningsmetoder för att möta varierade elevbehov
- Användandet av varierade bedömningsmetoder för att bedöma elevers förståelse
- Lämplig användning av feedback till eleverna och anpassning av undervisningen för att möta de behov som framkommer
- Aktivering av eleverna i lärandeprocessen

I studien konstateras att man, för att kunna arbeta i enlighet med punkterna ovan, behöver hitta sätt att koppla samman formativa och summativa syften och tillvägagångssätt, försäkra sig om att klassrumsbedömning kopplas till systematiskt kvalitetsarbete på skol- och systemnivå, investera i utbildning och stöd för lärare och skolledare och bygga starka broar mellan forskning, policy och praktik.

Sammanfattning – Policy

Kvantitativa studier, översikter och metastudier, som med hjälp av effektmått redovisar hur elevers lärande påverkas positivt av formativa arbetssätt, har under 2000-talet påverkat beslutsfattare världen över i riktning mot att vilja föra in mer formativ bedömning i skolan. De mest påtagliga exemplen på sådana studier är *Assessment and Classroom learning* (Black & Wiliam, 1998) och *Visible Learning* (Hattie, 2009). Mer eller mindre storskaliga initiativ på regional eller nationell nivå för att implementera formativa bedömningspraktiker på alla nivåer i utbildningssystemet har genomförts. De formativa arbetssätten har dock visat sig ha svårt att slå rot i system som fäster mycket stor vikt vid extern kontroll och ansvarsskyldighet som redovisas och jämförs via skolors resultat på externa mätningar. Accountability-diskursen är påtaglig i många länder, och testkulturen har snarare blivit mer än mindre påtaglig under 2000-talet.
Forskare kritiserar också policynivån för att vara kontextblind när det gäller tron på generaliserbarheten hos effekter av formativ bedömning som har kunnat konstateras i olika studier. Återigen betonas att effekten man får av formativa interventioner är kontextberoende. Generellt kan man dock se att det i huvudsak finns två hinder för ett mer utbrett användande av formativ bedömning, där det ena hindret utgörs av just spänningen mellan klassrumsbaserad formativ bedömning och summativa high-visibility-test. Det andra hindret utgörs av avsaknaden av samband mellan klassrumsbedömningar å ena sidan och systematisk skolutvecklingsarbete å den andra (CERI, 2005). För att komma till rätta med det senare krävs investeringar i utbildning av lärare och skolledare i anslutning till att formativa praktiker ska implementeras, samt att starka broar mellan forskning, policy och praktik byggs.

Verktyg

Den kategori vi här kallar för verktyg innefattar till stor del empirisk forskning som kan relateras till de nyckelstrategier (Wiliam & Thompson, 2007; Black & Wiliam, 2009) som redovisades tidigare i denna översikt, och som i hög grad beskrivs som *varandes* formativ bedömning. Vi förstår dem snarare som verktyg för arbete med bedömning i formativt syfte.

När vi talar om verktyg vill vi inledningsvis betona att effekten av olika formativa interventioner kan vara mycket svår att uttala sig om. Det krävs betydligt mer tid än vad vi har haft till vårt förfogande för att i detalj

kunna gå in och granska tillvägagångssätt och kvalitet i gjorda analyser. Vi väljer därför att lyfta tre artiklar som, var och en på sitt sätt, visar på att de storskaliga kvantitativa analyser som får stor påverkan på policy och skolpraktik inte är oproblematiska eller kan tolkas som givna sanningar.

Bennett (2011) hör till dem som i flera avseenden kritiserar de växlar som dragits på forskningen kring formativ bedömning. Han menar att begreppet inte är tillräckligt väl definierat i termer av vilka redskap och praktiker som egentligen avses. Därmed blir det också svårt att på ett trovärdigt sätt tala om hur effektiv formativ bedömning kan vara på en generell nivå. Under rätt omständigheter kan olika aspekter av formativ bedömning vara effektiva för att öka lärande, men – återigen – behöver man inse att detta varierar mycket beroende på olika kontextuella faktorer. Bennett menar vidare att flera av de existerande meta-analyserna som hänvisas till är problematiska. Vad gäller det innehållsmässigt vida begreppet formativ bedömning talar man ofta om att formativa interventioner innebär alltifrån medelstora till mycket stora positiva effekter för elevers lärande. Metaanalys i sin enklaste form är en metod som i huvudsak slår samman resultat från en uppsättning jämförbara studier som innefattar statistik – ofta beräknade effektstorlekar. För experimentella studier beräknas typiskt effektstorleken som skillnaden mellan en behandlingsgrupp och en kontrollgrupp, dividerat med standardavvikelsen. Man bör dock, menar Bennett, vara medveten om att meta-analyser kan producera tämligen meningslösa resultat, exempelvis om beräkningar är gjorda på basis av alltför disparata företeelser, eller om vissa av de studier som tas med i beräkningarna är opublicerade och därmed inte varit föremål för kritisk granskning. Bennett hävdar att dessa problem finns i exempelvis Black och Wiliams översikt från 1998. De studier som inkluderas är alltför disparata för att kunna sammanfattas meningsfullt i *en* meta-analys. Företeelser som återkoppling, målstyrning, självuppfattning, kamratbedömning, självbedömning, lärares val av bedömningsuppgifter och lärares användning av tester är bara några av de områden som finns representerade i de studier som innefattas i Black och Wiliams analys under paraplytermen formativ bedömning.

I artikeln *Meta-Analytic Methodology and Inferences About the Efficacy of Formative Assessment* ifrågasätter även Kingston och Nash (2011) de siffror som ofta redovisas kring effekterna av formativ bedömning. Särskilt riktar man sin kritik mot Black och Wiliam (1998), då denna artikel blivit så frekvent citerad sedan den publicerades och då den ständigt används av andra forskare för att underbygga argument för formativ bedömning. Kingston och Nash har gått igenom 300 artiklar som på olika sätt berör effekterna av formativ bedömning, och funnit att endast 13 av dessa gav den information som krävs för att kunna beräkna relevanta effektstorlekar. Utifrån detta menar de att man möjligen kan tala om effekter i storleksordningen[20] 0.20, snarare än de 0.70 som ofta hänvisas till. De hävdar också att två formativa interventioner framstår som särskilt positiva i jämförelse med andra, där den ena handlar om lärares professionella utveckling och den andra om användandet av datorbaserade formativa system. De 13 studier som Kingston och Nash hävdar är användbara är samtliga sådana där kontrollgrupper har använts och effektstorlekar har beräknats. Totalt 42 effektstorlekar redovisas i de 13 studierna: 19 inom matematik, 12 inom läsning eller skrivande, 10 inom naturvetenskap och en inom musik. Studierna involverar sammantaget elever från förskola, grundskola och gymnasium. 12 av de 13 studierna är från 2000-talet. Moderatoranalys visade att ämnesområde har störst betydelse för den effekt man når med formativa interventioner hos eleverna, och störst effektstorlek ser man inom matematik. I vilken åldersgrupp man genomför interventionerna spelar, enligt Kingston och Nash, mindre roll.

Kingston och Nashs metodologi kritiseras i sin tur av Briggs m fl (2012), som menar att Kingston och Nash inte gjort en tillräckligt vid sökning i det initiala stadiet (på grund av valda söktermer som inte med tillräcklig bredd täcker in det som kan rymmas under paraplytermen formativ bedömning). Briggs m fl framhåller att det finns andra studier som borde varit med i Kingston och Nashs metaanalys, för att de uppfyller alla kriterier för inkludering, men också att en av de 13 studier som är med i metaanalysen *inte* uppfyller de krav för inkludering

[20] Följande gränsvärden gäller som en vägledning för hur man ska förstå det standardiserade måttet: Själva måttet varierar från 0 till 1. För att man överhuvudtaget ska kunna tala om en viss effekt har man i samhällsvetenskapliga sammanhang enats om 0,20 som ett slags gränsvärde. Följande gränsvärde, som benämns måttlig effekt, är 0,35 (ibland anges 0,5 som krav på måttlig effekt). Först då man når effekter över 0,6 (ibland ställs krav på 0,8) ses effekten som god (jfr t.ex. Ellis, 2010).

som författarna själva skriver fram (Ruiz-Primo & Furtak, 2007), då den inte använder den kontrollgrupp som enligt beskrivning av inkluderingskriterier ska finnas med. Andra metodologiska aspekter kritiseras också, som exempelvis det faktum att vissa studier involverar både för- och eftertest av elevers kunskaper, medan andra involverar endast eftertest.

För- och eftertest är för övrigt ytterligare en aspekt som vi själva upplever som bekymmersam i de studier vi har gått igenom. Dessa tester utgör i någon mening grunden för att kunna uttala sig om effekten av interventioner, men ingen av de artiklar som gåtts igenom för den här översiktens räkning varken redovisar eller problematiserar för- och eftertesternas innehåll. Vi finner det anmärkningsvärt att man på basis av effektstorlek talar om hur elevers lärande har ökat, utan att någon definition görs av vad detta lärande består i eller med hjälp av vilka tester in- och utgångsvärde har mätts.

Kategorin verktyg är – vid sidan av kategorin teori – den största sett till innehållet i de artiklar vi har gått igenom. Verktygen för arbete med bedömning i formativt syfte är, som redan konstaterats, många och disparata. Studier som på olika sätt handlar om feedback är vanligast förekommande, men det finns också en betydande mängd studier som rör IKT-relaterade verktyg för formativ bedömning. Därutöver förekommer studier kring bland annat bedömningsmatriser, kamratbedömning, självbedömning, portfolio och klassrumsinteraktion. Olika verktyg kommer att redovisas under olika rubriker. I de fall där det finns översikter eller meta-studier gjorda utgår vi ifrån sådana, men exemplifierar också genom att lyfta in enskilda empiriska studier (med kvantitativ såväl som kvalitativ ansats).

Feedback

Begreppet feedback, eller återkoppling, upplevs ofta som synonymt med formativ bedömning. När gapet mellan elevers nuvarande nivå och den önskvärda målnivån ska överbryggas handlar det i hög utsträckning om att elever får olika former av feedback. Feedback är en ständig del av den kontinuerliga informella klassrumsinteraktionen, samtidigt som den ibland kan vara mer formaliserad och/eller uttryckas skriftligt (jfr Harlen, 2012).

Inte heller feedback har varit ett oproblematiskt område att studera. Redan 1989 beskrev Sadler hur feedback från lärare till elev – även då den var att betrakta som valid och reliabel – långt ifrån alltid ledde till utveckling eller förändring hos eleven. Perrenoud (1998) liknade en gång feedback från lärare till elev vid flaskposter som kastas ut i havet – ingen är säker på att de någonsin hittar en mottagare. Ett annat problem lyfts av Hargreaves (2011), som menar att många studier kring feedback har gjorts under experimentella omständigheter snarare än i autentisk klassrumsmiljö (t.ex. Azevedo & Bernard, 1995; Narciss & Huth, 2004), och att man utifrån dem inte kan uttala sig om hur feedback fungerar under andra omständigheter.

Feedback har dock skrivits fram som ett effektivt verktyg för att öka elevers lärande. Mest påtagligt syns detta i den frekvent citerade artikeln *The power of feedback* från 2007, där Hattie och Timperley granskar ett antal metastudier kring elevrelaterad feedback. Hattie och Timperleys artikel faller tidsmässigt inom ramen för studier som kommer i fråga för vår översikt, men det är viktigt att påpeka att de metastudier som ligger till grund för författarnas diskussion samtliga är gjorda innan millennieskiftet (i huvudsak 1980- och 90-tal). På basis av dessa metastudier gör Hattie och Timperley en kategorisering av feedback i fyra nivåer: den uppgiftsrelaterade, den processrelaterade, den självreglerande och den personliga. Uppgiftsrelaterad återkoppling är starkare när den handlar om felaktiga tolkningar snarare än en brist på förståelse, och den är mest som effektiv när den ger ledtrådar och information avseende felaktiga hypoteser och idéer och sedan pekar mot utvecklingen av mer effektiva strategier för behandling och förståelse av materialet. Processrelaterad återkoppling är mest fördelaktig då den hjälper eleverna bort från felaktiga hypoteser och ger ledtrådar till sökstrategier. Sådana ledtrådar medvetandegör elever om den kompetens och de strategier som behövs för att lösa en uppgift. I idealfallet rör sig sådan återkoppling från uppgiften till den process som krävs för att lösa uppgiften och vidare till vad som krävs för att fortsätta till mer utmanande uppgifter eller mål. Återkoppling på självregleringsnivån är kraftfull då den leder till engagemang/lust att investera ytterligare möda i uppgiften, och till att eleven får en ökad tilltro till att hen kan klara av uppgiften. När återkoppling uppmärksammar de regulatoriska processer som behövs för att lösa en uppgift, kan elevernas föreställning om betydelsen av den egna insatsen bli en viktig moderator i lärandeprocessen. Återkoppling på den personliga nivån (oftast beröm)

är, enligt författarna, sällan effektiv då den inte utgår från de frågor som huvudsakligen brukar relateras till effektiv feedback (Vart är jag på väg? Hur är jag på väg? Vad är nästa steg?). Det betonas att återkoppling inte är en enkel stimulus-respons-rutin, utan ställer höga krav på såväl ämneskunskaper som förmågan att hantera komplexiteten i bedömningsinformation. Det är dessutom endast när eleverna tar sig an mål som de förstår innebörden av som feedback kan fungera effektivt. Författarna betonar också att bedömningsuppgifter alltför ofta utformas för att ge ögonblicksbilder av lärande snarare än för att ge information som kan användas av lärare eller eleverna själva kring de tre feedbackfrågorna.

Yin, Tomita och Shavelstone (2014) har studerat effekten av formell feedback till mellanstadieelever i fysik. Genom att slumpvis dela in 52 elever i en experimentgrupp och en kontrollgrupp som fick precis samma undervisning av samma lärare under en tremånadersperiod, med undantag för att experimentgruppen fick formell formativ feedback vid tre tillfällen under perioden. Samtliga tre tillfällen utgjordes av en så kallad POE-aktivitet (predict-observe-explain) och en utmanande fråga som var designad utifrån varje individs förståelse. Utifrån sitt resonemang under de formativa kontrollstationerna fick eleverna ytterligare feedback från läraren. Både experimentgruppen och kontrollgruppen testades på fyra olika sätt före och efter tremånadersperioden; två multiple-choice-test, ett kortsvarstest och ett problemlösande test. Experimentgruppens resultat vid tre av fyra eftertest var signifikant bättre än kontrollgruppens. Vid kortsvarstestet var experimentgruppens resultat dock inte bättre än kontrollgruppens. En styrka i studien är att samma lärare undervisade i båda grupperna, och att man därmed undviker den så kallade lärareffekten. Studien är dock en av många där man beräknar signifikans på så små grupper som 26 elever utan att föra en problematiserande diskussion kring valet av analysmetod och statistiska beräkningar.

Hargreaves (2011) har i en studie låtit 88 lärare (främst från Storbritannien) resonera skriftligt kring feedback genom att fortsätta meningen ”Feedback becomes effective when…” (s 122) och kompletterat detta med intervjuer med sju elever i åldern 9-10 år. Resultaten visar att både lärare och elever tycker att feedback är som mest effektiv när den ges i daglig och stundlig interaktion. Det som i övrigt skiljer sig från tidigare existerande forskning kring ämnet, som ofta handlat om form och innehåll i återkopplingen, är att lärarna här betonar sociala och personliga aspekter av feedbackprocessen. Feedback tas emot bättre om den kontext feedbacken ges i bygger på förtroendefulla relationer och trygghet. Allal och Mottier Lopez (2005) menar att formativ bedömning är ett socialt medierat ramverk som främjar elevens ökande kapacitet att bli alltmer självreglerande i sitt lärande, och Hargreaves (2011) konstaterar mot bakgrund av resultaten i hennes studie att den relationella aspekten kan vara den pusselbit Sadler (1989) saknade när han undrade vad det kunde bero på att feedback baserad på bedömningar som till synes är valida, reliabla och rimliga ändå inte leder till utveckling hos mottagaren. Elevers uppfattningar kring feedback återkommer vi till längre fram i texten.

Kamratbedömning

Även kamratbedömning är en aspekt som ofta lyfts fram som central i formativt arbete, och den finns med som en av de fem nyckelstrategier som betonas av bland annat Black och Wiliam (2009). Kamratbedömning kan vara en högst varierande företeelse i termer av djup och utmaning (Harris och Brown, 2012), och det är svårt att mäta de konkreta effekterna av den (Strijbos & Sluijsmans, 2010). Kamrater, till skillnad från lärare, är inga ämnesexperter, och därmed varierar riktigheten i kamratbedömning – den kan vara allt ifrån korrekt till inkorrekt och direkt missledande. Helt uppenbart är att formativ feedback kräver mycket av den som ger feedbacken – en uppgift som ofta är avancerad även för lärare. Att ge effektiv feedback kräver såväl förståelse av uppgiftens mål och framgångskriterier som att bedömaren har förmågan att uttala sig om relationen mellan den aktuella uppgiften och vad som krävs för att överbrygga gapet mellan aktuell prestation och en god färdig prestation. Till det kommer förmågan att kunna uttrycka denna feedback, i tal eller skrift, på ett sätt som kamraten kan förstå. Det mest positiva med kamratbedömning anses vara att det ger större möjligheter för varje individ att få feedback än om bara läraren var den som gav den, samt att den som granskar någon annans prestation i avseende att ge feedback får en ökad förståelse för sin eget lärande på köpet, och att den metakognitiva förmågan därmed tränas:

> There are many more students than teachers in the classrooms. Consequently, feedback from peers can be more immediate, timely, and individualized than teacher feedback. True, it may take peers some time to develop the appropriate skills, and they will certainly need training, but their availability is overwhelming. (Topping, 2010, s 62).

Topping (2010) förespråkar starkt användandet av kamratbedömning, men erkänner samtidigt att det är ett område som ännu inte i särskilt hög grad är studerat. Systematiskt användande av kamratbedömning innebär att elever får ägna mycket tid åt bedömningsrelaterade aktiviteter. Om detta ger kognitiva, metakognitiva och sociala vinster för eleverna är tiden väl använd, menar han, men kring detta behövs mer forskning. Tiden som används till kamratbedömning hade ju naturligtvis kunnat användas på andra sätt, därför behöver vi veta mer om hur 'kostnadseffektiv' kamratbedömning är i förhållande till andra arbetssätt.

Antalet studier kring kamratbedömning i grundskolans kontext förefaller vara mycket begränsat, endast en handfull studier framkom i våra sökningar. Den stora majoriteten studier är gjorda i relation till högre utbildning (Topping, 2010), vilket naturligtvis är en helt annan kontext än grundskolan, där barnen är mellan 7-16 år. Då vi tillämpade våra inkluderingskriterier efter de initiala sökningarna var det tydligt att en mycket stor mängd av forskningen kring formativ bedömning är gjord i relation till äldre elever och högre utbildning. Det är dock långt ifrån alltid detta faktum problematiseras. Det finns översikter som tillkommit för att ge vägledning i bland annat grundskolan, där man kan få intrycket att den forskning som refereras till är gjord i grundskolekontext. Referenserna man stödjer sig på kring exempelvis kamratbedömning i skriften Forskning i korthet 2014:2 (Kommunförbundet Skåne, 2014) kommer nästan uteslutande från kontexten högre utbildning, medan illustrerande bilder leder tanken till yngre barns lärande. Man kan utifrån detta konstatera att kamratbedömning inte bara generellt är ett tämligen outforskat område, det tycks finnas ett specifikt behov av studier som riktar sig mot en grundskolekontext.

Majoriteten av de studier som finns tar dessutom sin utgångspunkt i studenters egna upplevelser av kamratbedömning, och det finns, enligt Strijbos och Sluijsmans (2010), ett stort behov av metodologisk utveckling i riktning mot studier som kan säga något om den faktiska effekten av kamratbedömning i olika utbildningskontexter. De talar främst om behovet av (kvasi)experimentella studier där kontrollgrupper används.

Med referens till ett tjugotal studier – som nästan uteslutande är gjorda före 2000 - lyfter Topping i en kort översikt (2009) fram kamratbedömning som en faktor som i hög grad är positiv och leder till att elever inte bara får mer feedback än vad de hade fått om det bara hade varit läraren som gett den, utan också att det rör sig om feedback av lika god kvalitet. Detta konstateras utifrån studier (t.ex. Falchikov, 2001; Sadler & Good, 2006; Falchikov & Goldfinch, 2000) där lärares och kamraters bedömningar av samma uppgifter har jämförts och man har sett en hög interbedömarreliabilitet. Sådana studier är främst gjorda i högre utbildning, men Topping menar att kamratbedömning kan och behöver introduceras i tidigare skolformer.

Gennip, Segers och Tillema (2009) har gjort en systematisk litteraturöversikt om studier av kamratbedömning från 1990-2007 (totalt 15 studier), och konstaterar att de studier som finns i hög grad har intresserat sig för överensstämmelse mellan lärares och elevers summativa bedömningar av prestationer snarare än kamratbedömning som en formativ och social praktik. Ingen av de studier som gjorts har intresserat sig för under vilka omständigheter kamratbedömning är ett gynnsamt verktyg, och de problematiserar heller inte kamratbedömares olikheter eller olika förmågor, utan betraktar dem som en homogen grupp. Topping (2010) konstaterar att elever är allt annat än en homogen grupp, men att man i kamratbedömningssammanhang principiellt bör para ihop elever som är på likartad kunskapsmässig nivå. Detta innebär att "duktiga" elever matchas med varandra, och att elever med större svårigheter matchas med varandra. Topping menar att läraren i det senare fallet behöver fungera som ett stöd i större utsträckning, men att vinsterna kan bli lika stora för mindre duktiga elever. Vi återvänder till detta under kategorin elevuppfattningar.

Bland våra 340 artiklar är de empiriska studierna kring kamratbedömning få. Ett exempel är Crane och Winterbottom (2008), som studerat en intervention innefattandes kamratbedömning i en klass bestående av 23 högpresterande elever i åldern 13-14 år. Interventionen genomfördes under en kortare period i ett undervisningsavsnitt som handlade om fotosyntesen, där eleverna delades in i kamratpar som kontinuerligt gav varandra feedback utifrån givna kriterier. Eleverna fick svara på enkätfrågor och fokusgruppsintervjuades både före och efter interventionens genomförande. Resultaten av enkäten visar att elevernas uppfattningar kring vem

som påverkade deras lärande i högst grad inte hade förändrats nämnvärt efter genomförandet, nästan alla svarade både före och efter att det är läraren som påverkar deras lärande i högst grad, och bara några enstaka elever menade att kamraterna eller de själva spelade roll för lärandet. I intervjuerna blev dock bilden något annorlunda, där menade 15 elever efter genomförandet att kamraternas respons varit viktig för deras lärande. Studien innefattar ingen kontrollgrupp och kan inte säga något om ifall elevernas lärande (i termer av resultat uppvisade på prov i slutet av området) har påverkats av kamratbedömning. Man konstaterar endast att resultaten på det avslutande testet var sådant som man hade kunnat förvänta sig utifrån elevernas kapacitet – oavsett kamratbedömning. Eventuella vinster skulle därmed i första hand ligga i att man hos vissa elever kunde skönja en utveckling mot att de efter perioden med kamratbedömning tydligare såg sin egen agens i lärandet.

Gielen m fl (2010a) genomförde en studie av kamratbedömning på högstadiet med avseende på textproduktion/essäer skrivna på modersmålet. 43 sjundeklassare från två olika klasser deltog i studien, som hade ett upplägg med textutkast (refereras till som förtest) – kamratrespons – färdig text (refereras till som eftertest). Eleverna var indelade i två olika grupper med olika villkor – den ena gruppen fick skicka en rapport till sin lärare kring vilka av kamraternas kommentarer de hade tagit till sig och hur de hade arbetat med sin text utifrån kommentarerna, medan den andra gruppen inte behövde skicka någon sådan rapport till sin lärare. Båda grupperna arbetade dock med kamratrespons. Kvaliteten på för- respektive eftertest avgjordes av två forskningsassistenter som oberoende av varandra gick igenom och bedömde alla texter utifrån en bedömningsmatris (interbedömarreliabilitet r=.74, p<.01.). I denna studie, till skillnad från i en studie av Gielen m fl (2010b), fann man inga kvalitetsskillnader i texterna beroende på om eleverna hade varit i den ena eller den andra gruppen. Man diskuterar om detta möjligen har att göra med att kamratresponsen i denna studie gavs under kortare tid än vad som var fallet i den andra studien. Resultaten visade, bortsett från det, att i de fall synpunkter från en kamrat är motiverade av kamraten kan detta leda till förbättrade prestationer för de som hade låga resultat på förtesten. För elever med höga resultat på förtesten kunde ingen sådan förbättring konstateras.

Självbedömning och självreglering

Kamratbedömning, som vi just behandlat, kan förstås som ett steg på vägen mot utvecklandet av självbedömande och självreglerande elever. Vikten av att elever utvecklas mot att bli goda självbedömare betonas av Black och Wiliam (1998). De bygger vidare på tankar från Sadler (1989) och Ramaprasad (1983) och menar i enlighet med dem att feedback handlar om att överbrygga gapet mellan individens nuvarande nivå och den önskvärda nivån. Sådan feedback kan dock inte efterlevas utan att eleven har utvecklat en egen förståelse av såväl den egna nivån som den önskvärda. På basis av detta hävdar Black och Wiliam (1998) att ett medvetet arbete med att utveckla den självbedömande förmågan hos elever är en nödvändig förutsättning för effektivt lärande.

Man skiljer ofta på de besläktade begreppen *självreglerat lärande* och *självbedömning*. Självreglerat lärande beskrivs som en process där eleven sätter upp mål för sitt lärande och sedan försöker övervaka, reglera och kontrollera kognition, motivation och beteende i syfte att nå dessa mål (Pintrich, 2000). Självbedömning är en formativ bedömningsprocess där eleven reflekterar över kvaliteten i sitt arbete, söker avgöra i vilken mån det svarar mot uppställda mål och kvalitetskriterier och sedan reviderar sitt arbete utifrån det (Andrade & Boulay, 2003). Andrade (2010) menar att man förenklat kan säga att studier kring självreglerat lärande har fokuserat på hur elever hanterar lärande*processer*, medan studier kring självbedömning har fokuserat på elevers bedömning av *produkterna* av sitt lärande. Oavsett om fokus är på process eller produkt handlar det om att involvera eleverna i reflektioner över sitt lärande och träna dem i att bli aktiva aktörer och ägare av sitt eget lärande.

Baserat på en modell av Goodrich (1996) listar Andrade (2010, s 93) åtta nyckelelement som samtliga behöver finnas med om elever ska utveckla självreglerande egenskaper och bli goda självbedömare:

1) Medvetenhet om värdet av självbedömning
2) Tillgång till tydliga bedömningskriterier
3) En specifik uppgift att bedöma
4) Modeller för självbedömning

5) Direkta instruktioner och stöd i att göra självbedömningen
6) Övning
7) Antydningar om när det är lämpligt att självbedöma
8) Möjlighet att revidera och förbättra prestationer

Om arbete med självbedömning inte fungerar tillfredsställande beror det, enligt Andrade, ofta på att en eller flera av ovanstående element inte finns på plats. Vidare lyfter Andrade fram några exempel på väl fungerande självbedömningsmetoder, bland annat projektbaserad portfoliobedömning, där såväl processorienterad självreglering som uppgiftsorienterad självbedömning är del. Trafikljusmetoden, som beskrivs av Black m fl (2003), där olikfärgade muggar användes av elever för att indikera sin förståelse av ett visst undervisningsmoment, eller genom att med olika färger markera sin förståelse av frågor på en text eller liknande, nämns också av Andrade som ett exempel på träning i självbedömning och självreglering.

Liksom i fallet kamratbedömning är även studier kring självbedömning i grundskolekontext få i våra sökningar. I en artikel av McDonald och Boud (2003) talas om ”den omfattande litteratur som finns kring självbedömning” (s. 210, vår översättning), men hänvisningarna är uteslutande till äldre studier/litteratur som faller utanför ramarna för våra inkluderingskriterier. McDonald och Bouds egen studie är från 2003, och undersöker huruvida studenters prestationer förbättras av medveten träning i självbedömning. 256 studenter som tränats i självbedömning jämfördes med en kontrollgrupp som inte fått samma träning. Man fann signifikanta skillnader mellan de båda grupperna, där den grupp som fått träning uppvisade bättre prestationer i samtliga av alla de olika skolämnen som berörts. Författarna diskuterar resultaten som en indikation på att systematiskt arbete med självbedömning ger positiva effekter på lärande, men menar samtidigt att man bör vara medveten om att studien genomfördes under speciella och gynnsamma förhållanden. Det finns många faktorer som kan ha bidragit till att interventionen fick så god effekt, och det är inte möjligt att särskilja inverkan av varje enskild faktor. Forskarna, med mycket goda kunskaper och erfarenheter kring självbedömning, var exempelvis de som tränade såväl elever som lärare i självbedömningens konst. Lärare i experimentgruppen kan också ha varit mer motiverade och engagerade än de som undervisade i kontrollgruppen, eftersom de kände att de gjorde något nytt och innovativt. Det ska i sammanhanget också tilläggas att detta var en studie som genomfördes i high-school-miljö.

Clark (2012) diskuterar i en forskningsöversikt baserad på 199 genomgångna artiklar självreglerat lärande (selfregulated learning/SRL) som en central aspekt av formativ bedömning. Översikten ger vid handen att självreglerande elever har en hög self-efficacy[21], vilket i sin tur stödjer tillägnandet av god studieteknik. Självreglerande egenskaper utvecklas, enligt Clark, hos eleverna när de aktiveras i de meta-kognitiva sub-processerna planering, uppföljning och kritisk reflektion. Clark menar att framtida forskning behöver riktas mer mot frågor som rör *under vilka omständigheter* elever bäst tillägnar sig sådana egenskaper.

Matriser

Matriser är ett verktyg som används av lärare och/eller elever i syfte att tydliggöra mål och bedömningskriterier. Då matriser används av elever är de att betrakta som en del av kamrat- och självbedömning. Panadero och Jönsson (2013) har gjort en översikt kring användandet av bedömningsmatriser i formativt syfte där totalt 21 studier ingår. Det konstateras att tidigare studier inte varit entydiga i sina slutsatser, och översikten syftar därför till att försöka finna svar på hur matriser som används i formativs syfte kan mediera förbättrade elevprestationer, och vilka faktorer som framstår som modererande för de effekter man kan få av matrisanvändning. På basis av de 21 studierna sammantaget fann man fem huvudsakliga kategorier som svarade mot den första frågan: hur matriser kan mediera förbättrade elevprestationer. Den första har att göra med att matriserna ökar transparensen genom att tydliggöra för eleverna vad som förväntas av dem. Relaterad till den första är den andra kategorin som handlar om att matriser reducerar elevers ängslan. Om förväntningar

[21] Self-efficacy kan översättas som omfattningen eller styrkan av ens tro på den egna förmågan att utföra uppgifter och nå mål

klargörs reduceras elevernas "prestationsångest", vilket i sin tur visat sig höja deras resultat. Tredje kategorin handlar om att feedback-processen underlättas om man har matriser att luta sig emot, både avseende feedback från lärare till elev och feedback elever emellan. Matriser har också visat sig öka elevers self-efficacy, deras tilltro till sin egen förmåga att prestera bra. Det refereras också till studier som visar att matrisanvändande stödjer utvecklandet av självreglerande förmågor, såsom planering av det egna arbetet och självbedömning.

Flera av studierna tar sin utgångspunkt i intervjuer med studenter och rör studenternas uppfattningar av olika matrisrelaterade fenomen. Relationen mellan uppfattningar och faktiska prestationer har studerats i vissa fall, men inte i alla.

Kring frågan som rör modererande faktorer har man i huvudsak funnit fyra kategorier. Den första handlar om att matriser tycks fungera bäst tillsammans med metakognitiva aktiviteter som exempelvis självbedömning. Skolnivå tycks också spela roll, liksom under hur lång tidsperiod matriserna används. Matriser tycks fungera även inom kortare segment i högre utbildning, medan en längre tids användning – där eleverna ges tid att aktivt arbeta med matriserna - tycks krävas på grundskolenivå för att få en god effekt. I vissa av studierna tycks kön vara en faktor som påverkar den effekt man får av matriser: flickor rapporterar högre self-efficacy än pojkar. Vilket ämne matriserna används inom tycks däremot spela mindre roll, studierna rapporterar om goda effekter i varierande ämnen.

Panadero och Jönsson diskuterar även det faktum att matriser och matrisanvändande kritiseras för att leda till ett instrumentellt lärande. De menar, i analogi med Biggs (1999) att en viss instrumentell syn på lärande är oundviklig från ett elev-/studentperspektiv, och att det inte behöver vara negativt. Som vi ser det är detta ungefär samma diskussion som att 'teaching to the test' inte behöver vara negativt om testen är väl genomtänkta och välkonstruerade. Samma diskussion kan också skönjas kring begreppet *assessment as learning*, som av vissa skrivs fram som negativt (t.ex. Torrance, 2007), medan andra (t.ex. Clark, 2012) skriver fram det som något positivt.

E-lärande/IKT

E-lärande är ett begrepp som refererar till användningen av elektroniska medier, pedagogisk teknik och information och kommunikationsteknik (IKT) i utbildning. Ett växande intresse för hur modern teknik kan stödja arbetet med formativ bedömning märks i våra sökningar framför allt från 2008 och framåt.

Den tidigaste studien inom detta område som finns bland de artiklar vi har gått igenom är en översikt av Hunt och Pellegrino (2002) där ämnet är hur nya teknik-baserade lärmiljöer kan underlätta arbetet med formativ bedömning i klassrummet. Författarna exemplifierar hur klassrumsmiljöer med ny teknik kan erbjuda möjligheter att följa och ge återkoppling på elevers prestationer på ett helt annat sätt än tidigare. Utifrån utgångspunkten att formativ bedömning är både svårt och mycket tidskrävande för lärare menar författarna att en fruktbar väg framåt skulle vara att avlasta lärarna genom användandet av datorteknik. De beskriver programmet DIAGNOSER som är byggt utifrån 'aspekt-baserad' undervisning. Aspekt-baserad undervisning beskrivs innefatta två stadier: identifiering av elev-aspekter och presentationen av situationer som ska leda eleverna bort från användandet av en begränsad/begränsande repertoar av aspekter i riktning mot ett mer 'expert-liknande' sätt att resonera.

När ett visst undervisningsområde introduceras får eleverna svara på ett antal inledande frågor om ett ämne (multiple-choice). Dessa frågor baseras på forskning kring vilka aspekter som kan tänkas visa sig i elevernas (naiva) förståelse av ett ämne (t.ex. "Om en kropp väger 2 kg under normala rumsomständigheter, skulle den väga mer eller mindre i ett rum där lufttrycket var dubbelt så stort?"). Efter detta får eleverna en fråga med ett antal resonerande alternativ, som ska visa varför eleverna har svarat som de har gjort. Dataprogrammet går sedan in i feedback-mode. Eleverna får återkoppling på om de har svarat rätt på de inledande frågorna och om svaren på de resonerande frågorna överensstämde med svaren på de inledande frågorna. Eleverna får också förslag på hur de ska gå vidare för att utmana sitt tänkande. När en hel session är slut får både elever och lärare en summering av hur det har gått. Denna summering är tänkt att vara till hjälp för lärarens kommande undervisning i ämnesområdet.

Författarna är noga med att betona att det är en komplicerad process att utveckla den här typen av program, då det krävs stor kunskap kring vilka olika aspekter som kan komma upp och hur dessa kan kopplas samman

med mer eller mindre effektiva sätt att resonera. Mycket forskning krävs om sådan teknologi ska kunna användas på ett gynnsamt sätt, menar Hunt och Pellegrino (2002).

I ett flertal av de asiatiska studier som fanns med i utfallet av våra sökningar beskrivs begreppet formativ bedömning som betydandes feedback till elever. Den som ger feedbacken är i dessa studier en dator eller någon annan mobil enhet. Ofta handlar det om att eleverna får svara på frågor och får omedelbar feedback i termer av ifall de har svarat rätt eller fel. Är svaret felaktigt får man någon form av repetitiva uppgifter eller en uppmaning att läsa på, och om svaret är rätt går man vidare till svårare uppgifter. Flera av studierna liknar varandra och får här exemplifieras med hjälp av en studie av Wang (2007). Denna studie omfattar 503 sjundeklassare i Tiawan, där eleverna fått arbeta med så kallad webbaserad formativ bedömning innehållandes sex strategier. Eleverna loggar in i ett program, får svara på ett antal frågor som kommer i slumpvis ordning och får sedan antingen repetera eller rätta till felaktiga svar. En fråga måste besvaras rätt tre gånger innan man anses ha klarat frågan. Feedbacken består i att eleverna får veta antalet rätta svar direkt efter att de har gjort klart testet och att de via datorprogrammet får tips om referensmaterial där de kan läsa mer. Inbyggt i programmet finns en funktion där man kan se hur kamrater har klarat testet och hur många försök som krävdes innan de klarade det. Tävlingsmomentet ses som ett extra stimulus. Efter att man har klarat testet belönas man med ett antal stjärnor motsvarande prestationsgraden samt animerade gratulationer. Detta beskrivs också som en typ av feedback som ska ge eleven känslan av att ha lyckats. Elevernas kunskaper mättes med för- och eftertest, och de jämfördes med en kontrollgrupp elever som haft traditionell undervisning och gjort tester med papper och penna. Slutsatserna i studien är att lärande blir effektivare för elever som finns i den miljö som använde webbaserad formativ bedömning jämfört med elever i den traditionella miljön.

En annan typ av studie, där interventionen är mer genomgripande, har gjorts av Ciampa och Gallagher (2013), som har genomfört en fallstudie kring användande av mobila enheter som del av undervisningen i grundskoleklassrum på en skola i Kanada. Skolan var relativt liten (250 elever) och samtliga 14 lärare och rektorn deltog i det projekt som innebar att lärande via iPod introducerades. Deltagarna deltog i åtta ”professional learning sessions” (Ciampa & Gallagher, 2013, s 313) där de inledande sessionerna handlade hur de mobila enheterna fungerade, samt hur man använde ett antal appar vars innehåll kunde kopplas till läs- och skrivinlärning, matematik, franska, musik, geografi och historia. I detta ingick att lärarna, utifrån apparna, planerade arbetsområden som kunde kopplas till läroplanen. De sessioner som pågick under tiden som de mobila enheterna och apparna användes syftade till att lärarna kontinuerligt skulle utvärdera och lära av varandra, samt få handledning av de externa mentorer som var inkopplade i projektet. Samtliga lärare och rektorn intervjuades efter hela projektets genomförande, och föräldrarna till skolans elever fick besvara en enkät, där frågorna handlade om de olika aktörernas uppfattningar av hur integreringen av mobil teknologi i undervisningen påverkar elevernas/barnens lärande. Det framkommer att samtliga aktörer ser stora vinster med att använda den mobila teknologin i lärandesituationen, och att lärarna anser att formativ bedömning underlättas. Endast i ett avseende uttrycks explicit att det är formativ bedömning det handlar om, men i vår tolkning kan många av de vinster lärarna lyfter fram kopplas till det som vi tidigare har definierat som existerandes under paraplytermen formativ bedömning. Lärarna säger sig ha upplevt markant förändrade roller och en förändrad klassrumsdynamik, där elever ofta klev in i rollen som undervisare/den som förklarade och läraren fick rollen som novisen/den som ställde frågor. Dessutom skedde ofta spontan kamratundervisning och kamratfeedback, då eleverna tog hjälp av varandra. Elevernas motivation beskrivs ha ökat kraftigt, vilket man bland annat ser som en effekt av det inbyggda tävlingsmomentet (mot sig själva) och att de via apparna får omedelbar feedback. Lärarna, liksom föräldrarna, beskriver att eleverna blir mer självständiga och självstyrande i sitt lärande när de använder de mobila enheterna. Motivationen beskrivs dessutom bidra till att sudda ut de gränser som annars ofta finns mellan det som sker i skolan och det som sker utanför. Elevernas lärande via användandet av undervisande appar visade sig pågå även utanför skolan och i hemmen. Det som uttryckligen beskrivs som kopplat till formativ bedömning är när lärarna via ipaden ser att en elev i exempelvis en quiz endast har ett fåtal rätt, på ett diskret sätt kan göra en notering om detta för att senare ta den eleven åt sidan och arbeta individuellt med denne. Lärarna beskriver, vidare, att arbete utifrån de mobila apparna är inkluderande i termer av att alla elever engageras, oavsett förmåga eller svårigheter. Här lyfts exempelvis fram att eventuell bristande förmåga/förståelse inte blir synlig på samma sätt för kamrater som det ofta blir i traditionell undervisning, att elever som normalt hade svårt att koncentrera sig med detta arbetssätt kunde

koncentrera sig längre, och att finmotoriska svårigheter som exempelvis kan påverka skrivande blev osynliga. Dessutom framhålls att den digitala klyftan minskar då man ger eleverna tillgång till den här typen av teknologibaserad undervisning i skolan, då även de elever som inte hade tillgång till motsvarande mobila enheter i hemmet blev i stort sett lika skickliga användare som de som hade tillgång i hemmet. I föräldraenkäten framkommer dock en oro från föräldrahåll att den här typen av teknologibaserad undervisning ska bidra till minskad likvärdighet i termer av att de som har tillgång till tekniken hemma får fördelar och ett försprång. Föräldrarna är generellt positiva, men uttrycker också en oro över att eleverna kan missa något om inte traditionella (papper och penna) metoder används som ett komplement.

Påfallande i studien är att inga svårigheter eller negativa aspekter över huvud taget lyfts fram från lärar- eller rektorshåll. I studien framgår inte hur intervjuerna varit upplagda eller hur frågorna har ställts, så det går inte avgöra vad det beror på att inga problematiska aspekter beskrivs av lärare och rektor. Forskarna själva problematiserar dock några aspekter. Genomförandet av projektet har skett på en liten skola, det har involverat alla, lärarna har fått kontinuerlig handledning, och tid har avsatts inom ramen för deras tjänster att kunna delta i denna handledning. Professionellt lärande har så att säga byggts in i projektet, vilket kan anses vara en viktig faktor för att positiv utveckling ska ske (jfr Gardner m fl, 2011, under nästa rubrik). Man diskuterar också de positiva effekter man har sett i termer av att de kan vara kopplade till 'nyhetens behag'. Elevernas och lärarnas entusiasm kan vara en effekt av att det händer något nytt, varför man menar att framtida forskning behöver innefatta longitudinella studier som fokuserar mer på vilken påverkan den här typen av teknologibaserad undervisning kan ha i det långa loppet. Studien bygger dessutom på lärares, rektors och föräldrars uppfattningar och innefattar inga mätningar av elevers lärande eller jämförelser med någon kontrollgrupp.

Helt klart är att vi i och med modern teknik och utvecklandet av den här typen av program får en ny dimension att fundera över i relation till den definition av formativ bedömning som varit rådande fram till nu. De IKT-relaterade studier som vi har gått igenom kan – möjligen något förenklat – tolkas som representerandes två olika infallsvinklar på hur tekniska hjälpmedel kan förstås i relation till formativ bedömning. Den ena infallsvinkeln, som bland annat representeras av ovan beskrivna studie av Wang (2007) kommunicerar en förståelse av datorn som en egen aktör. I den figur över nyckelstrategier som Wiliam och Thomson (2007) har skapat (se även Black & Wiliam, 2009) skulle vi i så fall inte längre enbart kunna tala om de tre aktörerna lärare, kamrat och elev som involverade i formativa processer. Datorn kommer in i bilden som varandes den som ger eleven, och ibland - som i den typen av program som beskrivs av Hunt och Pelligrino (2002) - även läraren feedback:

Aktörer	**Vart den lärande är på väg**	**Var den lärande befinner sig just nu**	**Hur gapet mellan nuläge och mål ska överbryggas**
Lärare	Tydliggöra mål och framgångskriterier för lärande	Skapa effektiva klassrumsdiskussioner och andra situationer som ger bevis för lärande	Ge återkoppling som tar den lärande framåt
Dator	Tydliggöra mål och framgångskriterier för lärande	Skapa situationer som ger bevis för lärande	Ge återkoppling som tar den lärande (och ibland även läraren) framåt
Kamrat	Förstå och dela mål och framgångskriterier för lärande	Aktivera elever som resurser för varandra	
Elev	Förstå och dela mål och framgångskriterier för lärande	Aktivera elever som ägare av sitt eget lärande	

Figur 4. Nyckelstrategier för formativ bedömning (efter Wiliam & Thompson 2009 – vår översättning och modifiering)

En annan infallsvinkel på hur tekniska lösningar kan förstås i relation till formativ bedömning representeras av Ciampas och Gallaghers ovan beskrivna kanadensiska studie, och av studier av bland annat Pachler m fl (2009) och Daly m fl (2010). De senare undersöker och problematiserar begreppet formativ e-bedömning i ljuset av ett projekt som syftar till att identifiera existerande praktiker där teknik på olika sätt kan *bidra till* formativ bedömning. De framhåller att ingen teknik-baserad bedömning *i sig själv* är formativ (medan däremot nästan all teknologi kan användas formativt), det är lärare, kamrat och elev i egenskap av mänskliga aktörer som i sista änden bestämmer de formativa effekterna av att använda teknologi. Framtida forskning, enligt Daly m fl (2010) bör därför inriktas på hur tekniska resurser kan fungera formativt *i samverkan* med andra kognitiva resurser, sociala såväl som individuella.

Vi förutspår detta som en av de stora frågorna inom fältet formativ bedömning framöver. Frågor som blir högintressanta (i synnerhet i relation till förståelsen av det tekniska hjälpmedlet som en egen aktör) är exempelvis vilken typ av feedback som kan ges och med vilken kvalitet den kan ges. En annan stor och viktig fråga handlar om konsekvensen av digitala prov och digital feedback för lärares professionalism, med tanke på att formativ bedömning ska tjäna syftet att utveckla inte bara eleven, utan även lärarens förståelse av hur undervisningsprocesser kan anpassas för att möta elevers behov. En tredje intressant fråga, kopplad till den förra, är naturligtvis också vem som utvecklar de program som används i detta syfte.

Verktyg – övrigt

De allra flesta studier som rör verktyg är koncentrerade till feedback, kamrat- och självbedömning, matriser och e-lärande/IKT. Det finns enstaka studier som behandlar det som enligt Wiliam och Thompsons figur över nyckelstrategier är kopplat till strategi två – att skapa effektiva klassrumsdiskussioner och andra situationer som ger indikationer på att lärande sker. Ett exempel är en studie av Ruiz-Primo och Furtak (2007), som med hjälp av videoobservationer utifrån ESRU-modellen studerat lärares informella formativa bedömningspraktiker i NO-undervisning i tre mellanstadieklassrum i USA. Totalt 30 lektioner filmades. Modellen innebär att man studerar hur läraren ställer en fråga (Elicit), hur eleven svarar (Response), hur läraren bekräftar elevens svar (Recognize) och hur läraren sedan använder informationen (Use) för att stötta elevens lärande. Forskarna använde sig av för och eftertest för att kunna säga något om påverkan på elevers lärande, och fann att när lärare använde sig av hela ESRU-cykeln i klassrumssamtal ledde detta till att elever fick bättre resultat. Anmärkningsvärt var att studien visade att den av de tre lärare som hade minst formella ämneskunskaper var den som var överlägsen de andra två i att använda sig av hela ESRU-cykler, och den lärare med längst lärarerfarenhet var den som i lägst grad använde sig av hela cykler. Mer storskaliga studier skulle dock behövas för att kunna dra några generella slutsatser kring hur olika kontextuella faktorer påverkar lärares mottaglighet och förmåga att ge respons på elevers svar i klassrummet.

Sammanfattning – Verktyg

Nyckelstrategierna, det vill säga tydliggörandet av mål och bedömningskriterier, skapandet av klassrumssituationer som genererar indikationer på elevers lärande, feedback från lärare till elev, kamratbedömning och självbedömning utgör det som vi kallar för verktyg för formativt klassrumsarbete/formativ bedömning. Många studier faller inom denna kategori, med en tydlig betoning på sådana som handlar om effekten av feedback från lärare till elev eller från elev till elev (kamratbedömning). Flera studier visar att formativa interventioner har positiv effekt på lärandet, men det finns också exempel på studier där man inte kan konstatera sådana effekter.

En problematik i dylika studier är att interventionerna ofta genomförs i kontexter som är utöver det vanliga, exempelvis i termer av att särskilda utbildningsinsatser för elever och/eller personal görs som del av interventionen, eller att valda skolor/klasser/elever kommer från särskilt gynnsamma förhållanden. En annan problematik som gör det svårt att som läsare förstå effekten eller bedöma värdet av de interventioner som genomförts, är att de för- och eftertester som används när elevers lärande ska mätas varken redovisas eller problematiseras. Meta-studier som handlar om generella effekter av formativ bedömning har dessutom kritiserats för att beräkningarna är gjorda på många disparata företeelser och inte problematiserar kontextens betydelse/modererande faktorer. Påtagligt, med tanke på att formativa nyckelstrategier i hög utsträckning

arbetas med i grundskolemiljö, är också att översiktsstudier som på ett mer generellt plan talar om effekten av exempelvis kamratbedömning eller självbedömning i mycket hög utsträckning är gjorda i högre utbildning.

En tämligen stor andel studier i kategorin verktyg – framför allt från Asien - rör IKT-relaterade verktyg för formativ bedömning. Då formativ bedömning skrivs fram som resurskrävande ur lärarsynpunkt betraktas IKT-verktygen som ett sätt att effektivisera undervisningen. Påtagligt i vissa av dessa studier är att man tenderar att betrakta datorn/den mobila enheten som en *egen aktör* vid sidan av övriga aktörer i klassrummet (lärare och elever). Datorn är den som analyserar elevens ”lärande” (svar på frågor) och ger eleven, och ibland också läraren, feedback. I andra studier positioneras IKT-verktyget som något som kan *bidra* till formativa praktiker, men det betonas att verktyget får sin formativa funktion först i händerna på mänskliga aktörer. Om IKT-verktyget ses som en egen aktör behöver man fundera över frågor som gäller kunskapssyn och lärarens roll; vilket lärande möjliggörs genom användandet av den här typen av verktyg, och vilka blir konsekvenserna för lärares professionalism?

Implementering/Professionellt lärande

En kategori studier handlar om hur formativ bedömning implementeras i praktiken, och problematiserar aspekter av professionellt lärande i relation till detta. Vi har i viss mån redan berört frågan om klassrumsimplementering av formativa praktiker under teori- och policyrubrikerna, och där kunnat konstatera att många anser att sådan implementering behöver stödjas av processer där lärares lärande ges utrymme. Olika typer av *professional learning communities (PLC)* har kommit till som ett sätt att organisera lärare i arbetsgrupper i syfte att främja kollaborativt lärande bland kollegor. Innehållet i studier som på olika sätt adresserar lärares professionella lärande och utveckling presenteras utförligare under denna rubrik.

I projektet *Analysis of Innovations in assessment* (ARIA) som pågick 2006-2008 gjordes en syntes av resultaten i 13 mer omfattande initiativ till förändrade bedömningspraktiker i England, Skottland, Wales och Nordirland utifrån frågorna *Hur kan lärares bedömningspraktik – summativ såväl som formativ – förbättras?* och *Vad underlättar spridningen av lärares förbättrade bedömningspraktiker?* I syntesen (Gardner m fl, 2011) framträdde professionellt lärande som en central aspekt. Man skiljer mellan termerna professionell utveckling och professionellt lärande (jfr också Timperley, 2013), då man ser att det finns exempel på hur lärare på ett ytligt plan kan tillägna sig tekniker utan att för den skull utveckla en fördjupad teoretisk förståelse av de processer formativ bedömning handlar om (jfr Sfard, 1998). Hållbara och verkningsfulla processer karaktäriseras av professionellt *lärande* menar forskargruppen som analyserat de 13 initiativen (Gardner m fl, 2011), varför implementeringsarbete behöver gå hand i hand med utbildning och organiserad reflektion bland lärare och skolledare. Utvecklingen behöver också ske över tid, där regelbundna förståelseskapande träffar varvas med möjligheter att pröva och utveckla bedömningsarbetet i praktiken.

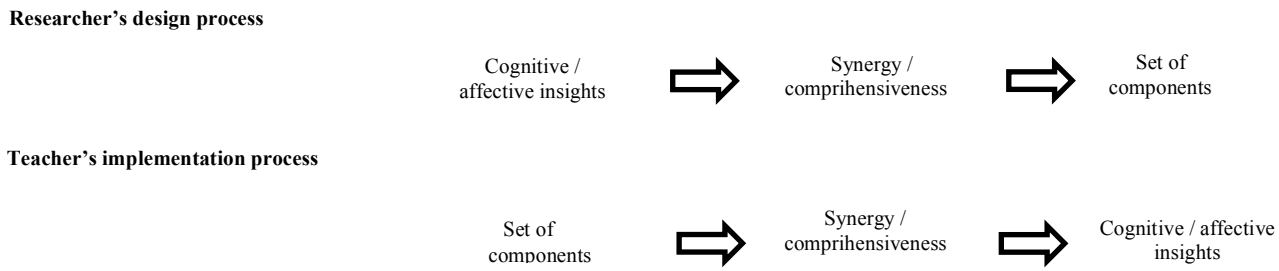
För att nå bästa möjliga spridningseffekt av diverse olika initiativ för förändring i skolan är det vanligt att man använder sig av olika varianter av ’cascade training’ (se Stobart & Stoll, 2005) där representanter från en skola får utbildning och sedan förväntas delge övriga på skolan samma utbildning. Modellen är inte optimal för utvecklandet av en fördjupad förståelse, men kan fungera tillfredsställande under förutsättning att tid för samarbete och samtal säkerställs på den lokala skolan. Gardner m fl (2011) skiljer mellan ”The transmission model” (överföring) och ”The transformation model” (omvandling) (s 110-111) när det gäller spridandet av nya bedömningspraktiker i de studerade initiativen. Överföringsmodellen innebär förenklat att lärare som fått särskild utbildning eller deltagit i pilotprojekt förväntas överföra sina kunskaper till andra lärare som inte deltagit. Överföringsmodellen visade sig inte lyckosam för spridning i de initiativ den använts, då den i låg grad lett till professionellt *lärande* för dem som inte varit med i pilotstadiet. De initiativ som tagit ett bredare grepp med intentionen att alla lärare ska utbildas, och att policy, forskare och praktiker samtliga ska vara involverade redan från planeringsstadiet av förändringsinitiativ, klassas som omvandlingsmodeller och har lyckats bättre med att få till stånd ett professionellt *lärande*.

För att förändringar ska bli hållbara över längre tid krävs att man förstår begreppet hållbar skolutveckling som dynamiskt snarare än statiskt. Detta bottnar i en fördjupad teoretisk förståelse av (formativ) bedömning. På det sättet kan även metoder utvecklas och anpassas samtidigt som de är en ständig del av det kontinuerliga arbetet och inte en tillfällig fluga. Hållbar förändring av bedömningspraktiker är också en kollektiv process som

kräver samsyn mellan olika aktörer på såväl praktiska som teoretiska arenor: forskare, policy, skolledare, lärare och inte minst lärarutbildare.

Gardner m fl (2011) hävdar också vikten av att man redan i planeringsstadiet av förändringsinitiativ/projekt planerar även för professionellt lärande och spridning. I flertalet av de 13 initiativ som studerats har sådan planering inte funnits med från början, vilket medfört att frågor kring professionellt lärande och spridning kommit att hanteras ad hoc. Man kan dessutom konstatera att förändringsinitiativ behöver utgå från en gemensam förståelse bland de olika inblandade aktörerna kring vilken betydelse som avses då man talar om *förbättrade bedömningspraktiker*.

En central fråga i sammanhanget – som vi tidigare berört - rör den relation mellan verktyg/metodik för formativ bedömning å ena sidan och den fördjupade teoretiska förståelsen av begreppet å den andra. Wiliam (2009) hävdar exempelvis att lärare kan agera sig in i fördjupad och forskningsbaserad teoretisk förståelse om de förses med verktyg/metodik. Forskarna börjar i teoretisk förståelse och landar i verktyg, medan processen är den omvända för lärare, enligt figur 5:



Figur 5. Lärares och forskares förståelseprocesser relaterade till formativ bedömning *(Wiliam, 2009, s 28)*

Lärarnas process, när de går från verktyg/metodik till fördjupad förståelse behöver, enligt Wiliam (2009) systematiskt stöd. Teacher Learning Communities (TLC), grupper där lärare träffas och tillsammans reflekterar över sin undervisnings- och bedömningspraktik enligt ett visst givet upplägg, är, enligt Wiliam (2009) en utprövad och lyckosam väg att gå. I en sådan grupp reflekterar man över det formativa arbete man har prövat i klassrummet och får råd och stöd av övriga. Man introducerar också nya element allteftersom, exempelvis via läsning av litteratur. Bennett (2011) menar att det ännu inte finns tillräckligt empiriskt stöd för att modeller som TLC är effektiva, men även han betonar att en djupare förståelse måste utvecklas. Lärarutbildningar har en stor och viktig roll att spela i sammanhanget, men även för redan verksamma lärare krävs tid för att förstå och lära sig att använda/anpassa domänbaserat material för formativ bedömning. Lärare behöver således tid för att reflektera över sina erfarenheter i iterativa cykler av *användning – reflektion – anpassning*, där verktyg/metodik och process blir till en helhet.

Vad gäller former för kollektiv lärarutveckling finns sådana som är stora i vissa delar av världen, men som ännu inte är så utbredda hos oss, trots att de kan sägas hänga samman med de aspekter av formativ bedömning som handlar om att använda bedömningsinformation för att forma undervisning. Det japanska konceptet *lesson study* är ett exempel som är tämligen välkänt. En besläktad form för kollektiv undervisningsanalys, som också är stor i Japan och vissa andra asiatiska länder, men även förekommer i USA, är så kallade *open lessons* (se t.ex. Miyakawa & Winsløw, 2013) eller *public research lessons* (se t.ex. Lewis m fl, 2006), där grundtanken är att lärare utifrån iakttar en kollegas klassrumsundervisning, och sedan deltar i en diskussions-session med läraren (och ibland inbjudna experter) där undervisningen analyseras, problematiseras och teoretiseras. En allmän beskrivning av konceptet *open lessons* kan ge intrycket av ett snävt fokus på en viss lektion i ett specifikt ämne, med specifika elever och under specifika villkor. Miyakawa och Winsløw (2013), som har närstuderat två fall där open lessons har genomförts i matematik, hävdar dock att dessa fall var tydliga exempel på att diskussionen lärare emellan (och därmed potentiella lärdomar) även var av mer allmän karaktär. De konkreta lektionerna och didaktiska metoderna var visserligen av intresse för alla deltagare, men efterföljande

diskussioner rörde också i hög grad teoretiska aspekter av matematikämnet som sådant, och även hur detta kunde relateras till skolans mer allmänna och övergripande mål. Forskarna betonar också, att varken lesson study eller open lessons i första hand syftar till att revidera eller förfina den didaktiska praktiken i den observerade lektionen, utan snarare att utveckla och stärka didaktisk praktik i vidare bemärkelse. De menar att etablerandet av en liknande ”para-didaktisk” infrastruktur, där den gemensamma observationen ligger till grund för efterföljande diskussion och kollektivt lärande, skulle kunna stärka lärares professionella utveckling i många västländer. I såväl Asien som i västvärlden har teknikbaserade plattformar för lärares gemensamma lärande i allt högre utsträckning börjat användas. Dessa kan naturligtvis betraktas som mer tidseffektiva än möten i realtid, och de möjliggör delning och i viss mån kollektiv utveckling av metoder. Mycket tyder dock på att de inte möjliggör samma fördjupade kunskapsutveckling som den som blir möjlig genom observationer och efterföljande diskussioner av realtidsundervisning (jfr Gueudet och Trouche 2009; Winsløw, 2011).

Vi kan konstatera att många tycks vara överens om att tid för kollegialt arbete med bedömningsfrågor är en nyckelfaktor om lärare ska kunna utveckla sina kunskaper och färdigheter med bedömning. Trots detta tycks just tid ständigt vara det som saknas. Atjonen (2014) har studerat finska lärares syn på bedömningspraktiker och konstaterar att gemensam tid – såväl för att diskutera bedömning med kollegor som elever – tycks vara svår att åstadkomma. I hennes studie fick 157 finska grundskollärare svara på en enkät, där de bland annat fick uppmaningen att beskriva bedömningsrelaterade situationer där de själva haft positiva erfarenheter av elevbedömning. På motsvarande sätt fick de också beskriva situationer där de haft negativa erfarenheter. De positiva erfarenheterna var i mycket hög utsträckning relaterade till muntliga bedömningstillfällen och diskussioner med eleverna kring bedömning. Många upplevde dock att tiden inte räcker till för sådana diskussioner, och att man därför ofta tar till ”traditionella” skriftliga prov för att mäta elevers kunskap. Känslan av att sådana bedömningstillfällen inte riktigt var valida i förhållande till den kunskap som efterfrågades var framträdande bland de negativa erfarenheterna. Detta kan jämföras med resultaten i en studie av Kirton m fl (2007), där lärarna också ansåg att arbete med formativa bedömningsmetoder var tidskrävande. Att involvera elever i exempelvis kamrat- och självbedömning, eller att ha mer diskursiva och interaktiva lektioner gav lärarna en känsla av långsammare processer, med en oro över att läroplanens alla områden inte skulle hinnas med som följd.

Atjonen (2014) diskuterar detta dilemma som en effekt av att olika aktörer har olika intressen av bedömning. Davies och Neitzel (2011) ser samma fenomen: i en studie av självreglerat lärande observerades klassrum hos 15 olika mellanstadielärare, och intervjuer genomfördes i syfte att fånga lärarnas uppfattningar om klassrumsbedömning. Resultaten visar att lärarna hade en komplex förståelse av klassrumsbedömning, men att den kontext de arbetade i inte var optimal för att stödja elevernas utveckling av självreglerande förmågor. Lärarna tenderade i stället att se sig själva som de som genomför och styr bedömning i klassrummet, vilket författarna menar är en konsekvens av de konkurrerande publiker som lärare med sitt bedömningsarbete måste tillfredsställa. Det finns en intern publik, som består av lärarna själva och eleverna, och en extern publik, som består av föräldrar och delstaten. Från den externa publiken finns kontinuerliga krav på summerade sammanställningar och resultatmått, vilket gör att lärarna upplever att de summativa bedömningspraktiker de ägnar sig åt blir dominerande på bekostnad av de formativa. I en sådan kontext utvecklas inte heller elevernas förmåga till självreglering, då lärarna inte upplever att det finns utrymme att arbeta med något så tidskrävande som att utveckla elevers metakognitiva förmågor. Ett liknande resonemang för Smith (2011) kring den norska kontexten. Hon menar att duktiga och engagerade lärare som *vill* stödja elevernas lärande genom formativt arbete hindras av såväl externa krav på en viss typ av bedömning, som av det faktum att de känner tidspress av att så måna moment i kursplanerna måste hinnas med. Vår egen erfarenhet säger oss att detta är högst påtagligt även i Sverige, då vi i otaliga möten med lärare hört deras frustration över att det centrala innehållet i de svenska kursplanerna är så omfattande att det känns som att man på ett ytligt sätt behöver stressa igenom områden för att hinna med.

Hermansen och Nerland (2014) har studerat arbetet med att integrera principer och verktyg för formativ bedömning på en högstadieskola i Norge. De förstår bedömning-för-lärande-konceptet som en medierande artefakt i lärares bedömningsarbete, och menar att sätten på vilka en ny artefakt tar plats och ges mening inte är givna, utan snarare starkt kopplade till redan existerande lokala praktiker. Implementeringen av nya koncept kommer därför sannolikt att ta sig olika uttryck i olika skolsammanhang. I denna specifika studie

observerades/intervjuades en grupp ämneslärare som fått i specifik uppgift att vara huvudansvariga för implementeringen av ett bedömning-för-lärande-koncept. Observationer och intervjuer med de åtta lärarna ägde rum sju gånger under en ett och ett halvt år lång period. Resultaten visar att en viktig inriktning i lärarnas strävan var relaterad till att utveckla bedömning-för-lärande-praxis i ”sann bedömning-för-lärande-anda”, med avseende på de generella principer och nyckelstrategier som tidigare har beskrivits i denna översikt. Detta var dock långt ifrån tillräckligt: arbetet med att integrera bedömning för lärande i praktiken handlade om mycket mer än att explicitgöra och anamma generella principer. Mycket arbete handlade om att varje lärare kritiskt behövde utforska sin redan existerande bedömningspraktik. Dessutom krävdes omarbetning av etablerade relationer mellan lärare, elever och föräldrar, då bedömning för lärande bygger på en annan arbetsdelning med mer aktiva elever. I detta hade olika lärare olika utgångspunkter, varför det blev påtagligt att det inte bara går att implementera bedömning-för-lärande som en generell metod. Frågan om vad i de formativa strategierna som skulle tillämpas likadant av alla och vad som skulle kunna lämnas till varje lärare att utforma/anpassa i relation till sig själv och de egna eleverna var ett återkommande dilemma. Hermansen och Nerland (2014) konstaterar att i en tid då föreställningar om kliniska metoder och standardisering av professionell praxis breder ut sig (jfr Evans, 2011), blir frågan om vad ”shared practice” faktiskt innebär i relation till undervisning central.

Harlen (2010) liksom Smith (2011) hävdar att en rad faktorer och nivåer måste samverka för att formativa bedömningspraktiker ska få genomslag i skolan. Det räcker inte att lärare involveras i förändringar, även skolledare och beslutsfattare måste fördjupa sin förståelse av vad som krävs för att formativt arbete ska få genomslag. Smith betonar också, med emfas, att lärarutbildningar inte håller tillräckligt hög nivå med avseende på utbildning i bedömningsfrågor. Hon tar situationen i Norge som utgångspunkt, men betonar att den på intet sätt är unik när det gäller hur blivande lärare utbildas i bedömning. Mansell och James (2009) hävdar exempelvis att en anledning till att brittiska lärare, trots att de sympatiserar med tankarna bakom formativ bedömning, har svårt att implementera formativa praktiker i sina klassrum är att de inte utbildas i detta i tillräckligt hög utsträckning under sin lärarutbildning. I ett forskningsprojekt initierat av dekaner vid fem lärarutbildningsinstitutioner i Norge (Smith, 2009), visade det sig att den utbildning studenterna får i bedömning i allmänhet och formativ bedömning i synnerhet är minimal. I bästa fall behandlas sådana frågor i en föreläsning följt av ett seminarium på ett par timmar, och integreras då och då i ämnesdidaktiska moment. I en annan, mer fördjupad studie, fann man att det är vanligt att lärarutbildningsinstitutioner bara har en person som kan sägas ha formell kompetens och forskningserfarenhet inom bedömning, och de flesta av dessa personer undervisar inte i bedömning i lärarutbildningskurserna. En annan alarmerande upptäckt är att det verkar finnas mycket liten enighet mellan lärarna på lärarutbildningarna kring hur bedömningsfrågor och bedömningskunskap ska göras tillgänglig för studenter. Som följd av detta får studenterna inom ramen för sin utbildning blandade och inte alltid särskilt välgrundade budskap (Andersen m fl, 2010). Den fråga som måste ställas är vem som ska utbilda lärarstudenter och redan verksamma lärare i bedömning, när det finns en brist på kompetens i bedömningsfrågor på lärarutbildningsfakulteterna. Smith (2011) menar att institutionerna måste initiera utbildningar bland lärarutbildare *inom* fakulteterna, där erfarenheter kan delas, externa experter kan kallas in för stöd, och där praxis kan analyseras och utvecklas. Masterprogram inom bedömning behöver utvecklas, och doktorandprojekt i bedömning bör ges finansiellt stöd. En stor grupp av lärarutbildare med expertis inom bedömningsfrågor måste med nödvändighet uppnås för att vi ska kunna ge lärarstudenter den nödvändiga kunskap och kompetens som de behöver ha redan från början av sin lärarbana.

Sammanfattning – Implementering/Professionellt lärande

I stort sett alla som forskar och skriver på temat professionellt lärande i relation till formativ bedömning är överens om att klassrumsimplementering av formativa praktiker behöver stödjas av processer där lärares (och skolledares) lärande ges utrymme. Gardner m fl (2011) skiljer mellan professionell utveckling och professionellt lärande, då det i studier visat sig att viss utveckling i termer av anammande av metoder för formativ bedömning är möjlig även om fördjupad reflektion och teoretisk kunskap saknas. Man har dock sett att bristen på fördjupad förståelse ökar risken för att så kallade pseudo-formativa praktiker (Birenbaum m fl, 2011) utvecklas, där förståelsen av formativ bedömning blir instrumentell och kopplad till ett mer rituellt arbetande med en rad metoder. Olika typer av *professionl learning communities*, där lärare organiseras i

arbetsgrupper i syfte att främja fördjupad teoretisk förståelse och kollaborativt lärande, är vanliga då formativa praktiker ska implementeras. Detta bör kompletteras med utbildningsinsatser av experter, och eventuellt andra metoder för att exempelvis observera och diskutera konkreta lektioner.

Forskare tycks överens om att en rad faktorer och nivåer av aktörer måste samverka för att formativa bedömningspraktiker ska få genomslag i skolan. Det räcker inte att lärare involveras i förändringar, även skolledare och beslutsfattare måste fördjupa sin förståelse för vad som krävs för detta. Lärarutbildningar spelar dessutom en central roll, och forskning (exempelvis Smith, 2009) har visat att inte tillräckligt många lärarutbildare har den kunskap i bedömningsfrågor som krävs för att kunna undervisa studenter på ett adekvat sätt.

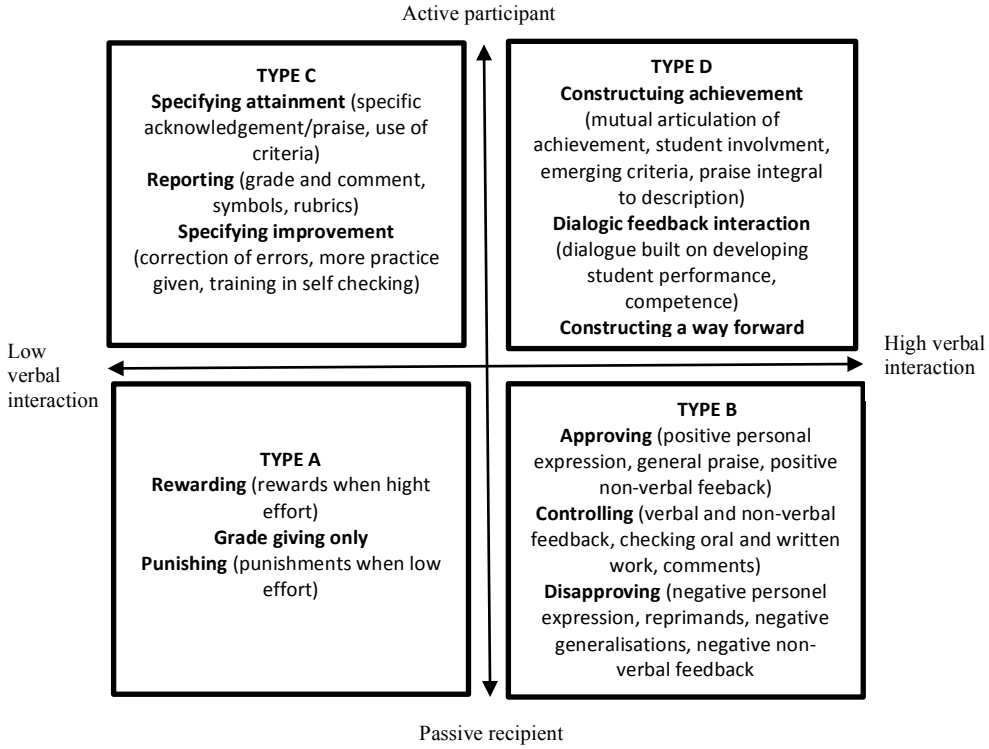
Elevers uppfattningar

I den kategorisering som gjorts i vår översikt är studier som rör elevers uppfattningar och erfarenheter av formativ bedömning den klart minsta kategorin. Elever har beskrivits som den "felande rösten" i pedagogisk forskning (Cook-Sather, 2002) generellt, och forskningen kring formativ bedömning tycks inte vara ett undantag. En handfull studier har dock gåtts igenom för den här översiktens räkning.

Peterson och Irving (2008) har genomfört fokusgruppsintervjuer med 41 högstadieelever från fyra olika högstadieskolor i Nya Zeeland i syfte att skapa kunskap kring elevernas uppfattningar av bedömning. Eleverna uppfattade att syftet med bedömning var att ge dem feedback, och att lärare och föräldrar ska kunna se vilka nivåer de ligger på. Eleverna gav generellt inte uttryck för att de tänkte i termer av lärandemål eller var medvetna om att den feedback de fick var riktad mot ett specifikt mål, varför Peterson och Irving menar att man endast i låg grad kunde tala om att feedbacken utvecklade elevernas självregulativa förmåga. Bedömning uppfattades av eleverna i vissa situationer som mindre relevant, till exempel när den inte uppfattades som rättvis, när den inte innefattade ett betyg, eller när den gjorts av en kamrat och inte av läraren. Eleverna gav starkt uttryck för att betyget var det som fick dem att förstå hur de låg till, samt att det var betyget som motiverade dem. Eleverna jämförde inte med egna tidigare prestationer för att avgöra om ett betyg var bra och om de var nöjda med sig själva eller inte, det bestämdes i stället i jämförelser med andra elevers betyg.

Ytterligare fyra studier från vår sökning handlar om elevers uppfattningar av feedback. Williams (2010) har studerat 56 nya zeeländska 12-13-åringars uppfattningar av effektiv feedback genom enkäter (där olika givna feedbackformer fick kryssas för och rangordnas och tillägg var möjliga att göra) och fördjupande intervjuer. Genomgående upplevde eleverna feedback som mest effektiv/givande i muntliga en-till-en-situationer mellan lärare och elev. Näst viktigaste feedbackformen var enligt eleverna betygssättning av arbeten, då de upplevde det var viktigt att veta var de låg nivåmässigt. Långa skriftliga kommentarer på arbeten upplevdes som effektiva av framför allt flickor. Endast ett fåtal tyckte att feedback som ges till hela klassen samtidigt var till någon hjälp, och ingen tyckte att feedback innehållandes korta kommentarer (som *well done* eller *good work*) gav dem något. Genomgående ansåg flickorna i högre grad än pojkarna att de fick feedback som hjälpte dem att förbättra sina prestationer.

Gamlem och Smith (2013) har gjort klassrumsobservationer och djupintervjuat 11 mellan- och högstadieelever i Norge. Baserat på elevernas utsagor har de skapat en fyrfältsfigur (figur 6):



Figur 6. En typologi för feedback i klassrum *(Gamlem & Smith 2013, s. 162).*

På ett generellt plan upplevs feedback som positiv när den kommunicerar att ens prestation och ansträngning är godkänd, samtidigt som man får veta vad som kan göras för att förbättra prestationen ytterligare. Feedback upplevs som negativ när läraren talar om för eleverna att de kunde ha gjort ett bättre jobb, samtidigt som de själva tycker att de har gjort sitt bästa, eller när lärarens kommentarer endast i generella termer säger att eleven bör jobba bättre eller hårdare i framtiden. ”If I knew more I would have written it – I don’t know what more to write. Teachers should tell me what is missing” säger exempelvis en elev (Gamlem & Smith, 2013, s. 159). Eleverna upplever också feedback som negativ om de får kommentarer kring förbättring, men inte ges tid till att arbeta med dem eller tid till ytterligare uppföljning av läraren. På basis av detta menar författarna att kraften i feedback som verktyg inte enbart kan sägas ligga i när och hur den ges, utan i hög grad också i när och hur den tas emot.

Vad gäller typologin av feedbacktyper ovan ger observationerna och intervjuerna vid handen att typerna A och B är de klart vanligaste i de här elevernas klassrumssituationer, och att typ D förekommer mycket sällan. I båda typerna A och B är eleverna mer eller mindre passiva mottagare av feedbacken. Typ B uppfattas av eleverna som positiv för motivationen och relationen mellan lärare och elev, men den uppfattas inte som någon direkt hjälp till bättre prestationer.

Typ C och D involverar eleven, men typ C ges ofta i slutet av ett arbete utan att tid ges för tillämpning i termer av möjligheter att förbättra aktuell prestation. Typ D beskrivs som den mest effektiva formen av feedback, då den involverar dialog och ger möjlighet till tillämpning här-och-nu. Observationerna och intervjuerna med elever visar dock att denna typ av feedback förekom mycket sparsamt.

Lee (2008) har genom klassrumsobservationer och enkäter till 76 högstadieelever i två olika klasser i Hongkong studerat hur eleverna uppfattar skriftlig feedback på uppsatser i språk. I analysen har en åtskillnad gjorts mellan hög- och lågpresterande elever. Resultaten visar att de högpresterande eleverna i hög grad ville ha feedback som innehöll markeringar av språkliga fel och betyg på prestationen. De lågpresterande eleverna efterfrågade i högre grad också kommentarer från läraren kring hur de kunde förbättra sina prestationer. Den feedback som försiggick i klassrummen, såväl som den som gavs i anslutning till uppsatserna, gavs från lärare till elev och innehöll ingen interaktion eller dialog parterna emellan. Lee menar att detta positionerar eleverna som passiva mottagare som snarare görs beroende av läraren än att de tränas i att bli självständiga och självreglerande.

Elevperspektiv på kamratbedömning beskrivs också i en fallstudie av Mok (2010), där fyra lågpresterande elever och en engelsklärare från Hongkong har observerats och intervjuats vid tre olika tillfällen. Eleverna ger starkt uttryck för att de – på grund av att de upplever sina egna engelskkunskaper som dåliga – känner sig otillräckliga i kamratbedömningssammanhang. Samtliga fyra ger uttryck för att dessa situationer därmed blir mycket besvärande för dem, som i följande exempel:

> **Kylie:** It's because our English proficiency is very low and we can't really assess the performance of others, for example, we don't know if grammar is good or not /…/
>
> **Yale:** You will have to think about how to translate what you want to say into English and say it… If not, you will be wasting a lot of time. But, sometimes you don't even know how to pronounce the word you want to use. What can you do then? /…/
>
> **Veronica:** So, I mainly write down things like body language and eye contact, because I think, in terms of ability, we are not good enough to assess others' performance.
>
> **Kylie:** We are not good enough to do so. (s 235).

Studien, i egenskap av fallstudie som enbart involverar fyra elever, gör inga anspråk på att säga något om hur vanliga sådana här uppfattningar är bland elever i allmänhet, men den är exempel på en aspekt som knappt har studerats alls, och har därmed ett klart värde. Som tidigare nämnts är studier som tar elevperspektiv ytterst få. Formativ bedömning talas i allmänhet om i positiva termer, och de olika nyckelstrategier som skrivs fram av bland annat Wiliam och Thompson (2007) är vida accepterade och etablerade bland lärare som varandes positiva för elevers utveckling. Elevröster, som dem citerade ovan, är viktiga bidrag till diskussionen om kamratbedömning, då de illustrerar att det inte går att ta för givet att kamratbedömning enbart har positiva effekter.

I en annan studie (Harris & Brown, 2013) har tre Nya Zeeländska lärare och 50 mellan- och högstadieelever observerats under en veckas klassrumsarbete i ämnena matematik och engelska i syfte att utröna vilka själv- och kamartabedömningspraktiker som kan observeras i klassrummen utan att några sådana specifikt har iscensatts eller initierats av forskningsprojektet. Dessutom utforskades vilka syften lärare och elever ser med själv- och kamratbedömning och vilka uppfattningar av dessa företeelser som finns via fokusgruppintervjuer med elever och individuella intervjuer med lärare. Resultaten visar att eleverna generellt inte uppfattar själv- och kamratbedömning som bedömning, snarare mer som klassrumsaktiviteter. Eleverna tycker att kamratbedömning i viss mån kan vara bra för att förbättra sina prestationer, men ger samtidigt uttryck för att de i betydligt högre grad sätter värde på lärarens feedback än kamraternas. Även i denna studie framgick att eleverna i hög utsträckning inte ansåg sig själva som kompetenta nog att vara bedömare, och att de kände obehag inför att visa sina arbeten för andra. Trots att vissa elever nämner positiva möjligheter med själv- och kamratbedömning är det generella intrycket dels att eleverna själva inte har en hög medvetandegrad kring tankarna bakom själv- och kamratbedömning, dels att det finns anledning att fundera över det faktum att kamratbedömning kan vara kopplad till känslor av obehag hos eleverna.

Vi anser att de processer som innefattas i formativ bedömning är komplicerade och ställer tämligen höga krav på de lärare som arbetar med dem. Om förståelsen av formativ bedömning är instrumentell och/eller inriktad på att det är eleven som ska formas finns fara för att man inte i tillräcklig grad ser till kontextuella och

relationella aspekter. Mok (2010) hävdar utifrån sin studie nödvändigheten av att lärare får fördjupad kunskap: ” It is crucial to align teacher professional development with the assessment reform so that school teachers can be equipped with the knowledge, awareness, experience, and strategies needed to maximize the learning potential of peer assessment.” (s 237).

Slutsatsen är således lik den som dras i flera studier som refererats till tidigare i denna översikt, där exempelvis Gardner m fl (2011) – på basis av studier av 13 större förändringsinitiativ – menar att hållbar förändring kräver *professionellt lärande* snarare än ett instrumentellt anammande av metoder.

I en studie av Leitch m fl (2007) problematiseras det faktum att pedagogisk forskning i så låg utsträckning tar elevers perspektiv. Den kritik forskarna framför handlar i första hand om att den forskning som finns gör elever till *objekt* för forskning, snarare än ser dem som aktiva agenter i forskningsprocessen. Elevers erfarenheter och attityder testas eller observeras, men det är forskarnas tolkning som medierar förståelsen. Leitch m fl (2007) menar att det är möjligt att i högre grad involvera elever som *medforskare* i studier som rör deras skolvardag. Svårigheter och möjligheter associerade med detta diskuteras i studien, där mellanstadieelever involveras i att själva filma klassrumssituationer som de upplever har med formativ bedömning att göra. Eleverna involveras också i analysen av de filmsekvenser som spelats in. Denna studies syfte är att utforska möjligheten att göra eleverna till medforskare, och redovisar inte vad eleverna tycker om formativ bedömning. Forskarna konstaterar dock att det, trots diverse svårigheter och en rad etiska aspekter som måste tas hänsyn till, är möjligt att ta till vara på elevernas upplevelser i högre utsträckning än vad som blir resultatet om situationer enbart bestäms av forskare och förståelse av situationer enbart medieras via forskare.

Sammanfattning – Elevers uppfattningar

Forskning som tar ett elevperspektiv på formativ bedömning är en bristvara – i synnerhet i grundskolekontext. Begreppet formativ bedömning är positivt laddat och en rad olika metoder för formativt arbete implementeras på skolor världen över. Vi har redan konstaterat att studier kring effekterna av detta, i termer av lärande och uppmätta kunskaper, på grundskolenivå är påtagligt få. Man kan därutöver konstatera att vi vet oerhört lite om hur eleverna *upplever* olika formativa arbetssätt.

Den forskning som vi har gått igenom visar att det finns aspekter att vara vaksam över och som behöver studeras betydligt mer. Det finns exempelvis studier som visar att elever föredrar lärarens feedback framför kamraternas, och att det i fallet kamratbedömning finns anledning att fundera över elevers eventuella känsla av utsatthet i situationer där de ska bedöma kamraters prestationer och/eller få sina egna prestationer bedömda av kamrater.

Vissa forskare menar också att det går att involvera elever mer i forskning kring hur de upplever undervisning och situationer där formativ bedömning är en del.

3.2 Formativ bedömning i svensk forskning

Inledningsvis kan vi konstatera att litteratursökningen via de tre i metodkapitlet beskrivna söktjänsterna (LIBRIS, SwePub och Skolporten) var magert i jämförelse med de avhandlingar och licentiatuppsatser som de facto berör ämnet. Resultaten via SwePub gav principiellt mest, men är ändå otillräckligt. Det resultat som presenteras i tabell 3 nedan omfattar därför en sammanfattning av sökningarna via de tre söktjänsterna, kompletterat med den kännedom vi har om forskningsfältet i förhållande till publicerade avhandlingar, licentiatuppsatser och ett kapitel i en forskningsrapport. För att förstå hur det totala resultatet uppstått anges i kolumnen *sökord/träff* huruvida titeln hittats med sökorden ”formativ bedömning” eller ”bedömning för lärande”. Markeringen --- anger att inget av sökorden gett resulterat utan titeln hör till den grupp avhandlingar och licentiatuppsatser vi känner till på annat sätt. En del av dessa är ett resultat av den bredare översikt Forsberg och Lindberg (2010) genomfört tidigare och ligger till grund för denna översikt.

För att synliggöra en del av den problematik som möter den som ska genomföra systematiska översikter på basis av svensk forskning presenterar vi i tabellen även de ämnesord som LIBRIS använt för att klassificera de svenska avhandlingar och licentiatuppsatser vi använt, liksom de nyckelord respektive författare själv angett för sitt verk. Vidare har vi använt funktionen *avancerad sökning* i pdf-dokumenten och räknat hur många gånger

formativ (*formative* i de fall texten är på engelska) förekommer, vilket anges i kolumnen längst till höger. Syftet med att lyfta in denna information i tabellen är att synliggöra den problematik som finns i det att administrativa klassificeringssystem används för forskningspublikationer där de som fastställer hur ett arbete klassificeras saknar tillräcklig kännedom om fältet. I två fall har vi använt avvikande strategier: 1) Jessika Karlssons avhandling är skriven i psykologi, så hennes terminologi skiljer sig delvis från de övriga. Hon använder dock begreppet *informative feedback*, vilket vi – utifrån hur hon använder begreppet – valt att likställa med sådana aspekter av formativ bedömning som fokuserar feedback. 2) Barbro Westlunds avhandling förekommer endast i tryckt form. I detta fall har vi använt oss av det manuskript som diskuterades under hennes slutseminarium, vilket innebär att antalet i kolumnen måste ses som ungefärligt.

Ytterligare en aspekt som vi vill peka på är att två av avhandlingarna bidrar med kunskap om eller av relevans för formativ bedömning utan att författarna kanske haft en sådan avsikt eller signalerat att de gör det. I den här översikten gäller detta Ingela Elfströms licentiatuppsats och Gunilla Granaths och Johan Hofvendals avhandlingar, vilka har inkluderats för att de handlar om dokumentation respektive individuella utvecklingssamtal som båda utgör aspekter av lärares formativa bedömningspraktiker. Utöver det har ett kapitel i en forskningsrapport med per review inkluderats (Löfgren & Lindberg, 2011).

Då vi redogör för det bidrag de olika texterna ger till svensk forskning om formativ bedömning kan det förefalla som om några arbeten får mera utrymme än andra. Detta är fallet t.ex. då det rör sig om sammanläggningsavhandlingar eller –licentiatuppsatser, där ingående artiklar är baserade på olika delstudier och där kappan dessutom tillför nya aspekter av relevans för vår översikt. Detta gäller exempelvis Hartell (2012; 2013; 2014) och Hirsh (2011; 2012; 2013; 2014; 2015) nedan. I andra fall gäller endast en del av en monografi formativ bedömning – t.ex. Krantz avhandling – eller, som i Strandbergs sammanläggningsavhandling, där det är två av artiklarna som utgör det huvudsakliga bidraget till kunskap om formativ bedömning. Detta avgör hur stort utrymme författarna fått. Notera även att antalet träffar i pdf-filen (tabell 7) enbart avser kappan i de fall det gäller en sammanläggningsavhandling.

Tabell 3. Sökord – ämnesord – nyckelord

Författare - avh/lic - pub. år	Sökord / träff	Ämnesord/LIBRIS	Författarens nyckelord	Antal träffar i text för formativ/ e
Jessica Berggren Lic. uppsats 2012	---	engelska språket, pedagogik, återkoppling, teaching, feedback, English language	peer feedback; peer review; feedback training; EFL writing; revision changes; genre-based writing instruction; classroom research	**52 st**
Lisa Björklund Boistrup Avh. 2010	---	matematikundervisning, lärare och elever, studie-omdömen, Teacher-student relationships, mathematics –study and teaching	matematikklassrum, klassrum, bedömning, diskurser, socialsemiotik, multimodalitet, återkoppling, matematisk kompetens	**45**
Ingela Elfström Lic. uppsats 2005	---	Utvärderingsmetodik, individuella utvecklings-planer Evaluation, methodology	*Saknas*	**0**
Gunilla Granath Avh. 2008	---	Utvecklingssamtal – grund-skolan, skoldisciplin, samtal – skolan – grundskolan –Sverige, lärare – auktoritet, elever – identitet – normer – genus, pedagogik –kommunikation School discipline	Pastoral powers, development talks, log books, disciplining techniques, communication, normalisation, individualisation, intimisation Performativity	**0**
Eva Hartell Lic. uppsats 2012	Bedöm ning för lärande	*saknas*	technology education, assessment, formative assessment, assessment for learning, primary education,	**103**
Åsa Hirsh Avh, 2013	Bedöm ning för lärande	grundskolan, individuella utvecklingsplaner, kunskaps-kontroll, education, elementary schools, examinations, Sverige	Individual development plan (IDP), individual education plan, (IEP), teachers' assessment practices, formative assessment, documentation, gender, mediated action, tool, inscription/translation, primary/secondary/tertiar	**72**

Författare - avh/lic - pub. år	Sökord / träff	Ämnesord/LIBRIS	Författarens nyckelord	Antal träffar i text för formativ/ e
			y artifacts, mastery/ appropriation, IDP practices, contradictions, dilemmas, national and local steering of schools	
Johan Hofvendahl Avh, 2006	---	Utvecklingssamtal –grundskolan, samtalsanalys, skolor, interpersonell kommunikation, risk, attityder, föräldrar, lärare, elever, Conversation analysis	Kvartssamtal, utvecklingssamtal, samtalsanalys, strategi, risk	**0**
Maria Hurtig Avh. 2008	---	Portföljer Digitala medier Utbildningsteknologi Datorstödd undervisning Computer-assisted instruction Computer managed instruction Educational technology Digital media	Digital portfolios, meaning making, artefacts, ethnographic method, socio cultural theory, multimodal theory	**26**
Jessika Karlsson Avh, 2008	---	musikpsykologi, musikinterpretatio n, musikundervisning, Music – psychological aspects, music – instruction and study	music performance, emotion, expression, feedback, computer-based teaching, music education	**12 (informa-tive ass.)**
Susanne Kjällander Avh. 2011	---	Datorstödd undervisning Utbildning via Internet Computer-assisted instruction Computer managed instruction Internet in education	interaction, learning, meaning-making, digital learning resources, Social Science, school, classroom, pupils, teachers, didactic design, digital interface, assessment, designs for learning, multimodality	**31**
Joakim Krantz Avh. 2009	---	Utbildningspolitik -- Sverige Lärarutbildning – Sverige Professionalism Vocational qualifications	Educational policy, Critical Discourse Analysis, validity claims, communicative action, professional practice, teacher education, secondary	**13**

Författare - avh/lic - pub. år	Sökord / träff	Ämnesord/LIBRIS	Författarens nyckelord	Antal träffar i text för formativ/ e
			school, higher education	
Christian Lundahl Avh. 2006	---	Kunskapskontroll – historia – Sverige, kunskapsprov –historia – Sverige, kunskap –teori, filosofi, läroplaner –teori, filosofi, skolor, elever, utvärdering, kunskaps-överföring, personbedömning, skolor, elever	---	**35**
Ragnhild Löfgren & Viveca Lindberg 2011 kapitel i forsknings-rapport med peer review	---	Kemiundervisning –grundskolan – Finland, kemiundervisning –grundskolan – Sverige	---	**20**
Eva Mårell-Olsson Avh. 2012		Individuella utvecklingsplaner, skolan, Sverige, education	Individuella utvecklingsplaner, digital, verktyg, delaktighet, inflytande, identitetsarbete, identitets konstruktion, utbildning, skola	**10**
Eva Nyberg Avh. 2009	---	Naturorienterande ämnen – studier och undervisning –grundskolan – mellanstadiet Biologiundervisning Hållbar utveckling –studier och undervisning Biology – Study and teaching	Life cycles, plants, animals, insects, cognition and aesthetics, teaching and learning, pedagogical content knowledge, formative assessment, education for sustainable development	**148**
Max Strandberg Avh. 2013	---	läxor, kulturell mångfald, föräldrainflytande, interkulturell pedagogik, skolan, hemmet, multicultural education, education, diversity in the workplace, diversity in the workforce, multiculturalism,	homework assignments, cultural diversity, parents, life experiences, feedback	**20**

Författare - avh/lic - pub. år	Sökord / träff	Ämnesord/LIBRIS	Författarens nyckelord	Antal träffar i text för formativ/ e
		education – parent participation, home, homework		
Charlotta Vingsle Lic. uppsats 2014	Formative assessment	--	--	**350**
Barbro Westlund Avh. 2013	---	läsförståelse, läsundervisning, utvärdering, grund-skolan, literacy, läsundervisning, Evaluation, reading comprehension, education, elementary schools, Sverige, Kanada		**64 (i prel. manus)**
Sanna Wettergren Lic. uppsats 2013	---	Matematikundervisning, utvärdering, Evaluation, mathematics –study and teaching	tankekollektiv, tankestil, formativ bedömning, matematikkunnighet, kollegialt lärande	**33**
Anna Wernberg Avh. 2009	---	Inlärning, kognitiva processer, kognitionsforskning matematikundervisning, learning, mathematics – study and teaching, kunskapsmål	teaching, learning, variation theory, object of learning, learning objective, learning study, lesson study.	**4**
Antal avhandlingar: 14 st	**Antal licentiatuppsatser: 5 st**			
Kapitel/forsknings rapport: 1 st				

Antalet artiklar som ligger till grund för översikten presenteras i tabell 4 nedan.

Tabell 4. Artiklar i tre tidskrifter

Tidskrift	Antal artiklar
Educare 2006-	3
Pedagogisk forskning i Sverige 2000-	2
Utbildning och demokrati 2000-	3

I den följande resultatbeskrivningen separeras inte avhandlingar, licentiatuppsatser och artiklar, sammantaget är de få och hanteras därför som ett enda material. Rent generellt kan vi konstatera att fyra av totalt fem licentiatuppsatser är ämnesdidaktiska och skrivna av lärare inom ramen för någon av de ämnesdidaktiska forskarskolorna (Berggren, 2012; Hartell, 2012; Vingsle 2014; Wettergren, 2013). Även tre av avhandlingarna är ämnesdidaktiska (Björklund Boistrup, 2010; Nyberg, 2009; Westlund, 2013) medan de övriga kan ses som allmändidaktiska eller pedagogiska (Granath, 2008; Hirsh, 2013; Hofvendahl, 2006a; Hurtig, 2008; Kjällander, 2011; Krantz, 2009; Strandberg, 2013). Principiellt ger Karlssons (2009) avhandling ett ämnesdidaktiskt bidrag till frågan om formativ bedömning, trots att den är publicerad som en avhandling i psykologi. Ingen av artiklarna fokuserar ämnesdidaktiska frågeställningar. Vi följer inte den kategorisering som tidigare presenterats för den internationella delen av resultatet, utan resultatet av den svenska översikten presenteras först i relation till tre övergripande teman, och därefter relateras resultatet av den svenska delen av översikten till resultatet av den internationella översikten.

De tre huvudsakliga teman som återkommer i svensk forskning relaterad till formativ bedömning är följande: dokumentation för formativ bedömning, formativ bedömning i klassrumskommunikationen, och skolans styrning och formativ bedömning.

Dokumentation som grund för formativ bedömning

Olika former för dokumentation används – och sprids – i samband med utvecklingssamtal med elever och deras föräldrar. Införandet av individuella utvecklingsplaner (IUP) år 2006 (SFS 2005:179), liksom förändringen av dess syfte 2008 (Utbildningsdepartementet, 2008), har bidragit till att nya teknologier har utvecklats: mallar producerade av lärare, skolchefer eller administrativ personal likaväl som olika kommersiella digitala mallar. Utformningen av dessa styr såväl dokumentationens innehåll som formen för hur dokumentationen görs. I förlängningen påverkas även relationerna mellan dem som medverkar i bedömningssamtalen och formen för kommunikation, nya bedömningspraktiker etableras.

Några av studierna (Elfström, 2005; Vallberg Roth & Månsson, 2006; Vallberg Roth, 2009) har granskat IUP som samhälleliga fenomen. Innan det fattas ett politiskt beslut om införandet av IUP introduceras de i Ds 2001:19 som en demokratisk rättighet – något alla elever har rätt till och som antas kunna bidra till att alla elever blir sedda. Då IUP gjordes obligatoriska från och med 2006, betonades dess formativa funktion i relation till bedömning. Två år senare förändrades förordningstexten, vilket innebar att IUP, från och med hösten 2008, utöver den formativa funktionen även gavs en summativ funktion och där beslutsfattarna öppnade för att dokumentet även kunde innehålla betygsliknande omdömen. I relation till sociologiska kategoriseringar kan IUP förstås som ett modernt/senmodernt fenomen, som handlar om individens självreglering, ansvar och val av framtid, men också i termer av en reglerad barndom. Skolan som samhällelig institution, normalisering, anpassning och vårdnadshavarnas respektive elevens ansvar är aspekter som lyfts fram för att belysa IUP (Elfström, 2005; Vallberg Roth, 2009).

Elfströms licentiatuppsats (2005) hör till de tidiga studierna om IUP i förskolan, förskoleklass och de första skolåren, och genomfördes innan den här typen av dokumentation blev nationellt obligatorisk. Studien syftade till att belysa varför IUP blev allt mer vanligt förekommande och vad som var syftet med dem, samt vilken syn på barn som representerades i dessa dokument. Metodiskt baserades studien på intervjuer med pedagoger i förskola och skola, intervjuer med ledningspersonal, samt dokument (mallar för IUP). Elfström konstaterar att beslutet att ta IUP i bruk på många håll fattas av

> förvaltningsledning eller förskolechefer och rektorer /…/ därefter följer politikerna efter och tar beslut om det i förskole- och skolplanen. I andra kommuner är det tvärtom politikerna som skriver in det som en punkt i planerna och det blir sedan förvaltningarnas uppgift att följa upp beslutet (a.a. s. 17).

Trots att studien berör förskola-åk 1-3 utgör ett av de fyra syften Elfström identifierade för användningen av IUP *att alla elever ska gå ut åk 9 med godkända betyg*. De tre övriga syftena är *att se alla barns/elevers positiva utveckling och hjälpa dem vidare*, *att tidigt identifiera barn med svårigheter*, samt *att underlätta för*

barnen i samband med övergångar mellan stadier. Innehållsligt identifierar hon mål för barnens utveckling som kan relateras till utvecklingspsykologiska teorier, sociala aspekter (barnens anpassning och beteende) och, främst till skolbarnen, mål relaterade till tidig läs-, skriv- och matematikutveckling. De mallar som utformats för IUP bär spår av tidigare observationsscheman för de yngre åren och, för barn från och med förskoleklass, även av läroplansformuleringar. Även Vallberg Roth och Månsson (2006), vilkas studie också baseras på dokument och intervjuer, kommer till ett motsvarande resultat. Deras studie omfattade 100 IUP i årskurserna 3, 6 och 8 inom fyra sydsvenska kommuner. En första analys av de dokument som utgjorde underlaget för IUP visade att de organiserade

> … innehåll och ämnesområden huvudsakligen som behovsområden styrda efter utvecklingspsykologisk teoribildning och/eller som innehållsområden strukturerade efter läro- och kursplaner. Resultatet visade också en tendens att vissa ämnesområden i läro- och kursplaner var berörda, medan andra reducerades och slogs samman eller inte kommenterades alls. Det informella lärandet och innehållet i den fritidspedagogiska verksamheten berördes i princip inte alls. Det tematiska innehållet berördes i mycket låg utsträckning, eller inte alls. Det fanns också en avsaknad på mångkulturellt och individualiserat innehåll. Vidare saknades exempelvis frågor som bejakade ett kritiskt, kreativt tänkande, initiativtagande och reflekterande (Vallberg-Roth & Månsson 2006, s. 47).

De kommuner som valdes ut för studien skilde sig från varandra med avseende på storlek, tätort–landsbygd och graden av kulturell mångfald (homogen-heterogen sammansättning av befolkningen). I ett följande steg analyserades dokumenten med avseende på skillnader och likheter mellan kommunerna och man fann att de demografiska skillnaderna mellan kommunerna återspeglades i den innehållsliga utformningen av utvecklingsplanerna (a.a.). I kompletterande studier (Vallberg Roth & Månsson, 2008a; 2008b) användes samma kommuner för insamling och analys av fler IUP. Enligt författarna visar de individuella utvecklingsplanerna dels ”en hög grad av styrande och standardiserade mallar i olika riktningar, dels önskad självreglering och låg grad av inflytande för barn och vårdnadshavare” (2008b, s. 97). I en syntetiserande artikel baserad på det totala materialet fann Vallberg Roth (2009, s. 208) att ”Barns och ungas positioner tycks röra sig om en viss likriktning med variation beroende på var barnen råkar växa upp”. Hon finner att IUPerna snarast kan ses som individorienterade normalplaner, men har även identifierat mönster av skillnader som de kunnat relatera till var IUPerna hämtats från. Tyngdpunkterna i de individorienterade normalplanerna varierar mellan en orientering åt kunskap, fostran eller omsorg. Innehållsligt förekommer alla tre kategorierna i alla kommuner, men vilken kategori som dominerar varierar mellan kommunerna.

Senare studier av IUP har bidragit till att nyansera innehållet i IUP. Åsa Hirshs avhandling (2013) baseras på fyra delstudier om IUP. Två av dem bygger på analyser av 379 IUP-dokument i tre kommuner för årskurserna 3, 5 och 8, de två övriga delstudierna baseras på djupintervjuer med 15 lärare om de dokumentationsmallar som används för IUP. I den första delstudien (Hirsh 2011) analyseras utvecklingsmålen (totalt 829 st) i de 379 IUP-dokumenten och resultatet beskrivs i tre kategorier: *being* (fokuserar elevens personlighet, beteende och förhållningssätt till skolan men har sällan en koppling till läroplanens eller kursplanernas mål)*, doing* (fokuserar vad eleven ska göra, dock vanligen i generella termer) och *learning* (fokus på vad eleven ska lära sig). Den sistnämnda är den kategori som innehåller det största antalet mål. Hirsh analyserade även de strategier (557 st) som eleverna rekommenderas i IUPerna och finner att även dessa kan beskrivas i termer av *being* och *doing* (men inte learning). Hon kopplar strategierna till de tre målkategorierna och får följande utfall (Hirsh 2013, s. 118):

- being by being (t.ex. “Bli mer uppmärksam på lektioner genom att vara mer koncentrerad”)
- being by doing (t.ex. “Bli mindre impulsiv genom att tyst räkna till tio innan jag sager något”)
- doing by doing (t.ex. “Räcka upp handen en gång per lektion genom att tänka på att räcka upp handen”)
- learning by being (t.ex. “Sikta på bättre resultat i engelska genom att tro på dig själv”)
- learning by doing (t.ex. “Utveckla din huvudräkning i allmänhet och multiplikationstabellen i synnerhet genom att använda datorprogram X varje dag i skolan och hemma”)

I den avslutande delen av analysen för den här delstudien granskar hon mål och strategier i förhållande till huruvida de är vaga eller konkreta å den ena sidan och individuella eller generella (d.v.s. skulle kunna gälla vilken elev som helst) å den andra. Här finner hon att mål och strategier tenderar att vara mer konkreta och individuella för de yngre barnen och mer generella och vagare för de äldre barnen.

I följande delstudie driver Hirsh (2012) analysen av de 829 målen i samma IUP vidare, denna gång med fokus på skillnader och likheter i relation till de mål som tilldelades flickor respektive pojkar. Resultatet av denna studie visade att fördelningen av en typ av mål (learning) var relativt jämn mellan flickor och pojkar, men vad beträffar kategorin being hade pojkarna tilldelats betydligt fler mål av detta slag än flickorna. Denna kategori av mål detaljanalyserades och kategoriserades vidare för att identifiera vad skillnaderna bestod av. Analysen resulterade i sex underkategorier: ”attityd, störande beteende, slarv/oordning, osäkerhet/tysthet, ineffektivitet och ’bibehåll det goda uppförandet’” (Hirsh 2013, s. 119). Inom fyra av kategorierna (attityd, störande beteende, slarv/oordning och ineffektivitet) var pojkarna överrepresenterade, medan flickorna var överrepresenterade inom en kategori (osäkerhet/tysthet). Resultatet av fördelningen diskuteras i relation till dels en förklaringsmodell som bygger på att pojkars och flickors olika sätt att uppföra sig förstås som relaterade till en strävan att bibehålla genusordningen, dels en förklaringsmodell som bygger på att sociokulturellt formade föreställningar om flickor och pojkar kan påverka lärares förväntningar och reflekteras till exempel i formuleringarna i IUP.

I den tredje delstudien (Hirsh 2015) utforskas lärares uppfattningar om IUP och deras syften, hur de använder dem och de kontextuella villkoren för användningen av dem. Resultaten formuleras i termer av tre IUP-praktiker: Praktik 1 karakteriseras av att IUP främst uppfattas vara till för föräldrar och elever och fungerar formativt enbart om de kan använda den information som finns i dem. Lärarna använder inte dokumenten annat än i relation till utvecklingssamtalen. De fyller plikttroget men motvilligt i de mallar de tilldelas – ett arbete som de upplever som tidsödande utan att resultera i motsvarande förändring. Förutsättningar för samverkan kring IUP finns i liten utsträckning och rektorernas bidrag är marginellt. Specifikt för praktik 2 är att IUP främst uppfattas ha ett informativt syfte, men den uppfattas även som ett redskap för att involvera elever och föräldrar i mål och utvärdering i relation till utvecklingssamtalen. Vidare karakteriseras praktik 2 av att lärarna, medan de formulerar IUP säger sig få syn på enskilda elevers behov vilket bidrar till att de bär dem med sig i sitt fortsatta arbete. Formuleringen av IUP upplevs som tidsödande men bidrar till ökad medvetenhet och reflektion. Givna IUP-mallar följs – det finns tendenser till viss frihetlighet men de kommunala riktlinjerna upplevs motverka detta. Rektorerna är tydligare involverade i IUP-arbetet än i praktik 1, främst i relation till formuleringar och organiseringen av arbetet. Inom praktik 3 ses IUP som ”ett formativt verktyg som påverkar lärarnas sätt att tänka kring, organisera och genomföra undervisning” (Hirsh 2013 s. 122) som både bidrar till att involvera elever i att förstå sitt lärande och till att lärarna samtalar om och utvärderar sin undervisning. IUP används aktivt och det finns tid för kontinuerlig uppdatering av dem i elevernas scheman. IUP upplevs ha bidragit till att lärarna förändrat sina förhållningssätt och sitt arbete jämfört med tidigare. Inom praktik 3 upplevs rektorerna vara aktiva medparter i arbetet med IUP.

Den sista delstudien fokuserar dels vad som framstår som dilemman i lärare arbete med IUP, dels vilka strategier lärare använder för att hantera dilemman. De tre dilemman som identifierades gällde 1) *dokumentation vs undervisning* och kan relateras till den tid lärarna lägger på dokumentation och som minskar den tid de tidigare lagt på planering och reflektion relaterad till undervisningen, 2) *officiellt korrekt vs elevvänligt språk,* ett dilemma som berör både mängden och komplexiteten i förhållande till innehållet i IUP men också språkdräkten – formell eller vardaglig, där det formella tar över på bekostnad av begriplighet för föräldrar och elever. 3) *summativ vs formativ bedömning* Den ”betygsliknande” delen av IUP upplevs stjäla föräldrars och elevers uppmärksamhet på bekostnad av de framåtsyftande delarna, vilket lärarna upplever riskerar påverka elevernas motivation på ett negativt sätt. De tre strategier som Hirsh identifierar att lärarna använde för att hantera dessa dilemman är följande: *Procedural display*, vilket huvudsakligen innebär att IUP reduceras till rituellt utförande, något som måste göras två gånger om året. *Enligt konstens alla regler* innebär egentligen att den här gruppen lärare saknar en strategi för att hantera dilemman – de försöker följa reglerna men upplever frustration i relation till dem. Den tredje strategin har benämnts *Mot en ny ’thirdness’*, vilket syftar på en riktning bort från det antingen-eller som ligger i dilemmats natur. Efter att ha försökt följa reglerna och upplevt frustrationen arbetar de lärare som gav uttryck för denna strategi med en kollektiv

förändringsprocess som bidrar till en ny IUP-praktik som främst karakteriseras av ett elevvänligare språk i dokumenten och av att ge större utrymme åt IUPs formativa syften för elevers lärande och lärarnas undervisning. Sammantaget bekräftar resultatet av Hirshs studier av innehållet i IUP (delstudierna 1 och 2) tidigare studier, men nyanserar dem också i var och en av de båda första delstudierna. Den dimension som hennes arbete som helhet bidrar till att synliggöra gäller lärarnas arbete (delstudierna 3 och 4), där hon finner att lärarna hamnar i kläm mellan nationell och lokal styrning. Den nationella styrningen ger principiellt de grundläggande riktlinjerna men samtidigt lämnar staten över ansvaret till lokala beslutsfattare (huvudmän och rektorer) att utforma mallarna – eller köpa in dem av tredje part. Genom decentraliseringen och marknadsanpassningen kan makten över konkretiseringen av uttolkningen av den nationella styrningen hamna hos konsulter vilkas kompetens är oklar med avseende på kunskapsbedömningar.

Eva Mårell-Olssons (2012) avhandling handlar om elevers, föräldrars och lärares arbete med att å ena sidan förbereda och genomföra utvecklingssamtal, samt, med stöd av ett digitalt verktyg, upprätta individuella utvecklingsplaner för elevernas lärande å den andra. Avhandlingen syftar till att dels bidra till en ökad förståelse för arbetsprocessen med de individuella utvecklingsplanerna, dels utforska betydelsen av digital dokumentation och arbetsprocessens innebörd för elevernas delaktighet, inflytande och identitetsarbete. Det empiriska underlaget för Mårell-Olssons studie omfattar en enkätstudie riktad till 52 föräldrar, intervjuer med 7 elever, 7 lärare och 15 föräldrar samt 49 elevers individuella utvecklingsplaner. Studiens resultat visar i sammanfattning att digitala verktyg påverkar arbetsprocessen – de mål som formuleras för eleverna förefaller ofta bestå av kortsiktiga ämnesmål eller momentmål snarare än av utvecklingsmål. Detta kan dock relateras till att de företag, vilkas produkter skolorna köpt, är de som utformar arbetsprocessen med IUP utifrån sina tolkningar av styrdokumenten. Vidare finner Mårell-Olsson att elever, föräldrar och lärare definierar arbetsprocessen utifrån sina förväntningar. Eleverna lär sig tidigt hur de förväntas uttrycka sig, uppföra sig och förhålla sig till sig själva, till sina prestationer men även till lärares skriftliga omdömen och föräldrars förhoppningar. Föräldrarna önskar en individanpassning av undervisningen för barnen och att de själva ska kunna ta aktiv del i barnets skolgång genom ökad delaktighet och inflytande i arbetsprocessen. Lärarna försöker att hantera arbetsprocessen effektivt genom att exempelvis formulera och anpassa elevmålen till den egna undervisningen men också genom återanvändning av färdigformulerade omdömen för flertalet elever (s. 210).

Studien visar även att arbetsprocessen med elevers individuella utvecklingsplaner handlar om mer än enbart elevens lärande. I studien framträder fenomen som makt, kontroll och förhandling. Liksom i andra studier relaterade till IUP förmedlas genom IUPerna sociala föreställningar om mönsterelever som förväntas bidra till ett gott arbetsklimat och arbetar för att prestera sitt bästa. Eleverna, föräldrarna och lärarna deltar i skapandet av mönstereleven – var och en utifrån sitt perspektiv och med olika motiv och strategier. Genom att elevernas IUP med skriftliga omdömen är offentliga handlingar kan även andra än eleven, föräldrar och lärare principiellt ta del av innehållet i dem. Detta kräver både en medvetenhet och kunskap om arbetsprocessen och vad språket och formuleringarna i de skriftliga omdömena om eleven kan få för betydelse.

Hartells (2012) licentiatuppsats syftar till att utforska hur bedömning i grundskolans teknikämne organiseras och genomförs på skolnivå och av lärare på klassrumsnivå. Den baseras på klassrumsstudier, dokumentation för bedömning (IUP) och intervjuer med lärare. Som vägledning för arbetet som helhet användes Wiliams (2009) begrepp *long* och *short cycle formative assessment.* Två delstudier ligger till grund för licentiatuppsatsen. I den ena delstudien (Hartell, 2014) fokuserades en längre cykel av formativ bedömning, exemplifierad med data från 351 IUP som redskap för bedömning i teknik från kommunala skolor i fem kommuner. IUP:erna har analyserats och kategoriserats med avseende på innehåll och utformning. Mallarna kategoriserades och resultatet blev 14 typer av IUP-mallar. Specifikt för teknikämnet var att det var det skolämne som mest sällan (5 av 14) fanns representerat i mallarna. Ingen av de individuella utvecklingsplanerna gav information om någon enskild elevs kunskapsutveckling i relation till styrdokumenten i teknik. Det saknades också förslag på vilket stöd eleven behöver för att nå de uppsatta lärandemålen i teknik. I de IUP-dokument där teknikämnet förekommer ges generella kommentarer om god måluppfyllelse i teknik där de flesta uppges nå de uppsatta målen. En slutsats av studien är att mallarna är bristfälliga, vilket bidrar till att det inte går att använda dokumenten som grund för utvecklingssamtal om elevers kunnande i grundskolans teknikämne. Vi återkommer till den andra delstudien under temat klassrumskommunikation nedan.

I en artikel om dokumentarism som fenomen i samtiden diskuterar Asp-Onsjö (2011) hur dokumentationen blir styrande i skolans verksamhet. Två parallella typer av dokumentation förekommer, den som tar form som IUP för alla elever och den som tar form som åtgärdsprogram för elever som bedömts vara i behov av särskilt stöd. På basis av sin avhandling och med konkretiserande empiriska exempel i artikeln hävdar hon att den dokumentation som produceras om eleverna får konsekvenser för individernas identitetsuppfattningar genom samtliga omdömen som eleverna får under sin skoltid. Men dokumentationen får också konsekvenser för den kunskapssyn som dokumentationen producerar. Med utgångspunkt i den kunskap som finns om vad som kommer till uttryck i dokumentationen drar Asp-Onsjö slutsatsen att denna ”nya kunskapssyn riskerar att kväva vissa aspekter av skolans arbete som uppmuntrar kreativitet och kritisk förmåga och kan också negativt påverka elevens förmåga att analysera och kritiskt granska sin omvärld” (s. 49).

Gemensamt fokus i två andra avhandlingar (Hofvendahl, 2006a; Granath, 2008) och en artikel (Hofvendahl, 2006b) inom ramen för denna kategori är utvecklingssamtalen. I båda fallen beaktas den dokumentation som ligger till grund för samtalen – Hofvendahl (2006c) benämner dem ”förpapprade” samtal. Hans studie baseras på 80 ljudinspelade kvarts- och utvecklingssamtal med elever i årskurs fem genomförda 1992/93 (45 st) och 2004 (35 st). Fokus i studien ligger dels på språkanvändning, hur samtalen inleds och avslutas, dels på innehållet i samtalen. Vid tidpunkten för avhandlingsarbetet var en utgångspunkt att det varje år genomförs cirka 2,6 miljoner elevsamtal i grundskolan och gymnasieskolan sammantaget, möten som involverar cirka 5,5 miljoner samtalsdeltagare. Ändå fanns det nästan ingen kunskap om hur de genomförs och vad som utgör fokus i dem. Avhandlingens syftade till att studera vilka kommunikativa strategier som togs i bruk för att hantera risksituationer i samtal mellan lärare, elever och föräldrar i grundskolan, det vill säga ett moment vars utgång är oviss och som potentiellt kan skapa eller leda till problem. Strategier sågs som en del (och materialiseringen) av vardagskulturella normer, umgängesregler och habitualiserade ”sätt att göra”. Tre specifika talsituationer bedömdes ha särskilt analytiskt värde: (1) hur samtalet startas, (2) hur man initierar tal om problem och (3) hur samtalet avslutas, eller närmare, möjligheten för elever och föräldrar att lyfta egna ”övriga frågor”. Hofvendahls resultat synliggör främst formerna för samtalen, innehållsligt framgår att fokus föreföll ligga på eleven som person (sociala och beteendemässiga aspekter) och i stor utsträckning på brister. Vad gäller innehållet i utvecklingssamtalen stöder hans resultat och resultaten från studierna om IUP ömsesidigt varandra. Utöver tolv ljudinspelade utvecklingssamtal fokuserar Granath (2008) den kommunikation mellan lärare och elev som pågår i loggböcker. Det empiriska underlaget kommer från två skolor, i båda fallen gäller utvecklingssamtalen elever i åk 6, medan loggböckerna representerar elever i åk 9. Både loggböckerna och utvecklingssamtalen är viktiga redskap för att påverka eleverna och deras förhållningssätt till sig själva, sitt skolarbete, sina liv och sin framtid. Interaktionen sker privat/individuellt – vid utvecklingssamtalet är elev, vårdnadshavare och lärare närvarande och loggböckerna skrivs av enskilda elever och kommenteras individuellt av läraren till eleven. Både utvecklingssamtal och loggböcker formar elevernas identiteter (”skoljag”) – enligt Granath sker detta individuellt men ändå i relation till tydligt normativ uppfattning. Reglering för självreglering står för Granaths resultat, som ligger i linje med resultaten från Vallberg Roth och Månssons (2006; 2008a; 2008b) studier.

Två studier som bygger på andra typer av dokumentation är Anna Wernbergs avhandling (2009) och Sanna Wettergrens (2013) licentiatuppsats. Wernberg genomför en learning study, en typ av interventionsstudie, i matematik och svenska i årskurs 4. Karakteristiskt för en learning study är att man påbörjar arbetet med interventionerna genom att först kartlägga vad eleverna kan i relation till det lärandeobjekt (det avgränsade innehåll) som man valt för studien. Innehållet väljs utifrån medverkande lärares erfarenheter om vad som är svårt för elever att lära sig, alternativ vad läraren erfarit som svårt att undervisa om. Kartläggningen beskrivs i termer av förtest, ett begrepp från klassiska experimentella studier, som sedan följs upp av ett eftertest i relation till varje learning study-cykel. Medan tidigare learning study-studier på ett ganska oproblematiskt sätt hanterat frågan om förtestet, dess konstruktion och användning, ramar Wernberg in förtestet och interventionen i termer av formativ bedömning. Hon pekar på att den här typen av studier dels hanterar förtestet i syfte att analysera

vad eleverna har svårt med och påpekar att dokumentationen används som utgångspunkt för lärarnas planering av lektionerna. Mellan varje intervention kartläggs elevernas kunskaper för att se vilka förändringar som framgår och de ligger sedan till grund för lärarnas fortsatta planering. Med den beskrivningen svarar learning study-typen av studier mot de kriterier som ställs på formativ bedömning[22].

Wettergrens studie syftar till att kartlägga vilka tankestilar om bedömning och matematikkunnighet som kommer till uttryck under en serie av fokusgruppssamtal med lärare i årskurserna 4-6 representerande två skolor. De dokument hon använder som underlag för serien av fokusgruppintervjuer är följande: grundskolans kursplaner i matematik (SFS 2010:37), kunskapskraven i matematik (SFS 2011:19) samt elevexempel i matematik. De sistnämnda består dels av ett prov för årskurs 4 (ur lärarhandledningen för ett specifikt läromedel), ett ämnesprov i matematik för årskurs 5 från år 2005/2006, samt en uppgift som kräver problemlösning (ur ett diagnostiskt material för elever i årskurserna 6-9). Resultatet konstrueras i form av två tankestilar, en som gäller bedömning och en som gäller matematikkunnighet. Tankestilen om bedömning är huvudfokus för vår översikt. Den karakteriseras av att lärarna ser bedömning som ett pågående arbete som bör genomföras såväl skriftligt som muntligt. I samtalen framstår formativ bedömning som överordnad betygssättande bedömning – man skulle också kunna säga att det förefaller som att lärarna känner stark press på att arbeta med formativ bedömning. Formativ bedömning kommer till uttryck i samtalen om elevernas läsningar i exemplen på följande sätt: a) elevernas lösningar på en uppgift används för att forma och anpassa undervisningen, b) efter explicit reflektion över elevers lösningar kan undervisningen justeras genom att rikta fokus mot ett specifikt område eller mot specifika typer av uppgifter. I detta fall används elevernas lösningar som återkoppling till den egna undervisningen. Vidare identifierar lärarna behov av c) feedback till elevens arbete, vilket kan innebära att de rekommenderar eleven att lära mer eller öva på specifika uppgifter. I detta fall ges ansvaret för förändring till eleven. Slutligen framgår det av samtalen att lärarna återkommande använder prov och diagnoser, men att eleverna inte alltid får tillbaka dessa. Återkopplingen till eleverna förefaller utebli i dessa fall. En fråga som det empiriska underlaget för studien inte ger svar på är hur systematiska bedömningarna är över tid, d.v.s. om lärarna har rutiner eller strategier för bedömningsarbetet som helhet och vilka de i så fall är. Lärarna ger också uttryck för att de tycker det är svårt att organisera för en typ av bedömning i matematikundervisningen som involverar eleverna och gör bedömningen till en självklar del av undervisningen.

Formativ bedömning i klassrumskommunikationen

Det som förenar de avhandlingar som presenteras nedan är att de tar sina utgångspunkter i formativ bedömning som något som tar form i klassrumskommunikationen och som har betydelse för såväl elevers lärande som för lärarens undervisning. De perspektiv de valt förenas av att de beaktar sociala situationer i klassrummet som centrala för elevers lärande – vad eleverna får tillgång till (alternativt erbjuds) bidrar till att forma det som blir centralt att fokusera för att få syn på hur elevernas kunnande förändras och vad lärarna gör av det för att justera sin undervisning.

Lisa Björklund Boistrups (2010) avhandling är en socialsemiotisk studie av matematik. Den baseras på klassrumsobservationer av matematikundervisning i fem klassrum i årskurs 4, och det hon fokuserar är uttryck för explicita och implicita bedömningshandlingar (*assessment acts*) i samtalen mellan lärare och elever. Särskilt tittar hon på vilken agens eleverna får tillgång till i relation till bedömningshandlingarna och vilka bedömningsdiskurser som kan uttolkas. Klassrumskommunikationen analyseras för att urskilja vilka möjligheter som finns för elevers aktiva agens, d.v.s. engagemang och deltagande i samtalen å den ena sidan och deras lärande å den andra. Baserat på de av Hattie och Timperley (2007) analytiskt identifierade nivåerna för feedback (uppgiftsnivå, processnivå, självreglering och personlig) i kombination med de av Skovsmose

[22] Ytterligare en avhandling som ger en ingående beskrivning av förtestets funktion och konstruktion i relation till Learning study är Angelika Kullbergs (2010) avhandling *What is taught and what is learned. Professional insights gained and shared by teachers of mathematics*. Gothenburg Studies in Educational Sciences 293.

(2005) identifierade centrala aspekterna av matematisk kompetens (matematiska begrepp, tillämpning av matematiska begrepp i olika sammanhang och kritisk reflektion) kopplas Hatties och Timperleys generella begrepp till matematiska kompetenser för att få ett ämnesdidaktiskt analysredskap. Utifrån detta finner hon två dominerande fokus i sitt material: uppgiftsfokus och processfokus relaterat till matematiska begrepp. Även självreglering förekom i hög utsträckning, medan feedback på processnivå relaterad till tillämpning av matematiska begrepp förekom i något mindre utsträckning. Ovanligast var fokus på eleven själv, följt av processfokus relaterad till kritisk reflektion. Genom att relatera förekommande bedömningshandlingar, fokus för feedback och den roll de semiotiska resurserna fick och vad eleverna genom dessa får tillgång till i termer av agens och matematiskt lärande, konstruerade Björklund Boistrup fyra bedömningsdiskurser som var och en resulterar i olika möjligheter för elevers aktiva agens och lärande: *Gör det fort och gör det rätt*, *Vad som helst duger*, *Allt kan tas som utgångspunkt för en diskussion*, samt *Resonemang tar tid*. Dessa är sinsemellan olika till sin karaktär och visar att lärares bedömningshandlingar bidrar till att skapa skilda förutsättningar för elevers lärande. Samma lärare kan dock representera fler än en bedömningsdiskurs, diskurserna är således relaterade till situationer snarare än personer. Björklund Boistrup konstaterar sammanfattningsvis att de diskurser som är resultatet av hennes studie "måste ses som en komplex samverkan mellan styrdokument, beslut fattade på olika nivåer i och utanför skolan som institution samt dominerande diskurser i den bredare institutionella kontexten" (s. 207) – alltså inte enbart som något som kan relateras till enskilda lärare.

Hartells (2013) andra delstudie, som ingår i hennes tidigare introducerade licentiatuppsats, fokuserade kortare cykler av bedömning. Det empiriska materialet för denna studie skapades genom klassrumsobservationer av två tekniklärares bedömningshandlingar i grundskolans årkurser 4 och 5 under lektioner då de undervisade om ett arbetsområde de själva valt, samt intervjuer med lärarna. Båda lärarna hade lärarutbildning och lång erfarenhet av undervisning, men ingen av dem var behöriga för undervisning i teknik. Som redskap för att klassificera lärarnas bedömningshandlingar använde Hartell Black och Wiliams (2009) fem nyckelstrategier (som har presenterats tidigare i rapporten). Det resultat hon kom fram till var att det gick att se spår av de fem nyckelstrategierna men huvudsakligen föreföll lärarnas klassrumsbedömningar vara baserade på mer eller mindre osystematiska minnesbilder. Det de främst gav uttryck för under intervjuerna var att de tittade efter glimten i elevernas ögon, som de tog som indikation på att eleverna hade förstått. Intervjuerna visade också att lärarna hade få möjligheter till kollegiala samtal om bedömning och att de saknade fortbildning om bedömning. Hartells (2012) slutsats, baserad på de båda delstudierna, är dels att ”skolledningen har delegerat allt ansvar för planering, genomförande av undervisningen och uppföljningen av elevernas kunskaper i teknik till läraren” (s. 157-158), dels att lärarna var beroende av sin tidigare erfarenhet när det gäller att bedöma elevernas kunskaper i syfte att säkerställa att eleverna når de kunskapsnivåer de har rätt till” (s. 158). Lärarna i Hartells studie förefaller således ha lämnats ensamma med ansvaret för bedömningsarbetet.

Maria Hurtigs (2007) avhandling fokuserar elevers och lärares arbete med den bedömningspraktik som är knuten till digitala portföljer, mer specifikt på vad som görs tillgängligt för eleverna i de situationer då de arbetar med sina portföljer. I hennes avhandling hanteras således digitala portföljer som redskap för bedömning, och det övergripande syftet är att beskriva och analysera elevers meningsskapande i arbete med digitala portföljer, vilket preciseras som att det handlar om såväl vad undervisningen erbjuder och hur arbetet organiseras som om vad eleverna skapar mening kring i relation till det som erbjuds och hur de använder till buds stående redskap. Det perspektiv som används är sociokulturellt, vilket riktar fokus på relationen undervisning-redskap-lärande och interaktionen mellan deltagarna. Elevernas (åk 2–6) meningsskapande synliggörs genom den kommunikation och de bedömningshandlingar som uppstår i interaktionen mellan lärare och elev(-er) och portföljen som artefakt. Flera skolämnen relateras till arbetet med portföljerna: svenska (läsning, skrivning), bild, engelska och idrott och hälsa. Studien är etnografisk, datamaterialet består dels av videodokumentationer av elevernas arbete med sina portföljer, dels av intervjuer med lärarna. Hurtig urskiljer fyra typer av aktiviteter i arbetet med portfolio: design, bedömning, redigering och rutinarbete. Aktiviteterna urskiljs analytiskt genom deras skilda syften. Sammantaget konstituerar dessa aktiviteter det som karakteriserar arbetet med digitala portföljer, nämligen att det handlar om att

- fatta beslut om hur portföljen ska struktureras och vad den ska innehålla (design),
- bedöma det material som läggs in i portföljen,

- redigera materialet och / eller strukturen

Dessutom förekom det rutinarbete i samband med var och en av de tre förstnämnda aktiviteterna. Det Hurtig betonar är dock att vad som framstår som rutinarbete är något som förändras över tid – till en början är det mesta förstås nytt och faller inom ramen för de övriga tre aktiviteterna, men med elevernas ökande erfarenheter av arbetet med de digitala portföljerna omvandlades delar av innehållet till rutinarbete. Bedömningsaktiviteten beskrivs enligt följande:

> När eleverna lagt in de objekt som ska finnas på en sida i portföljen och gjort de redigeringar av sidans layout som de vill göra, ska en bedömning av arbetet läggas in på sidan. Arbetet inleds med att pedagogen ber eleven berätta varför de valt ett arbete och vad för kunskaper och färdigheter arbetet representerar. Inledningsvis stöttar pedagogen eleverna med frågor som formas utifrån de bedömningskriterier som finns för arbetet. /…/ *Genom stöttningen lär sig eleverna vad det är som ska fokuseras i bedömningen men eleverna får också utrymme att ge egna motiveringar av vad arbetet visar*. När eleven och pedagogen har gjort bedömningen ska den realiseras i portföljen, och detta sker för det mesta genom att eleverna skriver ned bedömningen i form av en text. När detta är gjort och texten är inlagd på en sida i portföljen är arbetet med bedömningen slutförd. (s. 107 – *vår kursivering*).

I det samspel som Hurtig beskriver i citatet lär sig eleverna bedöma sina arbeten genom de modelleringar lärarna gör, där kriterierna så att säga tas i bruk för att granska elevernas arbeten. Med fokus på lärande i relation till portföljerna identifierar Hurtig tre meningsskapande förmågor som eleverna utvecklar i arbetet med sina portföljer: förmåga till självreflektion[23], förmåga till multimodalitet och förmåga att hantera informationsteknik. Dessa förmågor utvecklas i relation till arbetet med portföljerna och till det innehåll eleverna arbetar med i de skolämnen som fokuserades, och kan således ses som kompletterande det ämnesinnehåll som eleverna arbetade med. Ytterligare en aspekt av Hurtigs resultat som framstår som centralt med avseende på formativ bedömning, är de tre typer av stöttning som Hurtig urskiljer i sitt datamaterial. Hon benämner dem omfattande stöttning, kompletterande stöttning och minimal stöttning och beskriver dem enligt följande:

> Den första typen ges till elever som inte alls har arbetat med en funktion eller endast i ringa omfattning gjort det. Stöttningen innebär att pedagogerna stöttar eleverna i alla moment. Den andra typen av stöttning innebär att pedagogen hjälper eleven när denne frågar efter hjälp eller pedagogen märker att eleven är osäker. Den tredje typen av stöttning ges med hjälp av enkla ord eller fraser som bekräftar att eleven gör rätt. (s. 122).

Hurtig konstaterar att den stöttning lärarna ger resulterar i att eleverna klarar av att använda applikationerna på mer avancerade sätt. Vidare skiljer hon mellan stöttning och lotsning och konstaterar att lärarna förefaller använda lotsning då steget mellan elevernas aktuella kunnande och det applikationerna kräver är för stort. Till skillnad från lotsning, som gör eleverna beroende av läraren vid varje nytt tillfälle då samma moment ska utföras, resulterar samtliga tre typer av stöttning i att eleverna blir självständiga i sitt arbete. Hurtig kommenterar lotsning med att det förefaller vara en strategi lärarna i studien använder sig av för att lösa tillfälliga problem, till exempel tidsbrist.

Kjällander (2011) tar i sin avhandling, liksom Björklund Boistrup ovan, avstamp i multimodalitet för att studera hur elever interagerar, skapar mening och lär sig i klassrummet då de använder sig av digitala lärresurser i samhällsorienterande ämnen. Avhandlingen baseras på fem delstudier och det är främst den femte delstudien som är av intresse för vår översikt. Det arbetsområde (Europa) som studeras designas av läraren i en klass 8 didaktiskt så att det till sin karaktär påminner om ett dataspel. Av studien framgår dels att läraren ibland använder motstridiga teckensystem då hen bedömer elevernas kunskaper, dels att elevers tal och text erkänns

[23] Det Hurtig benämner bedömningsaktivitet resulterar i förmågan självreflektion – eller, uttryckt med hjälp av till exempel Black och Wiliam (2009), självbedömning.

som lärande i större utsträckning än vad deras användning av teckensystem såsom foton, färger och layout gör. Vidare visar studien att lärarens formativa bedömningshandlingar i huvudsak vägleder för lägsta betyg, medan lärarens didaktiska design av momentet och fokus i de summativa bedömningarna kretsar kring de höga betygen. Bedömningen sker med olika syften och mål. En slutsats Kjällander drar är att Eleverna själva förväntas forma sin förståelse i det glapp som uppstår mellan lärarens formativa intentioner riktade mot lägsta betyg och den avslutande summativa bedömningshandlingen.

I sin avhandling konstaterar Jessika Karlsson (2008) att en av de viktigaste aspekterna av musikutförande är förmågan att uttrycka känslor, men att detta är ett område som forskning relaterad till musikundervisning inte uppmärksammat i någon högre utsträckning. Avhandlingen syftade till att presentera och utvärdera ett nytt digitalt redskap för undervisning i känslomässigt uttryck i musikutförande och baseras på fyra delstudier. Den första delstudien undersökte hur musikundervisning går till i vardagen, med särskilt fokus på känslomässigt uttryck. Resultatet av denna studie visar att undervisningen bland de fem medverkande lärarna[24] karakteriserades av en brist på tydliga mål, specifika uppgifter och informativ feedback. Den andra delstudien var experimentell till sin karaktär. Ett nytt datorprogram introducerades, som analyserar musikutföranden för att ge informativ feedback med specifikt fokus på hur det känslomässiga uttrycket kan utvecklas. De tolv medverkande musikerna i åldern 15-35 år delades in i tre grupper: (1) feedback från datorprogrammet, (2) feedback från en lärare samt (3) repetition utan feedback. Resultatet från denna delstudie visade att den grupp som bäst utvecklade sitt känslomässiga uttryck vad den som fick feedback via datorprogrammet. Utvärderingen efteråt visade att musikerna tyckte att programmet var lätt att förstå och använda, men de ville ändå inte använda det för fortsatt lärande. Den tredje studien undersökte om oviljan att använda datorprogrammet handlade om generellt negativa attityder till datorer eller det handlade om innehållet i den feedback man fick från programvaran. Resultatet visade att feedback bedömdes som bättre av musikerna om de trodde att den kom från en lärare jämfört med om man trodde att den kom från ett datorprogram, men även att feedback som de facto gavs av en lärare bedömdes som bättre. Särskilt uppskattades lärares exempel och förklaringar till varför deras förslag skulle fungera. Sammantaget visar avhandlingen att musiker kan förbättra förmågan att uttrycka känslor genom datorstödd feedback, men också att programvaran behöver utvecklas så att de aspekter som lärarna kunde bidra med kan inkluderas i programmet.

Kapitlet av Ragnhild Löfgren och Viveca Lindberg (2011) i forskningsrapporten *Kemiundervisning, text och textbruk i finlandssvenska och svenska skolor: en komparativ tvärvetenskaplig studie* (Eriksson, red.) tar sin utgångspunkt i internationella och svenska studier av klassrumskommunikation och syftar till att dels kartlägga vilka frågor som ställs under en serie kemilektioner relaterade till det periodiska systemet och kemiska bindningar i två svenska lärares undervisning, dels jämföra resultatet med en tidigare studie av motsvarande undervisning i finlandssvenska lärares klassrum (Lindberg & Löfgren 2010). Ytterst handlar studien om vilket kemikunnande som efterfrågas och vilka slags svar eleverna förväntas ge i klassrumskommunikationen, som ses som mikrobedömningshandlingar ingående i kemilärarnas bedömningspraktiker[25]. Sammantaget har ca 130 frågor (88 frågor i de svenska och 43 i de finlandssvenska klassrummen) analyserats med avseende på kemiinnehåll och kemiskt kunnande, och frågornas karaktär och lärarens feedback till eleverna har jämförts. De likheter forskarna fann i klassrumskommunikationen inom svensk och finlandssvensk kemiundervisning var att bara en liten del av frågorna krävde mer än ett par ord av eleverna som svar, lärarna verkade inte ha utarbetade strategier för sina frågor med avseende på att använda frågorna för att få en uppfattning om vad som är svårt för eleverna att förstå och därmed vad läraren ska fokusera i fortsatt undervisning. Dessa delar överensstämmer resultat från tidigare internationella studier. Likaså var lärarnas feedback ofta kortfattad och visade främst hur eleven klarar sig i termer av bra/rätt eller dåligt/fel, medan feedback som mer nyanserat belyste uppgiftens syfte förekom sparsamt. Även denna del av resultatet bekräftar i stort internationella resultat. De skillnader som

[24] Eleverna valdes från den kommunala musikskolan, i gymnasieskolan och en musikhögskola. Förekomsten av elever i grundskoleålder har gjort att studien inkluderades.

[25] Inom ramen för forskningsprojektet som helhet bidrar studien till att synliggöra aspekter av kemilärarnas undervisningspraktiker (Eriksson, Ståhle & Lindberg 2011) i samma rapport, s. 76-113.

identifierades var att det som framstod som viktigt kemikunnande i det finlandssvenska klassrummet handlade om att kunna återge fakta och den ämnestradition som dominerade kan beskrivas i termer av disciplinär kemi d.v.s. fokus låg på naturvetenskapliga koncept, principer och lagar. I det svenska klassrummet fick eleverna flera frågor och läraren verkade ha intentionen att ge eleverna utrymme att förklara hur de tänkte. Principiellt efterfrågades samma typ av kunnande som i det finlandssvenska klassrummet, men på grund av att eleverna saknade redskap[26] för hur de skulle svara på den typen av frågor läraren ställde (t.ex. ”hur tänkte du?”) och på grund av de frågestrategier läraren använde hamnade resonemangen långt ifrån kemiämnet.

Eva Nybergs (2009) avhandling tar avstamp i formativ bedömning redan i formuleringen av syftet: ” att studera undervisning och lärande om växters och djurs livscykler i en lärandemiljö där formativ utvärdering betonas och där levande växter och djur är en utgångspunkt” (s. 26) och ramas in av ett socialkonstruktivistiskt perspektiv.

Hon betonar genomgående det krav på ömsesidighet som socialkonstruktivistiska perspektiv och formativ bedömning ställer på lärarens arbete: ” att läraren fortlöpande tar reda på hur eleverna förstår undervisningen och anpassar den till detta” (s. 39). Det handlar till exempel om att identifiera elevers framsteg och svårigheter, och att ge eleverna råd som gör det möjligt för dem att utveckla sitt lärande av ett specifikt innehåll, samtidigt som det handlar om att få syn på vad läraren, som en konsekvens av bedömningen, behöver göra för att utveckla sin undervisning. Nyberg baserar avhandlingen på en fallstudie av en lärares NO-undervisning i åk 5 i två klassrum. Hon följer först läraren i en klass och återkommer två år senare då läraren följande gång undervisar i årskurs 5. I avhandlingen ges rika beskrivningar av hur en lärare, genom att systematiskt följa elevernas kunskapsutveckling i en klass, lär sig vad som skapar svårigheter för eleverna med avseende på ett specifikt kunskapsinnehåll. Den systematiska uppföljningen ger läraren indikationer på vad hon behöver förändra i undervisningen för eleverna i denna klass, men även vad och hur hon ska förändra i sin planering till följande gång hon undervisar samma innehåll med en annan klass. Fallstudien fortsätter med elever i den klass som läraren undervisar två år senare med samma innehåll. En av huvudpoängerna i Nybergs (2008) avhandling pekar på den aspekt av formativ bedömning som ofta förbises: det handlar inte enbart om att identifiera elevers svårigheter för att hjälpa dem att förstå hur *de* ska utveckla sitt arbete. Det handlar också om att läraren får syn på hur *hen* behöver förändra sin undervisning för att på ett grundläggande sätt kunna bidra till att eleverna får bättre förutsättningar för sitt lärande.

Inom ramen för Max Strandbergs (2013) avhandling genomfördes två studier med bäring på den här översikten. Det huvudsakliga syftet för avhandlingen är relaterat till samarbetsläxor och ligger bortom översiktens syfte, men de två delstudierna har inkluderats. Den ena delstudien utgörs av en forskningsöversikt om studier av läxor å den ena sidan och läxor relaterade till formativ bedömning å den andra (Strandberg, 2013). Studien baseras på artiklar med peer review, handböcker, antologier och monografier. Framför allt i Sverige har läxor ända sedan husförhörens tid utgjort en aspekt av det som senare kom att bli skolans bedömningspraktiker, ändå har förvånansvärt lite uppmärksamhet ägnats åt just läxförhöret/genomgången av läxan som en del av de formativa bedömningspraktikerna. En stor del av läxgenomgången görs muntligt, varför Strandberg konstaterar att det behövs studier som förenar klassrumsinteraktion och läxgenomgångar. Den andra delstudien (Strandberg & Lindberg, 2012) berör undervisning för och om kulturell mångfald, med fokus på vilka typer av feedback som förekommer i en klassrumsdiskussion mellan elever i årskurserna 8 och 9, och vad återkopplingen kan innebära för elevers lärande. Med utgångspunkt i Tunstall och Gipps (1996) typologi identifierades sex typer av muntlig feedback från läraren som stödde elevernas lärande. Ytterligare en ny, sjunde kategori konstruerades, som inte tidigare dokumenterats. Denna sjunde kategori visade sig vara kontraproduktiv i förhållande till det läraren ville åstadkomma med undervisningen och kunde relateras till lärarens kulturella föreställningar. De övriga sex kategorierna bidrog, i högre eller lägre grad, till att stötta elevernas lärande i förhållande till syftet med undervisningen. Resultatet av studien pekar på behovet att uppmärksamma även oönskade effekter av den formativa bedömning som tar form som feedback till eleverna.

---

[26] Jfr Duschl (2000) som poängterar vikten av att eleverna guidas till att belysa ´the nature of science´ istället för att ge ett personligt svar.

Charlotta Vingsle (2014) följer en lärares undervisning i matematik (årskurs 5) under två och en halv månad i syfte att utforska vad som karakteriserar lärarens formativa bedömningspraktik, hur hon iscensätter olika aktiviteter under denna praktik samt vilka kunskaper hon använder i sitt utövande av formativ bedömning i samband med helklassundervisning i matematik. Fallstudien baseras huvudsakligen på ljudinspelade lektioner, men som stöd för ljudupptagning finns även videoinspelningar vilka används då oklarheter uppstår antingen i ljudkvaliteten eller om samtalet i klass rör sig om något specifikt (bild eller text på tavlan) som behövs för att förstå vad lärare och/eller elever hänvisar till. Såväl formativ bedömning (baserat på Black & Wiliam, 1998; 2009) som *pedagogical content knwoledge* (PCK) (Shulman, 1986; 1987) utgör studiens utgångspunkter. För analysarbetet har Vingsle konstruerat kriterier för urval av vilka episoder i datamaterialet som helhet som ska analyseras, d.v.s. sådana som utgör rika exempel och därigenom bidrar till att belysa det hennes studie handlar om. Genom detaljerade analyser av exemplen identifierar Vingsle fjorton aktiviteter vilka sammantaget bidrar till att få fram, tolka och använda information som läraren behöver för att bättre kunna svara mot elevernas behov i sin undervisning – allt detta sker inom ramen för Wiliams begrepp korta bedömningscykler, d.v.s. under ögonblick i klassrummet. Hon jämför aktiviteterna mellan korta, mellanlånga och långa cykler, och konstaterat att vissa aktiviteter går igenom oavsett bedömningscykelns längd, medan andra får större utrymme inom någon av de tre cyklerna. De kunskaper läraren behöver för sitt formativa bedömningsarbete är generellt och specifikt ämneskunnande, ämneskunnande och kunskap om eleverna, ämneskunnande och kunskap om undervisning samt pedagogiskt kunnande i kombination med kunskaper om eleverna och om kontexten. Dessa (kombinationer av) kunskaper relateras till bedömningscyklernas längd. Vingsle konstaterar att den mest krävande formen av formativ bedömning är den som representerar de korta bedömningscyklerna eftersom de är mest oförutsägbara – det går att planera för vissa aktiviteter, men förloppet är beroende av elevernas erfarenheter och förståelse. Vingsles slutsats är att formativa bedömningspraktiker är komplexa.

Den sista studien inom denna kategori skiljer sig delvis från de övriga genom att även elevernas perspektiv synliggörs. Utgångspunkten för Jessica Berggrens (2013) licentiatuppsats är relaterad till kamratbedömningens möjligheter i skolämnet engelska i högstadiet och att på så vis involvera eleverna i bedömningen av varandras och sina egna arbeten och skapa förutsättningar för eleverna att ta en aktiv roll i sitt eget lärande. Studien syftar till att bidra till förståelsen av vad elever kan lära sig genom att ge feedback på klasskamraters texter eleverna som kamratbedömare, samt att undersöka potentiella fördelar med att ge feedback. Studien baseras på två data från två klasser i årskurs 8. För det arbetsområde som studien är relaterad till var målet att eleverna skulle kunna skriva ett svarsbrev och undervisningen bestod av sex lektioner. Då arbetsområdet introducerades, användes modelltexter som utgångspunkt för att i klass diskutera vad som karakteriserar ett bra svarsbrev. Samtalen utmynnade i en gemensamt konstruerad kriterielista som utgjorde utgångspunkten för dels de svarsbrev eleverna skrev, dels den feedback de gav till sina kamraters brev. Sammantaget omfattade undervisningen således introduktionen som resulterade i de gemensamt formulerade kriterierna baserade på den feedback respektive klass och läraren resonerat sig fram till på basis av modelltexterna, en individuellt skriven text och kamratbedömning baserat på de gemensamt formulerade kriterierna. Data för studien omfattade de texter eleverna skrev, video och ljudbandade lektioner, elevernas feedback till varandra, enkäter till eleverna och intervjuer med dem om deras uppfattning om arbetsområdet och om sitt eget lärande. Eleverna kommentarer i feedbackgrupperna analyserades i syfte att identifiera formativ information, d.v.s. information som bedömdes kunde bidra till att ”främja mottagarens lärande genom att identifiera problem, förklara problemets natur, samt föreslå lösning” (s. 107). Vidare analyserades de ändringar eleverna gjorde för att revidera sitt första utkast, varefter förändringarna jämfördes med texterna i respektive feedbackgrupp och med de gemensamt skrivna kommentarerna i gruppen. Som indikation på elevernas lärande genom feedback sågs förändringar som kunde kopplas till innehåll och/eller kommentar. Även analyserna mellan de båda klasserna jämfördes och relaterades till såväl undervisningen som till de uppfattningar av lärande som eleverna gav uttryck för i enkäter och intervjuer. Studiens resultat visar att flera av de lösningar eleverna föreslog i sin feedback till varandra låg på en enkel och generell nivå – d.v.s. inte tog tag i specifika problem i texterna. Berggren menar de snarare fungerade som förstärkning av kriterierna än stöd för lärande. Av intervjuer och enkäter framgick att eleverna främst hade uppfattat att uppgiften syftat till att främja deras eget lärande, inte varandras. Det eleverna själva uppfattade att de lärt sig mest handlade om styckesindelning och disposition. De kopplade dock sitt lärande snarare till möjligheten att få läsa och bedöma sina kamraters texter än till den

feedback de fått. Så även om elevernas arbete kunde relateras till det arbete de gjort i sina feedbackgrupper var det ett fåtal elever som uppgav att de ansåg att de lärt sig något om innehållet genom att läsa och kommentera andra elevers texter. Berggren drar slutsatsen att de kriterier som utvecklades baserade på samtalen i klass snarare speglade elevernas aktuella kunskapsnivå än hade en utmanande funktion, vilket hon ser som en återkoppling till vad undervisningen i fortsättningen behöver beakta. Hon konstaterar vidare att även elever i den här åldern vinner på att ge varandra feedback, vilket är ett viktigt bidrag även till internationell forskning som, såsom vi tidigare påvisat. Slutligen konstaterar hon att studien visar att kamratrespons är viktig men inte kan ersätta den feedback läraren ger – de behövs båda men bidrar på skilda sätt till elevers lärande.

Formativ bedömning och skolans styrning

Christian Lundahls (2006) avhandling kan ses som en teoretisk studie av kunskapsbedömningar som fenomen, men den kan också ses som en empirisk studie baserad på textanalys av olika slags texter. Studien syftar till att, i relation till tidigare traditioner, förstå bedömningar ur ett kunskapssociologiskt och läroplansteoretiskt perspektiv. Fokus ligger på såväl interna (den pedagogiska praktiken) som externa (administrativt och psykometriskt baserade) bedömningspraktiker – hur de har formats över tid, manifesterats och förändrats. Textanalyserna görs av historiska material relaterade till politik, pedagogisk praktik, utbildningsvetenskaplig forskning, offentliga utredningar, tidskrifter som riktar sig till lärare och kursmaterial för fortbildning av lärare. Analyserna riktas mot hur de olika texterna samspelar i formeringen, manifesteringen och förändringen av skolans bedömningspraktiker. För vår översikt är det framför allt bedömningspraktiker inom den pedagogiska praktiken som är intressanta. Enligt Lundahl (s. 118) kan en pedagogisk praktik

> beskrivas utifrån hur den producerar kunskaper/information i syfte att strukturera, korrigera eller på annat sätt forma själva lärandet (formativa bedömningar). Det kan t.ex. handla om att läraren skaffar sig kunskaper om lärjungarnas kunskaper för att strukturera undervisningen, eller om hur lärjungar behöver kunskaper om sina framsteg och tillkortakommande för att strukturera sitt läsande (sic!)

Lundahl spårar ursprunget till vad som kan förstås som en formativ bedömningspraktik i svenska styrdokument till 1649 års skolordning, där argument för läxförhören förs fram på pedagogiska grunder i syfte att reproducera/memorera kunskapsinnehållet. Han följer sedan uttryck för formativa bedömningspraktiker genom styrdokumenten och visar hur och när de tonas ner och försvinner ur styrdokumenten i slutet av 1800-talet för att återkomma i ny form som ett senmodernt fenomen.

Joakim Krantz (2009) avhandling handlar om skolans styrning och om hur politiska anspråk (hur de riktas och motiveras) och professionell mening (förhandling om anspråken inom professionen och mellan professionen och styrningen) bidrar till att förändra lärares och lärarutbildares uppfattade autonomi. Anspråk och mening ses som relationella, varandra ömsesidigt konstituerande begrepp, som båda bidrar till vilken autonomi som tar form för lärares arbete. Avhandlingen baseras på två delstudier, en relaterad till lärarutbildning och en relaterad till grundskolan. Det är delstudien i grundskolan som är av intresse för vår översikt, då det innehållsliga exemplet berör IUP. Enligt Krantz pågår en ”särskild politisering kring lärares bedömningspraktik” (…) där det i ”den diskursiva kampen framträder särskilt ett spänningsfält mellan det summativa, mer kontrollinriktade och det formativa, mer utvecklingsinriktade” (s. 169). Krantz hävdar vidare att ”kunskaper då också framställs i perspektiv av att vara *bedömbara* och därigenom *styrbara* i kontrast till hur det förhåller sig för värden och mer vaga mål. Det är skolans demokratiska uppdrag som med den här framhävda tudelningen görs till det *ostyrbara*” (a.a. s. 169). I relation till detta argumenterar Krantz för att ”IUP definierat som bedömningspraktik gör det möjligt att resa anspråk på att bedömning är en professionell verksamhet som endast behöriga lärare kan och ska få ägna sig åt” (a.a. s. 169). Frågan om lärares professionella autonomi handlar enligt Krantz om att göra innehållsliga tolkningar av strävansmålen och skapa förutsättningar för kvalitativ fördjupning och reflektion under åren i grundskolan. Samtidigt finns det en spänning mellan de två typerna av mål (mål att sträva mot och mål att uppnå i Lpo-94) men också kritik mot hur skolan klarat uppdrag i allmänhet och specifikt hur lärares bedömningspraktiker har kritiserats på grund av

bristande likvärdighet[27] och måluppfyllelse. Resultatet av Krantz studie visar likhet med resultaten från Hirshs (2013) avhandling att lärares arbete befinner sig i spänningsfältet mellan nationell och lokal styrning, där dokumentationen ställer krav på lärarna att fatta beslut om komplexa pedagogiska frågor i relation till politiska och ekonomiska beslut. Sammantaget har den politiska styrningen av skolornas bedömningsarbete kringskurit lärarnas professionella autonomi vilket inneburit en ökad kvalitetskontroll, som i sin tur ökar kravet på dokumentation för bedömning. Även här finns en likhet i relation till resultaten från Hirshs (2013) studier: det förekommer tendenser till reducering av kunskapsbedömningarna relaterade till IUP som något instrumentellt, vilket också får konsekvenser för vad som bedöms – det som på ett enkelt sätt låter sig dokumenteras blir det som skrivs in i matriser medan det som är viktigt att bedöma är svårformulerat men också svårt att enas om. Ytterligare en aspekt av resultatet i Krantz avhandling är att de åtgärder som medverkande skolor genomförde för att öka likvärdigheten i lärarnas bedömningar mellan skolorna även ledde till att bedömningens administrativa, sorterande funktion slogs ut genom likriktningen.

Även Eva Forsberg och Christian Lundahl (2006) poängterar att den ökande användningen av kontroll- och utvärderingsinstrument på alla nivåer i systemet bidrar till att precisera – och därmed begränsa tolkningsutrymmet – av vad som anses som viktigt att kunna. Den ökade användningen av nationella prov och internationella jämförande kunskapsmätningar bidrar till detta (och preciseringen har ytterligare ökat sedan artikeln skrevs). I det offentliga samtalet om skolan har kommunikationen om elevers prestationer och skolors resultat ökat dramatiskt sedan millennieskiftet och den bidrar med konkretiseringar och tydliggöranden av vad skolan och eleverna förväntas/ges möjlighet ägna sig åt – samtidigt som den också begränsar vilket budskap som kan kommuniceras.

Barbro Westlunds (2013) avhandling baseras på en jämförelse mellan svenska och kanadensiska (från British Columbia) lärare för elever i årskurs 4. Jämförelsen syftar till att beskriva och analysera vilka bedömningsdiskurser av läsförståelse som är mest framträdande i intervjuerna med fem svenska och fem kanadensiska lärare, intervjuer som handlar om hur de stöder och bedömer sina elevers läsförståelse. Intervjuerna relaterades till klassrumsobservationer av lärarnas undervisning. Analysen av lärarnas svar gjordes i tre steg. I ett första steg analyserades varje enskild intervju för sig för att identifiera återkommande mönster, i följande steg jämfördes intervjuerna mellan de svenska lärarna för sig och mellan de kanadensiska lärarna för sig. Slutligen jämfördes intervjuerna mellan grupperna svenska och kanadensiska lärare. Resultatet visar att intervjusvaren inom respektive grupp av lärare bildar homogena mönster medan det förekommer betydande skillnader mellan grupperna. Westlund karakteriserar de svenska lärarna som *uppgiftsinstruktörer* och *kontrollanter av utförd uppgift* medan de kanadensiska lärarna karakteriseras som *instruktörer för att stödja läsförståelsen* och *kontrollanter av läsförståelsen medan den utvecklas*. I det fortsatta analysarbetet jämfördes lärardiskurserna med andra bedömningsdiskurser i respektive skolsystem. Här visar resultatet att lärarna ger uttryck för dominerande bedömningsdiskurser i respektive land och Westlund konstaterar att ländernas styrdokument föreföll ha stort inflytande på såväl lärarnas förhållningssätt till bedömning som till hur de hade tillägnat sig forskning av relevans för praktiken. En central skillnad mellan kanadensiska och svenska lärare var att de kanadensiska lärarna i högre utsträckning använde ett metaspråk då de talade om bedömning av sina elevers läsförståelse än vad som var fallet med de svenska lärarna. En annan central skillnad som Westlund identifierade var hur lärarna såg på vilken funktion feedback hade. Medan de kanadensiska lärarna i hög utsträckning beskrev betydelsen av att använda feedback för att förändra undervisningen, saknades den typen av beskrivningar helt bland de svenska lärarna. Vi ser Westlunds studie som betydelsefull även för att belysa de svenska resultaten inom detta tema – styrning får konsekvenser för lärares arbete och innehållet i styrningen är avgörande för formeringen av lärares formativa bedömningskompetens.

Den sista studien inom detta tema är en artikel av Anders Jönsson och Gunilla Svingby (2008) som baseras på ett uppdrag för Skolverket

[27] I relation till det betygssystem och den läroplan som Krantz avhandling handlar om kritiserade bl.a. Riksrevisionens utvärdering (RiR 2004:11) den bristande likvärdigheten i skolornas bedömningar, kritiken riktades dock specifikt mot staten (Utbildningsdepartementet och Skolverket) för att de inte i tillräcklig utsträckning skapat förutsättningar för likvärdighet i samband med införandet av 1990-talets betygssystem.

att ta fram underlag för att formulera kraven på en svensk nationell provbank för grundskolan med utgångspunkt i: (1) regeringens uppdrag, (2) mål i läroplan och kursplaner, (3) teoretisk och empirisk prov- och bedömningsforskning samt (4) nationella och internationella erfarenheter av provbanker. Kraven ska kunna användas för att utveckla provbankens innehåll och dess tekniska lösning (s. 57-58).

Översikten tar sin utgångspunkt i aktuella styrdokument och kritik mot framför allt bristande likvärdighet i bedömningar inom svensk skola, vilka bidrar till författarnas tolkning av sitt uppdrag. Själva översikten baseras dels på aktuella studier om bedömning, dels på konstruktionen av och studier om internationella och nationella provbanker. De slutsatser som formuleras på basis av analysen av provbanker handlar om de krav som provbanken ska möta. Det som Jönsson och Svingby ser som ett grundläggande krav är provbankens syfte, som de menar bör vara ”att stödja skolan och lärarna i ansträngningen att hjälpa eleverna att nå målen” (s. 77). Vidare menar de att provbanken ska kunna bidra till att förtydliga/konkretiserar målen. Denna förtydligande funktion kan även gälla frågan om likvärdighet i relation till lärarnas bedömning av elevers svar om provbanken inte enbart exemplifierar målen utan även är förebildlig genom att visa på olika sätt att uttrycka och bedöma elevers kunskaper. Principiellt tänker sig författarna att detta även kan resultera i kvalitetshöjning av lärarnas bedömningar, vilket internationella erfarenheter visar som en möjlig uppgift för en provbank. Ytterligare riktlinjer för provbanken gäller uppgifternas innehåll och utformning i förhållande till variationen av mål och bredden i innehåll, en kategorisering av uppgifterna avseende uppgiftstyper, bedömningsanvisningar, organisation (expertbaserad), användningsmöjligheter (flexibla), utbildningsmöjligheter (inbyggda i tekniken och valbara inom systemet) samt tekniska krav och kvalitetskriterier, de sistnämnda är beroende av vilket syfte som slutligen fastställs för provbanken.

Som en avslutande kommentar till Jönssons och Svingbys studie konstaterar vi att det i dagsläget finns provbanker i ett flertal av grundskolans skolämnen, flera av dem byggdes dock upp i relation till Lpo-94.

Sammanfattning av de svenska studierna

Även om studierna här är tämligen grundligt beskrivna och därmed utrymmesmässigt utgör en betydande del av denna översikt, är alltjämt svenska studier av lärares bedömningsarbete få, och mot den bakgrunden skulle man kunna säga att det enda vi kan bidra med är *indikationer* om lärares formativa bedömningspraktiker. På ett övergripande plan kan vi konstatera att de flesta studierna gäller yngre barn – upp till årskurs 6 – och att de flesta studierna är genomförda av lärare i forskarskolor eller lärare som antagits till traditionell forskarutbildning. Vad gäller resultaten så har vi funnit att dessa i flera studier stöder varandra, också över de tematiska gränser vi upprättat. Studierna av IUP framstår som starkast genom att datamaterialet totalt sett omfattar flera hundra IUP:er och genom att flera av studierna kombinerat dokumentstudier med intervjuer om IUP-mallar eller ifyllda IUP. Här finns också en tydlig indikation på att ämnesdidaktiska studier av IUP behöver komplettera de mer generella studierna – exempelvis visar Hartells studie relaterad till teknikämnet att det finns skolämnen som marginaliserats i åtminstone de IUP-mallar hon haft tillgång till. Studierna av styrning relaterad till formativ bedömning förstärker resultaten av studierna om IUPer – och tvärtom. Sammantaget visar studierna inom dessa två teman att Riksrevisionens (RiR 2011:23) senaste utvärdering av likvärdigheten i bedömning inom den svenska skolan behöver studeras i relation till den mellan stat och huvudmän delade styrningen av svensk skola. Klassrumsstudierna visar på komplexiteten i de situationer och det kunnande lärarna behöver hantera för att använda formativa bedömningar för att ömsesidigt utveckla sin undervisning och elevernas lärande.

3.3 Nordiska utblickar

Som tidigare nämnts har vi också valt att göra en kort nordisk utblick. Vi har redan tidigare redovisat tre av de artiklar Kari Smith står som författare till – ensam, eller som medförfattare (Gamlem & Smith, 2013; Smith 2009; 2011). Utöver dessa finns ytterligare en artikel (Engelsen & Smith, 2010), i vilken författarna utforskar norska lärares bedömningspraktiker, specifikt med fokus på feedback, inom ramen för ett 3-årigt FoU-projekt.

De finner att särskilt lågpresterande får väldigt lite feedback som skulle kunna bidra till deras lärande, men också att den feedback de får är vag och att den vanligen handlar om att förstärka deras självförtroende snarare än fokusera kunskapsinnehållet på ett sätt som kan informera eleverna om hur de ska komma vidare. Vi kan konstatera att Smith tills vidare verkar vara den forskare som engagerat sig mest i relation till formativ bedömning i grundskolekontext.

Även en studie av Päivi Atjonen (2014) ingår i den internationella delen av översikten. Här kompletterar vi med Atjonens (2007) analys av etiska aspekter på bedömning och med en artikel av Räty m.fl. (2000) som berör föräldrars uppfattningar om skolans bedömningar av deras barn. Atjonen relaterar aktuell forskning om bedömning till olika synsätt på etik och forskning om etiska bedömningar. Nyckelstrategierna för formativ bedömning ses som centrala, liksom frågan om rättvisa. Frågan om rättvisa måste ses i relation till såväl individen som samhället. För att säkerställa rättvisa räcker det inte med att enbart se till olika delaspekter av elevers prestationer, det behövs tillräcklig information om helheten i relation till det som ska bedömas. Delarna kan framstå som enkla att bedöma – här ställs krav på att lärare försäkrar sig om att det är det som är centralt för det kunnande som ska åstadkommas som också är det som bedöms, inte det som är enkelt att bedöma. Även kollegialitet och inkludering av eleverna i bedömningsarbetet ses som centrala aspekter för etiskt hållbar bedömning. Bedömning i sig är varken god eller ond/bra eller dålig, utan bedömningens kvalitet kan enbart avgöras i förhållande till de mål mot vilka bedömningarna relateras. De värden Atjonen kommer fram till som centrala för allt bedömningsarbete att bidra till är följande: de ska främja omsorgen om eleverna och vara rättvisa. För att detta ska blir möjligt bör bedömning bygga på principer som ömsesidighet, dialog och inkludering och vara framåtriktad. De krav som kan ställas på personer som genomför bedömningar är att de både ställer krav och är storsinta. Storsinthet utan krav innebär inte ett erkännande av individens kunnande, och krav utan storsinthet reduceras till ensidig kritik.

Artikeln av Räty m.fl. baseras på en enkätstudie som besvarades av 936 föräldrar (av 1700 utskickade enkäter d.v.s. svarsprocenten var 59) till barn på lågstadiet om deras uppfattningar om skolans bedömningar av deras barn. Bland frågorna ingick även sådana som syftade till att kartlägga vilket slags information föräldrarna upplevde att de ville ha om sina barns skolframgång och hur de helst skulle se att deras barn blev bedömda. Resultatet visar att föräldrarna var nöjda med skolans etablerade bedömningsformer – samtal med lärarna, skriftliga omdömen[28] och sifferbetyg. Bland de nya bedömningsformerna uppfattades självbedömning som relativt positivt, medan nationella jämförelser dels delade upp föräldragruppen i de som var för och de som var emot detta, dels påvisade att föräldrarna kände sig osäkra på dem. Vissa skillnader i svaren kunde relateras till föräldrarnas ålder, utbildning och kön: mammorna var signifikant mer intresserade av det enskilda barnets prestationer än papporna, ju äldre föräldrarna var desto mer intresserade var de av den relativa bedömningen av sitt barns (i relation till klassen, parallellklasserna) och av hur skolan klarade sig i förhållande till landets övriga skolor, och kvinnor gillade självbedömning signifikant mer än män. För självbedömning gällde också att ju längre utbildning föräldrarna hade desto mer positiva var de i förhållande till självbedömning.

Avslutningsvis en sammanfattning av de delar av Kristine Kousholts (2009) avhandling *Evalueret: Deltagelse i folkeskolens evalueringspraksis* som specifikt berör formativ bedömning. Studien baseras på klassrumsobservationer på tre skolor i årskurserna 4, 5 och 7 under sammanlagt 25 dagar och individuella intervjuer med lärarna, grupp- eller individuella intervjuer med elever i de deltagande klasserna, samt på en av skolorna även en intervju med skolledaren. Under de dagar Kousholt genomförde sina observationer och intervjuer förekom bedömning som del av undervisningen (I-R-E mönster i klassrumssamtalen, beröm och kritik) på alla tre skolorna (i mindre grad på en av dem än på de två övriga) i form av användande av mallar för

[28] Skriftliga omdömen får användas i Finland som alternativ till sifferbetyg (skalan 4-10) under årskurserna 1-7. De skriftliga omdömena har därmed primärt en summativ funktion. Det finns en finsk avhandling om dessa skriftliga omdömen (Mäensivu 1999) och innehållsligt förefaller de snarlika innehållet i IUP i termer av att en rätt stor andel av omdömena handlar om elevens person och beteende snarare än om kunskapsinnehållet och att det förekommer innehållsliga skillnader mellan de omdömen pojkar och flickor får. Avhandlingens titel: *Opettaja määrittelijänä – oppilas määriteltävänä: sanallisen oppilaan arvioinnin sisällön analyysi* [Läraren som bedömare – eleven som bedömd: en innehålls analys av skriftliga omdömen, på finska]. Jyväskylä studies in education, psychology and social research, 157. Jyväskylän yliopisto.

elevernas självbedömning, projektpresentationer med kollektiv/kamratbedömning av redovisningarna och utvecklingssamtal med föräldrar. Beskrivningarna av dessa bedömningspraktiker påminner om beskrivningarna från svenska skolor och klassrum. Under intervjuerna får lärarna frågor om hur de följer sina elevers kunskapsutveckling, och även i detta fall finns det likheter med resultaten från de svenska studierna:

> Kristine: Du siger, at du mærker, hvor de ligger henne. Hvordan gør du det?
>
> Birthe: Det gør jeg ved mit kendskab til dem, ved at se, hvor de er henne lige nu. Det gør jeg ved vores elevsamtaler, det gør jeg ved de tests, som jeg giver dem indimellem, ved vores.. hele tiden ved vores evalueringer. Det gør jeg ved at have berøring med alt, hvad de arbejder med hele tiden, at alt hvad de laver, det ryger i mine hænder på et eller andet tidspunkt, uanset hvad det så end er, de laver. Så jeg hele tiden kan se, at der er en eller anden form for progression i det, de laver. (…)
>
> Det Birthe peger på, at det at mærke elevernes udvikling er en bredere måde at orientere sig på end ved enkeltevalueringer. Birthe knytter det at mærke eleverne til forskellige former for evaluering (tests og andre evalueringer), men også til en bredere orienteringsproces. (s. 127).

Av elevintervjuerna framgår att eleverna tar över skolans förståelse av vad värderingar och bedömning går ut på. De går in i skolans bedömningspraktik och klassificerar sig själva och varandra som ’duktiga’ eller ’mindre duktiga’ i relation till skolans förväntningar. I och med att duktigheten alltid relateras till skolan (är kontextuell), får både lärarna och eleverna ensidiga intryck av de elever – ofta med utländsk bakgrund/annat modersmål – som framstår som mindre duktiga i skolsammanhang. Det präglar inte bara dessa elevers självförståelse – de tar också skolans bedömningar av sig själva och andra med sig in i andra sammanhang. Skolans bedömningar får således konsekvenser även utanför skolan genom att de bidrar till den identitet eleverna konstruerar om sig själva. Kousholt konstaterar att skolans institutionella strukturer och organisationen av undervisningen bidrar till att skapa (och återskapa) de sammanhang där eleverna framstår som duktiga eller mindre duktiga – det förefaller som att undervisningstraditionen bidrar till att lärarna återskapar situationer som baseras på föreställningar om en homogenitet som saknas, exempelvis i presentationer av grupparbeten där elever med annat modersmål än danska inte har tillräckliga språkliga kunskaper för att genomföra den redovisning de förväntas göra och inte kan svara på de frågor de får. Detta bidrar till att redovisningen framstår som ett fiasko och eleverna som okunniga. Genom att eleverna tar över skolans förståelse och förhållningssätt till bedömning bidrar de till att upprätthålla den.

Sammanfattningsvis kan vi konstatera att de nordiska studierna kompletterar översikten med resultat som påminner om dem från svenska studier, men också kompletterar framför allt genom delvis andra fokus och frågeställningar.

4 SLUTSATSER OCH DISKUSSION

Det övergripande syftet med föreliggande översikt har varit att systematiskt kartlägga och redovisa forskningsläget i Sverige och internationellt vad gäller formativ bedömning i grundskolan eller motsvarande. På ett generellt plan har vi kunnat konstatera att empiriska studier från grundskolemiljö är betydligt färre än studier gjorda på motsvarande gymnasienivå eller i högre utbildning. Bland de svenska studier som inkluderats i översikten ser vi däremot att det finns en tyngdpunkt på studier bland grundskolans yngre barn, årskurserna 1-6. Sett ur ett svenskt perspektiv - och med tanke på det stora genomslag begreppet formativ bedömning fått i svensk skola under senare år - är befintliga studier dock få och bristen på forskning högst påtaglig.

Vårt uppdrag har varit att syntetisera 2000-talets forskning kring formativ bedömning på en generell nivå – alltså inte i relation till specifika skolämnen. En av våra forskningsfrågor gällde dock huruvida det går att se likheter och skillnader mellan länder och mellan skolämnen i relation till formativ bedömning när man gör en läsning av så pass många studier som vi har gjort. På basis av de texter som inkluderats är skillnader mellan skolämnen svåra att konstatera, då ytterst få studier gör sådana jämförelser. Kingston och Nash (2011) konstaterar i sin metastudie kring effekterna av formativ bedömning att man tycks nå längst med formativa interventioner i matematik, medan Jönsson och Panadero (2013) utifrån sin översikt om matriser talar om att ämnesområde tycks spela mindre roll och att man får goda effekter av matrisanvändande oavsett ämne. Bland de artiklar vi har gått igenom kan konstateras att flest empiriska studier är gjorda inom ämnena matematik och modersmål (främst läs- och skrivinlärning). Svensk forskning på området är så knapp att det inte heller på basis av den går att dra några slutsatser om skillnader mellan skolämnen. Vi kan däremot konstatera att det finns studier vars resultat tyder på vikten av att studera formativ bedömning i relation till specifika skolämnen (exempelvis Hartell, 2012, som vi återkommer till längre fram i diskussionen).

Vad gäller likheter och skillnader mellan länder har vi i resultatkapitlet på flera ställen kunnat konstatera att det finns sådana. Begreppet formativ bedömning är inte nytt, men har - framför allt i och med Black och Wiliams översikt *Assessment and Classroom Learning* (1998) - under 2000-talet fått stor spridning internationellt på ett sätt som påverkat såväl policy som klassrumspraktik på ett liknande sätt i många länder (Berry, 2011a; Klenowski, 2011; Black & Wiliam, 2005; CERI, 2005). Den formativa bedömningens genomslag på policynivå är i hög utsträckning en följd av de kvantitativa mått på elevers ökade lärande som presenterats i översikter och metastudier (exempelvis Black & Wiliam, 1998; McDonald & Boud, 2003; Hattie, 2009). Hur lyckad och hållbar implementering av formativt arbete blir i praktiken tycks i hög grad vara beroende av de ramar och förutsättningar som ges, och här kan vi också konstatera att det på många håll tycks vara svårt att få till förutsättningar som ger tillräckligt utrymme för professionellt lärande och fördjupning. Följden har i många fall blivit att metoder anammas, men att fördjupad förståelse saknas, vilket vi återvänder till längre fram i detta kapitel. Den formativa bedömningens plats i summativa system är också ett genomgående tema som tycks medföra en likartad problematik i många länder. Vi har exempelvis sett att USA utmärker sig som varandes en ovanligt långt driven test- och mätkultur där det är svårt för formativa praktiker att få fäste (Flaitz, 2011; Black & Wiliam, 2005; Stiggins, 2002), men kan också konstatera att sådana tendenser i olika utsträckning även finns i flertalet andra länder. Vi ser också exempel på att enskilda länders policy och kultur kan begränsa och skapa mening i lärares arbete, och en jämförande studie som visar på intressanta skillnader är Westlunds (2013) avhandling, som vi också återkommer till längre fram i diskussionen.

Olika aspekter av kunskapsbedömning utgör centrala frågor internationellt, och den svenska kontexten torde inte skilja sig nämnvärt från den i många andra länder; också här har frågor som gäller bedömning varit starkt fokuserade i samband med en mängd reformer på skolområdet, och i synnerhet i samband med skrivandet och implementerandet av den senaste läroplanen i grundskola och gymnasium. Lärare och skolledare i Sverige har på ett allt mer påtagligt sätt fått ta sig an frågor kring bedömning i såväl summativa som formativa syften; å ena sidan kommer nu de formella betygen redan i årskurs sex samtidigt som nationella prov har införts i fler ämnen än tidigare och dokumentationen i form av omdömen om elever har ökat, å andra sidan har Skolverket tämligen tydligt propagerat för den formativa bedömningens plats i skolan genom allmänna råd och diverse andra skrifter/stödmaterial i bedömning (exempelvis Korp, 2011; Skolverket, 2011). Man kan se att svenska lärare i jämförelse med lärare i flera andra länder har stor makt över elevers betyg/fortsatta utbildningsbana, då det är

lärarna själva som bedömer nationellt standardiserade prov och avgör hur stor roll dessa ska spela vid betygssättning (jfr Black & Wiliam, 2005). Samtidigt kan man konstatera att Sverige är ett land som under det senaste decenniet i allt högre grad gått mot ökad extern kontroll av såväl elevers kunskaper som lärares arbete. Sverige hör exempelvis till de länder som är allra mest öppna med att publicera statistik över skolors resultat på nationella prov. I kombination med det fria skolvalet och antalet friskolor, som utmärker Sverige jämfört med många andra länder i Europa och övriga världen, gör detta att vi har en situation där extern kontroll och ansvarsskyldighet ändå är tämligen påtagliga fenomen (EACEA, 2009; jfr också Lundahl, 2009). Måhända kan man tycka att detta inte har med formativ bedömning att göra, men ett återkommande tema i de artiklar som gåtts igenom rör den formativa bedömningens plats i summativa system, och det faktum att det finns svårigheter i att förena bedömningsfunktionerna när redovisningskuluren är påtaglig. Vad gäller formativ bedömning specifikt tycker vi oss, utifrån egna erfarenheter, kunna konstatera att problematiken i det svenska sammanhanget inte skiljer sig nämnvärt från den som beskrivs i flertalet internationella studier. Även i Sverige (och övriga Norden) brottas lärare med problematik som rör den formativa bedömningens plats i summativa system, och även här kan man se att såväl kunskaperna hos olika nivåer av aktörer i skolsystemet som organisatoriska faktorer får konsekvenser för hur formativa praktiker utvecklas (jfr Hirsh, 2013; Kousholt 2009; Krantz, 2009; Westlund 2013).

4.1 Formativ bedömning – teori, begrepp och metoder

I vår kartläggning ingick att inventera vilka teorier som legat till grund för forskning om formativ bedömning, hur formativ bedömning har uppfattats och använts i forskningssammanhang, samt vilka metoder som utprövats/använts. I översiktens inledning skrev vi att olika tanketraditioner och teorier om lärande har format/formar förståelsen av bedömning och får konsekvenser för hur begreppet formativ bedömning har uttolkats. Under analysförfarandet kunde vi konstatera att en stor del av vårt material utgjordes av icke-empiriska artiklar som syftar till att problematisera och definiera innebörden av begreppet formativ bedömning. Något förenklat skulle man kunna säga att de flesta är överens om att formativ bedömning handlar om att forma och utveckla lärande. Innebörden och konsekvenserna av detta kan dock te sig olika beroende på den förståelse av lärande man har, vilket är viktigt inte bara för definitionen av begreppet formativ bedömning, utan också för det praktiska arbete som utvecklas på basis av hur man förstår lärande (Black, 2001).

Vi förstår lärande som något som inte enbart försiggår inne i elevens huvud, utan snarare som något som äger rum i interaktionen mellan lärare och elever i en viss miljö som är ordnad på ett visst sätt. Med den förståelsen handlar formativ bedömning inte enbart om att forma eleven, det handlar i lika hög grad om att forma undervisning och undervisningsmiljö. I de artiklar och bokkapitel som definierar och positionerar begreppet formativ bedömning teoretiskt finns genomgående en förståelse av begreppet som vilandes på två ben, där bedömningsinformation kring enskilda elever används för formandet av eleverna å ena sidan, och där bedömningsinformation kring såväl enskilda elever som elevgrupper ligger till grund för analys och förändring/formande av undervisning å den andra. Vi hävdar också, i likhet med flera forskare i vår översikt, att man inte kan isolera förståelsen av elevers lärande från den undervisning de får eller den miljö de finns i när de ska lära i skolan. Vi kan inte heller isolera förståelsen av lärares arbete med undervisning från den större kontext som bidrar till att forma deras arbete, i termer av organisation och villkor som styr dem på olika nivåer.

Teoretiskt skrivs alltså formativ bedömning fram som vilandes på två ben, men i vår genomgång av empiriska studier kan vi konstatera ett tämligen stort fokus på det ena benet: det är eleven som ska formas snarare än undervisningen eller miljön. Man studerar exempelvis hellre effekten av en viss given feedback från lärare till elev genom olika kvantitativa för- och eftertester, än kvaliteten i den feedback som ges. I än mindre utsträckning studerar man vilka förändringar som görs i undervisning eller undervisningsmiljö på basis av elevernas resultat. Den figur över nyckelstrategier (se exempelvis Wiliam & Thompson, 2007) som kommit att spela så stor roll för såväl empirisk forskning som det storskaliga implementerandet av formativa bedömningspraktiker i klassrum världen över (inte minst i Sverige) handlar i hög grad om att forma eleven. Eleven ska förstå målen för lärande, eleven ska ges feedback och eleven ska tränas i metakognitivt tänkande. Naturligtvis är detta viktiga aspekter som behöver studeras i högre utsträckning än vad som hittills har gjorts med tanke på hur utbredda sådana praktiker är i klassrum världen över, men vi menar att man i högre grad

också måste betona och beforska hur undervisning och miljö formas på basis av den formativa bedömningen. Då handlar det om hur lärare använder bedömningsinformation för att analysera sin undervisning och göra förändringar i den, eller om hur skolledare analyserar aggregerade resultat och använder dem som en del i skolans kontinuerliga systematiska kvalitetsarbete. Det sistnämnda kan även gälla huvudmän på en än mer övergripande nivå. Sådana frågor, som rör hur analyser görs i syfte att förändra och förbättra undervisning utifrån elevers resultat – vare sig det är på konkret klassrumsnivå eller övergripande skolnivå – vet vi väldigt lite om. Här finns dock några svenska klassrumsstudier som bidrar med kunskap om den komplexitet det handlar om (jfr t.ex. Vinglse 2014) och de varierande förutsättningar som olika lärare har (jfr t.ex. Hartell 2012, Hirsh, 2015).

De metoder som används för att studera formativ bedömning är av varierande slag. Som vi vid ett flertal tillfällen konstaterat finns en tydlig tendens att studier i vilka man använder sig av kvantitativ ansats får större genomslagskraft i relation till policy. Meta-studier, som slår samman resultat från olika kvantitativa studier, förekommer tämligen frekvent. Vi har berört att det i fallet formativ bedömning finns en viss problematik relaterad till meta-studier, då paraplybegreppet innefattar en rad olika företeelser. Vi återkommer till detta längre fram i denna diskussion. Kvantitativa studier för att studera effekten av en viss intervention är ofta uppbyggda enligt modellen *förtest – intervention – eftertest*, där mellanledet ofta är en process som pågår under en kortare tid (1-3 månader). Inte sällan använder man sig av en blandning av metoder för att samla in data, där exempelvis ovan nämnda typ av studie kombineras med enkäter till elever och/eller lärare/skolledare och intervjuer enskilt eller i grupp. På ett generellt plan kan sägas att i de få fall elever intervjuas sker det oftast i fokusgrupper, medan det är vanligare att lärare/skolledare intervjuas individuellt. Naturligtvis finns också många exempel på studier som enbart har kvalitativ ansats, där olika varianter av intervjuer (face-to-face eller genom skrivna frågor där öppna svar förväntas) eller en kombination av intervjuer och observationer har använts. I det svenska fallet har dokumentanalys varit en vanlig metod, då man i flera fall har studerat hur formativ bedömning kan komma till uttryck genom olika typer av skriftlig dokumentation kring elever. En variation av metoder finns således representerade bland de studier vi har tagit del av, och det förefaller också uppenbart att en variation av metoder såväl som studieobjekt är vad som måste till för att kunna bredda såväl som fördjupa forskningsfältet framöver.

4.2 Formativ bedömning – hinder och möjligheter för ett framgångsrikt arbete

Genomgående i många av de teoretiska artiklar som gåtts igenom i föreliggande översikt framgår vikten av att de olika nivåerna av aktörer som implementerar arbetet i praktiken har en fördjupad teoretisk förståelse av vad begreppet innebär. Centralt är att begreppet formativ bedömning behöver förstås som kopplat till att forma inte bara elever utan även undervisning. Att använda bedömningsinformation för att analysera och förändra undervisning är en komplicerad process, vilket är en anledning till att implementering behöver innefatta robusta utbildningsinsatser och kontinuerligt professionellt lärande. Vi har redan kunnat konstatera att empirisk forskning hittills har fokuserat mest på formandet av elever, och inte alls i lika hög grad på formandet av undervisning och undervisningsmiljöer, vilket kan tänkas vara en orsak till att även implementerande nivåer av aktörer uppfattar begreppet så. En annan anledning kan förmodas hänga samman med sätten på vilka forskningsresultat sprids till policy- och skolnivå. Då man vill tillgängliggöra forskning för verksamma lärare och skolledare populariseras och förenklas den för att de lättare ska kunna ta till sig den. Hultman (2015) har studerat överföring av forskningsbaserad kunskap till skolan, och menar att något ofta går förlorat när forskning ska populariseras och spridas. Dels handlar det om att det något komplicerade forskningsspråket översätts till ett mer populärvetenskapligt språk, och dels beror det på att delar av den teoretisering och problematisering som finns i originalstudierna skalas bort och aspekter av analysen försvinner. Risken är då att forskningen blir vad Hultman kallar ”herrelös”, det vill säga att metoder anammas på ett något oreflekterat sätt. Vi menar att forskningen kring formativ bedömning är ett exempel på att det i viss mån har blivit på detta sätt. Vi kan i sammanhanget dra paralleller till Gardner m fl (2011), som beskriver skillnaden mellan professionell utveckling och professionellt lärande, där interventioner som innefattar professionellt lärande i samband med

implementeringen av formativa praktiker tenderar att leda till en mer medveten och därmed också hållbar förändring. Skillnaden mellan professionell utveckling och professionellt lärande torde hänga samman med hur implementeringsarbetet går till: om all personal som förväntas anamma formativa arbetssätt involveras i utbildningsinsatser och grupper organiseras där kollegialt lärande ges utrymme menar Gardner m fl (a. a.) att effekten av implementeringen blir en annan än om bara vissa nyckelpersoner får sådan utbildning och sedan förväntas delge övriga (jfr även Timperley, 2013). Birenbaum m fl (2011) menar att den internationellt spridda diskursen kring de positiva effekterna av formativ bedömning leder till att så kallade pseudo-formativa praktiker utvecklas bland många lärare, då de ”vet att det är bra” med formativ bedömning och därför använder sig av viss metodik (relaterad till någon/några av de nämnda nyckelstrategierna), men förhållningssättet tenderar att vara ateoretiskt. När det gäller frågan om under vilka villkor arbete med formativ bedömning tycks fungera framgångsrikt är således *ett* svar som är möjligt att ge, att arbetet tycks fungera mer framgångsrikt då de implementerande aktörerna inte enbart förses med instrument/metoder, utan även fördjupar sin teoretiska kunskap kring processer och får möjlighet till kontinuerliga kollegiala diskussioner. För detta krävs en organisation som ser värdet av fördjupning och ger de ramar som krävs. Vad gäller utvecklandet av formativa klassrumspraktiker kan det finnas fog för att prova och utforska olika modeller för kollegialt lärande/kollegial reflektion. *Professional learning communities*, där grupper av lärare regelbundet träffas och diskuterar sin undervisnings-/bedömningspraktik med varandra är en väg att gå. En annan – möjligen kompletterande – metod är det så kallade *open lesson*-formatet, där ett antal lärare sitter med och observerar givna aspekter av en kollegas lektion för att sedan, i en efterföljande session göra en kollektiv analys av dessa aspekter (jfr Miyakawa och Winsløw, 2013).

Vi har konstaterat att bedömningsfrågor (relaterade till såväl summativa som formativa syften) är helt centrala för en rad olika aktörer på olika nivåer i skolsystemet; lärare arbetar med bedömningsrelaterade frågor dagligen och stundligen, skolledare behöver fördjupade kunskaper i bedömningsfrågor för att kunna ta vara på skolans aggregerade resultat i det kontinuerliga kvalitetsarbetet, och huvudmannanivån behöver kunna använda bedömningsdata på ett omdömesgillt sätt i övergripande analyser. Samtliga nämnda är också centrala aktörer då det gäller implementering av formativa praktiker, och vår översikt har visat att alla nivåer av aktörer behöver vara inblandade om vi ska få ett förändrat tänkande kring och arbete med formativa bedömning i praktiken. Starkare broar behöver dessutom byggas mellan forskning, policy och praktik, då det gäller att hitta sätt att arbeta formativt inom ramen för summativa system, och försäkra sig om att klassrumsbedömning kopplas till systematiskt kvalitetsarbete på skol- och systemnivå (jfr CERI, 2005). Ytterligare en aktör som spelar avgörande roll för den kunskap som utvecklas kring bedömningsfrågor är naturligtvis lärarutbildarnivån. Studier vi har gått igenom (Andersen m fl, 2010; Mansell & James, 2009; Smith, 2009; 2010; 2011) visar att det finns alldeles för få personer på lärosäten som har den fördjupade och forskningsbaserade kompetens som krävs för att adekvat kunna utbilda blivande lärare och fortbilda redan verksamma. Smith menar med referens till den norska situationen att läget är allvarligt och att omfattande insatser måste till för att höja kompetensen på landets lärosäten. Även kring denna fråga finns lite svensk forskning att relatera till.

Det finns över huvud taget ett stort behov av studier som på olika sätt undersöker och problematiserar såväl organisation av som arbete med formativa praktiker i klassrummen. Sådan forskning bör vara av olika slag: det handlar om allt från interventionsstudier som över tid mäter effekten[29] av olika aspekter av formativa praktiker till kvalitativa undersökningar som fördjupar vår kunskap om olika aktörers förståelse, erfarenheter och uppfattningar[30] av formativt arbete. Den stora bristen på studier som tar elevers perspektiv talar dessutom för att sådana är nödvändiga.

---

[29] Dock med förbehållet att operationaliseringen av begrepp måste kunna göras tillgänglig – något vi sett som ett genomgående i de studier som baserats på kvantitativa data, d.v.s. den kvalitativa grunden för storskaliga effektstudier.

[30] I relation till kvalitativa studier ser vi ett behov av uppskalning – såväl internationella som svenska studier baserade på intervjudata bygger ofta på småskaliga studier – vilket principiellt är nödvändigt för att kunna hantera mängden data, men här krävs samverkan mellan lärosäten för att bygga storskaliga projekt baserade på kvalitativa data.

Genom att gå tillbaka till den förteckning över studier som Black och Wiliam presenterade som en rekommendation baserad på deras översikt från 1998, kan vi konstatera att dessa fortfarande äger giltighet, även om det finns studier som bidrar till att svara på en del av frågorna:

- Vilka antaganden ligger till grund för läroplanen och för lärares didaktiska arbete?
- Vilken/vilka rationalitet/-er ligger till grund för hur det elever förväntas lära sig designas och presenteras?
- Vad karakteriserar det som tas som intäkt för kunnande i elevers svar i relation till olika typer av bedömning?
- Vilka tolkningsramar använder sig lärare och elever av i relation till elevernas svar?
- Vad karakteriserar det lärande som kommer till uttryck i relation till de uttolkningar som kan identifieras?
- Hur fördelas ansvaret mellan lärare och elever i dessa processer?
- Vilka uppfattningar och föreställningar har elever om sig själva som lärande personer, om sitt eget lärande, om syftet för sina studier och om sina studiemetoder?
- Vilka uppfattningar och föreställningar har lärare om lärande, om förmågor, om sina elevers framtid och om sin egen roll som bedömare?
- Vad karakteriserar de sociala sammanhangen i klassrum, såsom de utvecklats av elever och lärare och av de yttre ramar dessa föreställer sig att skolsystemet skapar?
- Frågor relaterade till ras, samhällsklass och genus – dessa är föga uppmärksammade i forskning om formativ bedömning
- I vilken mån är kontexten för någon studie överhuvudtaget artificiell och vilka möjliga effekterna har det för resultatens generaliserbarhet? (s. 58-59).

På den första frågan visar studier att de antaganden det rör sig om är knutna till tid och plats kollektivt. Läroplaner utvecklas i samspel med samtida föreställningar om kunskap, lärande och bedömning. Samspelet involverar flera typer av aktörer – vilka aktörerna är varierar mellan länder och i relation till de system de ingår i. Samspelet innebär att det inte finns en linjär relation från policynivån till lärares didaktiska arbete på klassrumsnivå. Här framstår de aspekter av resultaten av Kousholts (2009) avhandling som intressanta att ha som utgångspunkt även för studier i svenska sammanhang, med tanke på elever som aktörer i betygssystemet. Likaså ger resultatet från Westlunds avhandling intressanta utgångspunkter: om skillnader i policy och kollegial praxis manifesteras som olika förståelse av formativ bedömning och praxis i relation till vad formativ bedömning används för: hur agerar eleverna i det kanadensiska systemet (eller i vilket annat betygssystem som helst), om de, likt de danska eleverna, bidrar till att upprätthålla och återskapa det system de utgör en del av? Vad betyder detta för elevers lärande i olika system? Vad karakteriserar skolframgång (”duktighet”) i olika system? På motsvarande sätt kan diskussionen föras i relation till de övriga punkterna.

Utöver de punkter som Black och Wiliam pekar ut som viktiga finns andra frågor som blivit angelägna att studera i relation det utbredda arbetet med formativa metoder i svensk skola. Man arbetar idag exempelvis i mycket hög utsträckning med pedagogiska planer och matriser, utan att vi vet särskilt mycket om kvaliteten på dessa. Frågor som framstår som väsentliga att studera är bland annat följande: Hur ser de dokument man använder ut i termer av kunskapssyn, innehåll och kommunicerbarhet? Hur skapas och används dokumenten av lärare? Hur uppfattas och används de av elever? Leder användandet till förändrad undervisningspraktik – på vilka sätt? Får man de effekter man eftersträvar? Vet man vilka effekter man eftersträvar? Under vilka kontextuella omständigheter får man de effekter man eftersträvar? Samma frågor kan ställas kring arbetet med olika former för feedback, där det dessutom framstår som viktigt att se närmre på relationella/emotionella aspekter, framför allt hos eleverna.

4.3 Värdet av befintlig forskning om formativ bedömning i relation till svenska sammanhang

Följande text handlar om vilka slutsatser det går att dra utifrån svensk och internationell forskning. Utgångspunkten för det inledande avsnittet är vad som karaktäriserar modeller och förutsättningar för formativ bedömning för vilka det finns indikationer på att måluppfyllelse ökar.

Den svenska forskningen kring formativ bedömning är dels knapphändig, dels fokuserar de få studierna skilda aspekter av formativ bedömning, vilket gör att även de internationella studierna måste beaktas för att se

vad som är möjligt att uttala sig om. Vi har tidigare påtalat att det finns en problematik i att generalisera forskningsresultat mellan kontexter. I en jämförande fallstudie av tyska och svenska lärares kompetensutvecklingskultur (representerade av delstaten Berlin i Tyskland och Stockholmsregionen i Sverige) visar Forsberg och Wermke (2012) och Wermke (2013) att den samtida övertron på att erfarenheter från en kontext på ett enkelt sätt kan överföras till en annan kontext är problematisk. Faktorer som till exempel vilka aktörer inom kompetensutvecklingsfältet som de båda lärargrupperna som ingår i studien har förtroende för skiljer sig åt. Det är en av de kontextuella faktorer som identifierats som betydelsefull. En annan faktor handlar om skillnader i den autonomi lärarna tillskrivs och upplever, och som Wermke relaterar till historiskt framvuxen utifrån skilda villkor. Wermkes studie har särskild relevans i detta sammanhang, då det exempel Forsberg och Wermkes artikel bygger på handlar om kompetensutveckling relaterad till kunskap och kunskapsbedömning. Förutom redan tidigare nämnda studier av Hirsh (2013) och Krantz (2009) visar exempelvis Westlunds (2013) avhandling på ett påtagligt sätt vad kanadensisk och svensk kontext betyder för de båda lärargruppernas kunskaper om och förhållningssätt till formativ bedömning av läsutveckling och vad det kan vara. Till exempel måste de effekter man har sett i internationella studier prövas både i och i relation till svensk kontext innan det går att påvisa att det som gäller i ett land, med ett specifikt betygssystem och med vissa villkor för lärares arbete, även skulle kunna gälla i ett annat land med ett annat betygssystem och med andra villkor för lärarnas arbete. På motsvarande sätt är det problematiskt att generalisera effekter man sett i studier genomförda i högre utbildning till exempelvis barn på lågstadiet. Det finns ingen anledning att tro att högkvalitativ feedback från lärare till elev skulle fungera sämre med yngre barn än med vuxna studenter, men omständigheterna är så pass olika att högkvalitativ feedback i så disparata kontexter skulle vara två helt olika företeelser. Exempel på sådana omständigheter är lärarnas ämnesmässiga och didaktiska kompetenser, där behörighetskraven skiljer sig åt för tjänster i skola och högre utbildning, grundskolan som ett obligatorium för eleverna och studenternas val av (och urvalet till) högre utbildning och skillnader i förväntningar på eget ansvar. Ytterligare ett exempel handlar om hur man inom högre utbildning kan hantera självvalda eller uppifrån implementerade förändringar genom FoU-projekt som följs upp, dokumenteras och rapporteras, medan detta förfarande – tills vidare – hör till ovanligheterna inom skolan (jfr Lindberg 2005).

Vi ifrågasätter inte att arbete med de olika nyckelstrategierna skulle kunna leda till positiva resultat för eleverna på varierande sätt. Stora översikter (exempelvis Black & Wiliam, 1998; Hattie, 2009) presenterar effektstorlekar som de uppger är betydande. Vi har dock också sett att man behöver förhålla sig kritiskt till sådana siffror av olika anledningar. Många disparata företeelser slås samman under paraply-begreppet formativ bedömning, det förekommer att statistiska analyser är gjorda på mycket små underlag, liksom att studier som inte varit utsatta för kritisk granskning tas med, och att kontextuella/modererande faktorer inte tas hänsyn till – åtminstone framgår detta inte i flertalet av de texter som ingått i översikten.

Eftersom paraply-begreppet formativ bedömning (utifrån de fem nyckelstrategierna) innefattar så varierande aspekter som tydliggörande av mål och kriterier för lärande, skapandet av klassrumsmiljöer som på ett effektivt sätt involverar eleverna och ger dem möjlighet att kontinuerligt visa så mycket som möjligt av sin kunskap, feedback från lärare till elev och elever emellan, samt feedback till lärarens undervisning, innebär det stora svårigheter att studera effekterna av formativ bedömning som helhet. Det torde vara såväl omöjligt som oetiskt att skala bort allt klassrumsarbete som kan klassas som formativt för en grupp elever under en lång period, i syfte att kunna studera och jämföra med en annan grupp elever som fått ta del av alla formativa praktiker under samma period. Några sådana studier har, oss veterligen, heller aldrig gjorts. I boken *Assessment for Learning: Putting it to practice* (Black m fl, 2003) beskrivs ett tvåårigt projekt där 19 lärare (företrädelsevis i matematik och NO) på motsvarande högstadiet i England genomför en rad förändringar i sin undervisningspraktik. Förändringarna var relaterade till sättet att ställa frågor i klassrummet, feedback på elevuppgifter, arbete med kamrat- och självbedömning samt formativt användande av summativa test. Hur lärarna valde att arbeta med dessa aspekter var individuellt och inte systematiserat, och projektets data var i första hand kvalitativa och byggde på intervjuer med lärare och observationer av klassrumssituationer. På något sätt ville man ändå försöka mäta presumtiva kunskapsvinster, och gjorde därför uppskattningar på basis av standardiserade tester som genomförs i givna årskurser i det engelska skolsystemet. Resultaten i klasser som inte deltagit i projektet jämfördes med resultaten i de klasser som deltagit, efter att diverse kalibreringar hade gjorts. Även om sådana jämförelser i de flesta fall visar på positiva effektstorlekar i storleksordningen 0.2-0.3 är detta inte att betrakta

som robust forskning (vilket forskarna själva problematiserar), då sådan hade ställt helt andra krav på forskningsdesign och reliabilitet. Vi menar att frågan om vad som karaktäriserar ”modeller” för arbete med formativ bedömning för vilka det finns indikationer på att måluppfyllelsen ökar i stort sett är omöjlig att svara på, just för att paraply-begreppet innefattar för många disparata företeelser, av vilka många är helt naturliga delar i de allra flesta lärares undervisningspraktiker.

Även om den mångfald av aspekter som kan relateras till formativa bedömning svårligen kan slås samman till en given modell med vars hjälp man skulle kunna mäta effekterna på ett sätt som leder till att man kan uttala sig om den i relation till elevers måluppfyllelse, finns påtagliga behov av att i svensk kontext studera effekterna av olika interventioner. Vår erfarenhet är att det i svenska klassrum i hög grad arbetas med metoder som har sitt ursprung i nyckelstrategierna, och det finns fog för att undersöka vad sådant arbete leder till på flera olika plan. Vi bedömer att bedömningsforskning relaterad till svenska förhållanden är så knapp att större insatser behövs än enskilda studier. Tills vidare saknas även klassrumsstudier relaterade till frågor om genus. Här har till exempel Murphy och Ivinson (2005) visat att även formativ bedömning kan bidra till att pojkar och flickor får tillgång till olika kunskapsinnehåll. Murphy (2008) visar dock också att det inte behöver vara så, utan lärare som förstår vad som bidrar till att skapa genusstereotypa fördelningar av agens och därmed sammanhängande tillgång till kunskap kan motverka detta genom strategisk planering av specifika situationer. Inom antologin *Cultural Validity in Assessment* (del Rosario Basterra m. fl., 2011) pekar två av kapitlen specifikt på behovet av att använda formativ bedömning för att klargöra vad som handlar om bristande språkförståelse och vad som handlar om bristande förståelse av ämnesinnehållet (Kapriva & Sexton, 2011; Durán, 2011). Vi kan konstatera att studier av olika slag behövs: dels rörande eventuella effekter (dock med förbehållet att den kvalitativa grunden för operationaliseringen av studien redovisas) av formativt arbete för elevers lärande/måluppfyllelse, men också rörande lärares och skolledares förståelse/uppfattningar/arbete och huvudmännens roll i att skapa förutsättningar. Ytterligare en aktör är också central i sammanhanget, och behöver studeras för att få fördjupad kunskap om läget i svensk bedömning – landets lärarutbildningar. Vi har i översikten kunnat konstatera att man i flera länder ser brister hos den nivå som ska utbilda blivande lärare och fortbilda redan verksamma. Fördjupad kunskap saknas eller finns hos allt för få, med påföljd att studenterna får alltför knapp och i vissa fall inte särskilt välgrundad kunskap i bedömningsfrågor.

En sista fråga för diskussion gäller de metodiska aspekterna av studierna. Vad gäller de studier som baserats på kvantitativa data har vi konstaterat att de metodiska aspekterna sällan är redovisade på ett tillfredsställande sätt. De flesta är måna om att ange vilka statistiska analysmetoder metoder de använder, men argumentationen för dem i relation till variablernas informationsvärde tas för given. På motsvarande sätt saknas grunden för hur frågeområden och enskilda frågor operationaliserats t.ex. för enkätstudier, vilket gör att läsaren lämnas åt sitt öde att försöka avgöra rimligheten i valet av såväl statistiska metoder som resultatens validitet – den kvalitativa grunden för studierna redovisas inte på ett tillfredsställande sätt. Just detta utgör en aspekt av den kritik som riktats mot de så kallade meta-studierna: att de statistiska beräkningarna som gjorts i varje studie hanteras som om de var jämförbara med varandra utan att man granskat operationaliseringsprocessen – eller om man gjort det, så redovisas inte de delarna. Detta ser vi som en kvalitetsbrist i flera av studierna. På motsvarande sätt kan flera kvalitativa studier kritiseras för att varje forskare vill utforma sina studier individuellt, vilket gör att de blir svåra att jämföra i relation till varandra. Vi kan se att det finns fördelar med att inte bara producera nytt utan också bygga vidare på och pröva befintlig forskning. För det tredje så saknar vi en typ av studier, nämligen samverkan mellan studier baserade på kvalitativa respektive kvantitativa data.

4.4 Kartläggningsmodellens starka och svaga sidor

Både vad gäller internationella och svenska databaser finns det svårigheter med avseende på systematiska översikter. Systematiken i de systematiska översikterna begränsas av databaserna som administrativa söksystem, konstruerade för forskare inom vitt skilda forskningsfält av administratörer som är kunniga på bibliotekssystem. De administrativa söksystemen utgör både stöd och begränsningar, och en sådan begränsning handlar om att forskningsfält förändras över tid vilket gör att klassificeringar skapade under en tidsperiod kan skapa hinder för att komma åt det söksystemen är avsedda att hjälpa till med, särskilt om klassificeringarna inte förändrats i takt med forskningen.

Förutsättningarna för att söka svenska avhandlingar och artiklar är än så länge otillfredsställande. Det saknas ett enhetligt söksystem, vilket gör att man är tvungen att använda separata sökningar och sedan jämföra dem för att få en sammantagen bild, ändå ger inget av sökresultaten ett fullgott resultat. Sammanfattningsvis vill vi lyfta följande punkter som betydelsefulla för vårt metodiska arbete:

- Kunskap om forskningsfältet är en förutsättning för att kunna värdera de resultat man får fram via databaser, om de artiklar man fått som utgångspunkt är rimliga i termer av innehållsliga områden, om centrala forskare finns representerade, om sökningarna behöver kompletteras eller avgränsas. För kompletteringar är det nödvändigt att kunna avgöra vad som behöver kompletteras för att välja eller pröva ut andra sökstrategier som kan bidra med detta. Specifikt för den här översikten har det handlat om att vi inte fått fram de material vi redan på förhand kände till inom det svenska forskningsfältet. Vad gäller avgränsningar handlar det om att identifiera material som dyker upp i sökresultatet som inte har relevans för översiktens syfte.
- Kunskap om databaser/söktjänster: denna punkt är relaterad till föregående punkt på så sätt att när sökresultaten skiljer sig från det man förväntat sig är man tvungen att gå in och läsa om hur respektive databas/söktjänst är konstruerad. Den här typen av information hjälper en att förstå till exempel skillnader i sökresultat eller vilka alternativ som kan hjälpa en att precisera sökningen – men också om man får finna sig i att arbeta vidare manuellt.
- De svenska söktjänsterna är tills vidare otillfredsställande. Det förefaller som att det finns tre faktorer som bidrar till detta. För det första är söktjänster uppbyggda via LIBRIS baserade på klassifikationssystem istället för på den information som forskarna ger om sina publikationer. I något skede genomför administratörer uttolkningar av de abstrakt och/eller nyckelord31 respektive forskare angett för avhandling, licentiatuppsats eller artikel som anpassas till den uppsättning ämnesord som systemen erbjuder (ämnesord, ämneskategori). På vilka grunder transformeringen av den information forskarna själva uppger till det administrativa systemets sökord och klassificeringar görs framgår inte. Det saknas information om hur Skolportens söktjänst är uppbyggd. För det andra så har svenska lärosäten och forskningsinstitutioner tills vidare inte en gemensam söktjänst/databas för forskning. För det tredje är den information som finns i till exempel DiVA dels beroende av den information forskarna uppger om sina arbeten, dels beroende av att enskilda forskare registrerar sina publikationer. Angående det förstnämnda (abstrakt och nyckelord) har vi konstaterat att informationen ofta är knapphändig och ibland utelämnas central information helt. Vad gäller registreringen av publikationer så har vi erfarit att detta inte alltid görs. Principiellt verkar det som att lärosätenas och institutionernas hemsidor fortfarande är ett viktigt komplement till generella söktjänster, en extra kontroll vi inte haft tid med för denna översikt. Det finns således ett behov av att enas om och bygga ut en gemensam söktjänst för samtliga lärosäten och forskningsinstitutioner, men också av att enskilda forskare registrerar (alternativt får hjälp med att registrera) sina publikationer.
- För vår översikt gäller även att den tid vi hade till förfogade vad kort i relation till materialets omfång, vilket har viss betydelse för resultatet. Det hade varit optimalt att kunna pröva fler sökningar – till exempel i relation till var och en av de huvudkategorier som organiserar våra resultat, men även i relation till asiatiska länder vilkas forskning vi är betydligt mindre förtrogna med än västerländsk forskning. Sådana kompletteringar hade rimligen kunnat nyansera våra resultat något, eventuellt också kompletterat dem.

4.5 Rekommendationer

Utifrån denna syntetisering/diskussion/konklusion kring fenomenet formativ bedömning vill vi avsluta med några korta rekommendationer till skolforskningsinstitutet:

---

[31] För denna process använder LIBRIS dels en trädstruktur för en grundläggande ämnesklassificering i 25 bokstavskategorier (http://libris.kb.se/subjecttree.jsp). Dessutom används den internationella Web Dewey Search (DDK) som gör en ämnesklassificering baserad på tio numeriska ämneskategorier (000-900).

- Initiera och stöd forskning som syftar till fördjupad förståelse av det arbete med formativ bedömning som pågår i svenska klassrum, även sådan forskning som tar elevernas perspektiv
- Initiera och stöd forskning och utvecklingsarbete där fokus riktas mot analys och (om)formande av undervisning på basis av bedömningsinformation
- Initiera och stöd forskning som involverar olika inblandade aktörers roll i skolutveckling med avseende på bedömningsfrågor – vår genomgång visar tydligt att starka broar behöver byggas mellan forskning, lärarutbildning, beslutsfattare, huvudmannanivå och lokal skolnivå
- Stöd forskning som rör IKT-lösningar i relation till formativ bedömning, särskilt sådan som problematiserar konsekvenser för kunskapssyn och lärares professionalism när digitala verktyg ses som en egen aktör snarare än ett verktyg i händerna på mänskliga aktörer
- Kartlägg lärarutbildningarnas roll och läget på landets lärarutbildningsinstitutioner vad gäller den bedömningskunskap som finns och tillgängliggörs för studenter (kartläggning som grund för vilka insatser som kan komma att behövas)
- Bidra till att stärka kunskapen kring formativ bedömning bland annat genom problematisering av den förenklade/instrumentella uppfattningen av formativ bedömning som i viss utsträckning råder
- Granska och bevaka den forskning som i förenklad form sprids, så att sådan även innehåller problematiseringar kring de kontexter i vilka studier är gjorda och kring i vilken utsträckning generella slutsatser kan dras
- Granska och problematisera talet om effekter av formativ bedömning – begreppet innefattar oerhört disparata företeelser och det kan vara mer fruktbart att tala i termer av de olika aspekterna specifikt och vilken påverkan de kan tänkas ha på elevers lärande och välbefinnande under olika kontextuella villkor (exempelvis avseende elevers ålder, olika skolämnen)

REFERENSLISTA

Airasian, P. W. & Gregory, K. D. (1997) The Education Reform Act of 1998 (Review essay), *Assessment in Education*, *4*(2), 307–314.

Allal, L., & Mottier Lopez, L. (2005). Formative assessment of learning: A review of publications in French. I J. Looney (Red.), *Formative assessment: Improving learning in secondary classrooms* (241-264). Paris: OECD.

Andersen, T. E., Blair, B., Engelsen, K. S., Måseidvåg-Gamlen, S. T., Olsen, J. I., Sandal, A. K., m fl. (2010). *Kartlegging av vurdering som tema innenfor grunnutdanningen i UHnett Vest.* Research Report, Bergen: UH-nett, Vest.

Andrade, H. L. (2010). Students as the definitive source of formative assessment. I H. Andrade, & G. J. Cizek (Red.), *Handbook of Formative Assessment* (90-105). New York: Routledge.

Andrade, H., & Boulay, B. (2003). Gender and the role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*(1), 21-34.

Andrade, H. L. & Cizek G.J (Eds.), Handbook of Formative assessment New York: Routledge.

Asp-Onsjö, L. (2011). Dokumentation, styrning och kontroll i den svenska skolan. *Educare Vetenskapliga skrifter*, 2011:2, 39-56.

Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal, 25*(2), 238-259.

Atjonen, P. (2007). *Hyvä, paha arviointi* [Bedömning på gott och ont. På finska]. Helsinki: Tammi.

Ausubel, D. P. (1972). *Learning theory and classroom practice*. Toronto: The Ontario institute for studies in education, 1.

Azevedo, R., & Bernard, R. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

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

Berggren, J. (2013). *Learning from Giving Feedback: Insights from EFL Writing Classrooms in a Swedish Lower Secondary School*. Department of English, Stockholm University. (licentiatuppsats).

Berry, R. (2011a). Assessment trends in Hongkong: seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice, 18*(2), 199-211.

Berry, R. (2011b). Assessment Reforms Around the World. I A. Berry & B. Adamson (Red), *Assessment Reform in Education, Policy and Practice*. Dordrecht: Springer.

Biggs, J. (1999). Teaching for quality learning at University. Buckingham: SRHE and Open University Press.

Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment *for* learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation, 37*, 35-48.

Björklund Boistrup, L. (2010). *Assessment Discourses in Mathematics Classrooms: A Multimodal Social Semiotic Study*. Stockholm

Black, P. (2001). Formative assessment and curriculum consequences. Ingår i D. Scott (red.), *Curriculum and assessment* (7-23).

Black, P. (1998). Testing – friend or foe? Theory and practice of assessment and testing. London: Falmer.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it into practice*. Maidenhead: Open University Press.

Black, P., & Wiliam, D. (2012). Developing a Theory of Formative Assessment. I J. Gardner (Red.), *Assessment and Learning, 2nd edition*. London: Sage.

Black, P. & Wiliam, D. (2009). *Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Black, P. & Wiliam, D. (2006). Developing a theory of formative assessment. I J. Gardner (red.), *Assessment and learning. Theory, policy, and practice* (s. 81-100). London/Thousand Oaks, CA/New Dehli: Sage.

Black, P. & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16*(2), 249-261.

Black, P. & Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5(1), 1–74.

Bloom B.S. & Krathwohl, D.S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain.* New York: David McKay Co Inc.

Bloom, B. S.; Hastings, J. T. & Madaus, G. F. (1971). *Handbook on formative and summative evaluation*. New York: McGraw-Hill Inc.

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

Broadfoot, P., & Black, P. (2004). The first ten years of Assessment in Education. *Assessment in Education: Principles, Policy & Practice, 11*(1), 7-26.

Carlgren, I. (2002). Det nya betygssystemets tankefigurer och tänkbara användningar. I Skolverket (red.), *Att bedöma eller döma: Tio artiklar om bedömning och betygssättning* (s. 13-26). Stockholm: Liber Distribution.

Carlgren, I., Forsberg, E. & Lindberg, V. (2009). *Perspektiv på den svenska skolans kunskapsdiskussion*. SKIP-rapport nr 7. Stockholm, Stockholms universitets förlag.

CERI. (2005). *Formative Assessment: Improving learning in secondary classrooms*. Paris: OECD Publishing.

Chappuis, S. (2005). Is Formative Assessment Losing Its Meaning? *Education Week, 24*(44), p 38.

Chappuis, S., & Chappuis, J. (2008). The best value in formative assessment. *Educational Leadership, 65*(4): 14-19.

Ciampa, K., & Gallagher, T. L. (2013). Getting in Touch: Use of Mobile Devices in the Elementary Classroom. *Computers in the Schools: Interdisciplinary Journal of Practice, Theory, and Applied Research, 30*(4), 309-328.

Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review, 24*, 205–249.

Cook-Sather, A. (2002). Authorizing Students' Perspectives: Toward Trust, Dialogue, and Change in Education. *Educational Researcher, 31*(4), 3–14.

Cooper, B., & Cowie, B. (2010). Collaborative research for assessment for learning. *Teaching and Teacher Education, 26*, 979-986.

Cowie, B. & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education, 6* (1), 101-116.

Cowie, B. (2000). Formative assessment in science classrooms. Unpublished PhD Thesis, Hamilton: University of Waikato.

Crane, L., & Winterbottom, M. (2008). Plants and photosynthesis: peer assessment to help students learn. *Educational Research, 42*(4), 150-156.

Cronbach, L.J. (1963). Course Improvement through Evaluation. *Teachers College Record, Vol. 64*, s. 672-83.

Crooks, T., Kane, M. T. & Cohen, A. S. (1996) Threats to the valid use of assessment, *Assessment in Education*, 3(3), 265–286.

Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation, 37*, 71-77.

Crossouard, B. (2012). Absent presences: the recognition of social class and gender dimensions within peer assessment. *British Educational Research Journal, 38*(5), 731-748.

Daly, C., Pachler, N., More, Y., & Mellar, H. (2010). Exploring formative e-assessment: using case stories and design patterns. *Assessment & Evaluation in Higher Education, 35(* 5), 619–636.

Davies, D., & Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *The Journal of Educational Research, 104*(3), 202–215.

del Rosario Basterra, M., Trumbull, E. & Solano-FLores, G. (red.), (2011). *Cultural validity in assessment. Addressing linguistic and cultural diversity.* New York & London: Routledge.

Duschl, R. (2000). Making the nature of science explicit. I R. Millar, J. Leach & J. Osborne (red.), *Improving science education* (s. 187-196). Buckingham, UK: Open University Press.

Durán, R. (2011). Ensuring valid educational assessment for ELL students: Scores, score interpretation, and assessment uses. I M. del Rosario Basterra, E. Trumbull. & G. Solano-FLores (red.), *Cultural validity in assessment. Addressing linguistic and cultural diversity* (s. 115-143). New York & London: Routledge.

EACEA (Education, Audio-visual and Culture Executive Agency). (2009). *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. DOI 10.2797/18294

Elfström, I. (2005). Varför individuella utvecklingsplaner? En studie om ett nytt utvärderingsverktyg i förskolan. Individ, Omvärld och Lärande/Forskning 26. Lärarhögskolan i Stockholm (licentiatuppsats).

Ellis, P. D (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. New York: Cambridge University Press.

Elwood, J. (2006). Formative assessment: possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice, 13*(2), 215-232.

Engelsen, K.S. & Smith, K (20). Is "Excellent" good enough? *Education Inquiry, Vol. 1*, (4), 415–431

Engeström, Y. (1987). *Learning by expanding. An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.

Engeström, Y., & Miettinen, R. (1999). Introduction. I Y. Engeström, R. Miettinen & R. L. Punamäki (Red), *Perspectives on Activity Theory (1-6)*. Cambridge University Press.

Engeström, Y. (2001). Expansive Learning at Work: toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-156.

Eriksson, I., Ståhle, Y. & Lindberg, V. (2011). Kemikemi eller samhällskemi? Skilda betoningar i finlandssvenska och svenska kemiklassrum. I I. Eriksson (red.), *Innehållet i fokus – kemiundervisning i finlandssvenska klassrum* (s. 76-113). SKIP-rapport nr 8/2010. Stockholms universitetsförlag.

Evans, L. (2011) The 'shape' of teacher professionalism in England: Professional standards, performance management, professional development and the changes proposed in the 2010 white paper, *British Educational Research Journal, 37*(5), 851–870.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. London: Routledge Falmer.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287-322.

Firestone, W. A. (1998) A tale of two tests: tensions in assessment policy, *Assessment in Education*, *5*(2), 175–191.

Flaitz, J. (2011). Assessment for Learning: US Perspectives. I A. Berry & B. Adamson (Red), *Assessment Reform in Education, Policy and Practice*. Dordrecht: Springer.

Forsberg, E. & Lindberg, V. (2010). *Svensk forskning om bedömningar – en kartläggning*. Vetenskapsrådets rapportserie 2010:2. Stockholm: Vetenskapsrådet.

Forsberg, E. & Lundahl, C. (2006). Kunskapsbedömningar som styrmedia. *Utbildning & Demokrati, Vol. 15*, (3) 7-29.

Forsberg, E. & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education, Vol. 38* (5), 741-758.

Gagné, R. (1965) *The conditions of learning.* London: Holt, Rhinehart and Winston.

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice, 20*(2), 150-169.

Gardner, J. (2012). *Assessment and learning*. London: Sage.

Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2011). Engaging and Empowering teachers in Innovative Assessment Practice. I A. Berry & B. Adamson (Red), *Assessment Reform in Education, Policy and Practice*. Dordrecht: Springer.

Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*, 41–54

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010a). Improving the effectiveness of peer feedback for learning. Learning and Instruction, 20, 304-315.

Gielen, S., Tops, L., Dochy, F., Onghena, P., & Smeets, S. (2010b). A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British Educational Research Journal, 36*, 143-162.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education, Vol. 24*, 355-392

Gipps, C. V. (1994). *Beyond testing. Towards a thery of educational assessment*. London: The Falmer Press.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist, 18*, s. 519-521.

Goodrich, H. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment*. Opublicerad doktorsavhandling, Harvard University, Cambridge, MA.

Granath, G. (2008). *Milda makter! Utvecklingssamtal och loggböcker som disciplineringstekniker*. Göteborg Studies in Educational Sciences, No. 263.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. Psychological Bulletin, 124, 75–111.Gueudet, G., & Trouche, L. (2009). Towards new documentation systems for teachers? *Educational Studies in Mathematics, 71*(3), 199–218.

Hargreaves, E. (2011). Teachers' Feedback to Pupils: "Like so Many Bottles Thrown Out to Sea"?. I A. Berry & B. Adamson (Red), *Assessment Reform in Education, Policy and Practice*. Dordrecht: Springer.

Harlen, W. (2012). On the Relationship between Assessment for Formative and Summative Purposes. I J. Gardner (Red.), *Assessment and Learning, 2nd edition*. London: Sage.

Harlen, W. (2010). What is quality teacher assessment? I J. Gardner, W. Harlen, L. Hayward, G. Stobart, & M. Montgomery (Red), *Developing teacher assessment* (29–52). Berkshire: Open University Press.

Harlen, W. & Deakin-Crick, R. (2003) Testing and motivation for learning, *Assessment in Education*, 10(2), 169–208.

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101-111.

Hartell, E. (2014) Exploring the (un-)usefulness of formative assessment documents in primary technology education. *International Journal of Technology Education, Vol. 24* (2), 141-161.

Hartell, E. (2013). Looking for a glimpse in the eye.I I-B. Skogh & M.J. de Vries (red.), *Technology Education – Practicing Teachers Researching Teachers Practice* (s. 255-281). Sense Publishers

Hartell, E. (2012). *The Inefficient Loneliness. A Descriptive Study about the Complexity of Assessment for Learning in Primary Technology Education.* Stockholm: KTH (licentiatuppsats).

Hattie. J. (2009). *Visible Learning*. London: Routledge.

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

Hermansen, H., & Nerland, M. (2014). Reworking practice through an AfL project: an analysis of teachers' collaborative engagement with new assessment guidelines. British Educational Research Journal, 40(1), 187–206.

Hill, M. F. (1999). Assessment in self-managing schools: Primary teachers balancing learning and accountability demands in the 1990s. *New Zealand Journal of Educational Studies, 34*, 176–185.

Hill, M. F. (2000). *Remapping the assessment landscape: Primary teachers reconstructing assessment in self-managing schools*. Unpublished DPhil thesis. Hamilton, New Zealand: University of Waikato.

Hirsh, Å. (2015). IDPs at work. *Scandinavian Journal of Educational Research, Vol. 59***,** (1) 77-94. (i-first publicerad online 2013-10-07)

Hirsh, Å. (2014). The Individual Development Plan: supportive tool or mission impossible? Swedish teachers' experiences of dilemmas in IDP practice. *Education Inquiry Vol. 5*, (3) 405-427.

Hirsh, Å. (2013). *The individual development plan as tool and practice in Swedish compulsory school.* School of Education and Communication, Jönköping University (avhandling)

Hirsh, Å. (2012): The individual education plan: a gendered assessment practice? *Assessment in Education: Principles, Policy & Practice, Vol. 19* (4), 469-485.

Hirsh, Å. (2011). A tool for learning? An analysis of targets and strategies in Swedish Individual Education Plans. *Nordic Studies in Education, Vol. 31*, 14–30.

Hirst, P. H. (1974). *Knowledge and the curriculum: a collection of philosophical papers*. London: Routledge.

Hofvendahl, J. (2006a). *Riskabla samtal – en analys av potentiella faror i skolans kvarts- och utvecklingssamtal.* Linköpings universitet, Institutionen för Språk och Kultur & Arbetslivsinstitutet.

Hofvendahl, J. (2006b). "Noa har inga fel": om bristfokus i skolans utvecklingssamtal. *Utbildning & Demokrati, Vol. 15* (3) 61–81.

Hofvendahl, J. (2006c). Förpapprade samtal och talande papper. Ingår i E. Forsberg & E. Wallin (red.), *Skolans kontrollregim – ett kontraproduktivt system för styrning?* (s.). Stockholm: HLS Förlag.

Hultman, G. (2015). *Transformation, Interaktion eller Kunskapskonkurrens. Forskningsanvändning i praktiken*. Vetenskapsrådet (vr.se).

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice, 16*(3), 269-290.

Hunt, E., & Pellegrino, J. W. (2002). Issues, Examples, and Challenges in formative Assessment. New Directions for teaching and Learning, 89, 73-85.

Hurtig, M. (2007). *Jag vågar visa att jag kan. Om meningsskapande med digitala portföljer*. Doktorsavhandling 2007:45. Luleå tekniska universitet, Institutionen för pedagogik och lärande.

James, M., McCormick, R., Black, P., Carmichael, P., Drummond, M-J., Fox, A., MacBeath, J., Marshall, B., Pedder, D., Procter, R., Swaffield, S., Swann, J., and Wiliam, D. (2007). *Improving Learning How to Learn - classrooms, schools and networks*. Abingdon: Routledge.

James, M. (2012). Assessment in Harmony with our Understanding of Learning. I J. Gardner (Red.), *Assessment and Learning, 2nd edition*. London: Sage.

Jönsson, A. & Svingby, G. (2008). Underlag till ramverk för en provbank i grundskolan. Educare – Vetenskapliga skrifter, 2 (57-93).

Kapriva, R. & Sexton, U. (2011). Using appropriate assessment processes: how to get accurate information about the academic knowledge and skills of English language learners. I M. del Rosario Basterra, E. Trumbull. & G. Solano-FLores (red.), *Cultural validity in assessment. Addressing linguistic and cultural diversity* (s. 95-114). New York & London: Routledge.

Karlsson, J. (2008). *A Novel Approach to Teaching Emotional Expression in Music Performance*. Uppsala universitet, Institutionen för psykologi.

Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and Call for Research. *Educational Measurement: Issues and Practice, 30*(4), 28-37.

Kirton, A., Hallam, S., Peffers, J., Robertson, P., & Stobart, G. (2007). Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools. *British Educational Research Journal, 33*(4), 605-627.

Kjällander, S. (2011). *Designs for Learning in an Extended Digital Environment: Case Studies of Social Interaction in the Social Science Classroom*. Doctoral thesis in didactic Science, Stockholm university.

Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation, 37*, 78-83.

Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16*(3), 263-268.

Klenowski, V. & Wyatt-Smith, C. (2013). *Assessment for Education Standards, Judgement and Moderation*. London: Sage

Kommunförbundet i Skåne. (2014). *Bedömning för lärande – en vägledning utifrån aktuell forskning.* Forskning I korthet 2014:2. Hämtad 2014-10-27 från http://www.kfsk.se/download/18.249dc2c0148adc97571ef510/1417765641929/Forskning+i+korthet+2014-2+slutversion.pdf

Korp, H. (2011). *Kunskapsbedömning – vad, hur och varför?* Stockholm: Skolverket. Hämtad 2014-09-05. http://www.skolverket.se/om-skolverket/publikationer/visa-enskild-publikation?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2FRecord%3Fk%3D2666

Kousholt, K. (2009). *Evalueret: Deltagelse i folkeskolens evalueringspraksis*. Institut for Læring, Danmarks Pædagogiske Universitetsskole, Aarhus Universitet (opublicerad doktorsavhandling).

Krantz, J. (2009). *Styrning och mening – anspråk på professionellt handlande i lärarutbildning och skola*. Växjö universitet, Pedagogik, No. 181. (Doktorsavhandling)

Kullberg, A. (2010). *What is taught and what is learned: Professional insights gained and shared by teachers of mathematics*. Gothenburg Studies in Educational Science 293 (Doktorsavhandling).

Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*, 144–164.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? A case of lesson study. *Educational Researcher, 35*(3), 3–14.

Liedman, S-E. (2002). *Ett oändligt äventyr: om människans kunskaper*. Stockholm: Bonnier.

Leitch, R., Gardner, J., Mitchell, S., Lundy, L., Odena, O., Galanouli, D., & Clough, P. (2007) Consulting pupils in Assessment for Learning classrooms: the twists and turns of working with students as co-researchers, *Educational Action Research, 15*(3), 459-478.

Lindberg, V. (2009). Kunskapsuppfattningar och inter/nationella redskap för bedömning. Ingår i I. Carlgren, E. Forsberg & V. Lindberg, *Perspektiv på den svenska skolans kunskapsdiskussion* (s. 78–119.) SKIP-rapport nr 7. Stockholms universitetsförlag.

Lindberg, V. & Löfgren, R. (2010). Vilket kemikunnande efterfrågas och görs tillgängligt för eleverna? – Frågor, svar och feedback i kemiklassrummet. Ingår i I. Eriksson (red.), *Innehållet i fokus – kemiundervisning i finlandssvenska klassrum* (s. 145-173). SKIP-rapport nr 8/2010. Stockholms universitetsförlag.

Lundahl, C. (2009). *Varför nationella prov? Framväxt, dilemman, möjligheter*. Lund: Studentlitteratur.

Lundahl, C. (2006). *Viljan att veta vad andra vet: kunskapsbedömning i tidigmodern, modern och senmodern skola*. Stockholm: Arbetslivsinstitutet & Uppsala universitet (avhandling).

Löfgren, R. & Lindberg, V. (2011). Bedömningshandlingar i två klassrum. Likartat kemiinnehåll men skilda inramningar. Ingår i I. Eriksson (red.). *Kemiundervisning, text och textbruk i finlandssvenska och svenska skolor – en komparativ tvärvetenskaplig studie* (s. 238-274). SKIP-rapport 9/2011. Stockholms universitets förlag.

Mansell, W., & James, M. (2009). *Assessment in schools—Fit for purpose?* Assessment Reform Group. London: TLRP, Institute of Education, University of London.

McDonald, B. & Boud, D. (2003) The impact of self assessment on achievement: the effects of self assessment training on performance in external examinations, *Assessment in Education, 10*(2), 209–220.

Miyakawa, T., & Winsløw, C. (2013). Developing mathematics teacher knowledge: the paradidactic infrastructure of ‘‘open lesson’’ in Japan. *Journal of Math Teacher Education, 16*, 185–209.

Mok, J. (2010). A case study of students' perceptions of peer assessment in Hong Kong. *ELT Journal, 65*(3), 230-239.

Morris, P., & Adamson, B. (2010). *Curriculum, Schooling and Society in Hongkong*. Hongkong: Hongkong University Press.

Morrison, K. & Tang Fun Hei, J. (2002) Testing to destruction: a problem in a small state, *Assessment in Education: Principles, Policy & Practice, 9*(3), 289–317.

Moss, P. A., Pullin,D. Gee, J.P Haertel, E. & Young, L. (red.), (2008). *Assessment Equity and Opporutnity to Learn*. New York: Cambridge University Press.

Murphy, P. (2008). Gender and subject cultures in practice. I P. Murphy & K. Hall (red.) *Learning and practice: Agency and identities* (s. 161-172). London/Thousand Oaks, CA/New Delhi: Sage Publications.

Murphy, P. & Ivinson, G. (2005). Gender, assessment and students' literacy learning: implications for formative assessment. *Teacher development, Vol. 9* (2), 185-200.

Mårell-Olsson, E. (2012). *Att göra lärandet synligt? Individuella utvecklingsplaner och digital dokumentation*. Doktorsavhandlingar i pedagogiskt arbete, 48. Institutionen för tillämpad utbildningsvetenskap, Umeå universitet.

Mäensivu, K. (1999). *Opettaja määrittelijänä oppilas määriteltävänä - sanallisen oppilaan arvioinnin sisällön analyysi* [Läraren – den som bedömer, eleven – den som blir bedömd]. Jyväskylä studies in education, psychology and social research 157, Jyväskylän yliopisto.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media-learning. I H. Niegemann, R. Brunken, & D. Leutner (Red.), *Instructional design for multimedia learning* (181-195). Munster: Waxman.

Nilholm, C. (2013). Det är dags att kritisk granska John Hattie. Pedagogiska Magasinet 2013-11-21. Hämtad 2014-09-14 på http://www.lararnasnyheter.se/pedagogiska-magasinet/2013/11/21/ar-dags-kritiskt-granska-john-hattie

Nyberg, E. (2009). *Om livets kontinuitet. Undervisning och lärande om växters och djurs livscykler – en fallstudie i årskurs 5*. Göteborgs universitet, Institutionen för pedagogik och didaktik.

Pachler, N., H. Mellar, C. Daly, Y. Mor, D. Wiliam, and D. Laurillard. 2009. *Scoping a vision for formative e-assessment: A project report for JISC.* London: Institute of Education, WLE Centre.

Panadero, E., & Jönsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129-144.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning. Towrds a wider conceptualisation of the field. *Assessment in Education: Principles, Policy & Practice, 5*(1), 85-102.

Peterson, E. R., & Irving, S. E. (2007). Secondary school students' conceptions of assessment and feedback. Learning and Instruction, 18, 238-250.

Pintrich, P (2000). The role of goal-orientation in self-regulated learning. I M. Boekaerts, P. Pintrich, & M. Zeidner (Red), *Handbook of self-regulation,* (454-502). San Diego CA: Academic.

Popham, W. J. (1978). *Criterion referenced measurement*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Preece, P. F. W. & Skinner, M. C. (1999) The national assessment in science at Key Stage 3 in England and Wales and its impact on teaching and learning, *Assessment in Education*, 6(1), 11–26.

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford review of Education, 34*(1), 1-20.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science, 28*(1), 4-13.

Richardson, V. (red.) (2001). *Handbook of Research on Teaching.* 4. ed. Washington, D.C.: American Educational Research Association.

Riksrevisionen (RiR 2011:03). *Lika betyg, lika kunskap? En uppföljning av statens styrning mot en likvärdig betygssättning i grundskolan*. Dnr: 31-2010-1551. Hämtad 2014-02-15 på http://www.riksrevisionen.se/PageFiles/13875/Anpassad_11_23%20Lika%20betyg%20lika%20kunskap.pdf

Roos, B. & Hamilton, D. (2005). Formative assessment: A cybernetic viewpoint. *Assessment in Education: Principles, Policy & Practice, Vol. 12*, (1) 7-20.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry. *Journal of Research in Science Teaching, 44* (1), 57-84.

Räty, H., Snellman, L., Leinonen, T. & Maksimainen, A. (2000). Miten lasta arvioida? Vanhempien käsityksiä lapsenas kouluarvioinnista [Hur ska barn bedömas? Föräldrars uppfattningar om skolans bedömning av deras barn – på finska]. *Kasvatus 31*, (4), 332-338.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment. 11*, 1-31.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally.

Sellergren, U. (1975). *Läraren, planeringen och undervisningen*. Malmö: Liber Läromedel.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4-13.

SFS 1993/100, Bilaga 2. Högskoleförordningen. Utbildningsdepartementet: Stockholm.

Shen, C. (2002) Revisiting the relationship between students' achievement and their self-perceptions: a cross-national analysis based on TIMSS 1999 data, *Assessment in Education*, 9(2), 161–184.

Shepard, L. (2001). The role of assessment in a learning culture. Ingår i Richardson, Virginia (red.) (2001). *Handbook of Research on Teaching.* 4. ed. Washington, D.C.: American Educational Research Association.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review, 57* (1), 1-23.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, 4-14.

Skolverket. (2011). *Kunskapsbedömning i skolan. Praxis, begrepp, problem och möjligheter*. Stockholm: Fritzes.

Skovsmose, O. (2005). *Travelling through education: Uncertainty, mathematics, responsibility*. Rotterdam, The Netherlands: Sense Publisher.

Smith, K. (2009). Samspillet mellom vurdering og motivasjon. In S. Dobson, A. B. Eggen, & K. Smith (Eds.), *Vurdering, prinsipper og praksis-nye perspektiver på elev og læringsvurdering* (23–38). Oslo: Gyldendal Akademisk.

Smith, K. (2011). Professional development of teachers—A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation, 37*, 55-61.

Stiggins, R. J. (2002). Assessment Crisis: The Absence of Assessment FOR Learning. *Phi Delta Kappan, 83*(10), 758-765.

Stobart, G. & Stoll, L. (2005) The Key Stage 3 Strategy: What kind of reform is this? *Cambridge Journal of Education, 35*, 2, 225-238.

Strandberg, M. (2013). Homework – is there a connection with classroom assessment? A review from Sweden. *Educational Research, Vol. 55* (4), 325-346.

Strandberg, M. & Lindberg, V. (2012). Feedback in a multi-ethnic classroom discussion – a case study. *Intercultural Education, Vol. 23* (2), 75-88.

Strijbos, J-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. Learning and Instruction, 20, 265-269.

Stufflebeam, A.L. & Shinkfield, D.J. (2014). *Evaluation Theory, Models, and Applications*. 2. Uppl. San Fransisco, CA: Jossey Bass.

Sundström, U. (2011). Hattie: Är du expert, erfaren eller novis? Lärarnas Tidning 2011-10-25. Hämtad 2014-09-14 på http://www.lararnasnyheter.se/lararnas-tidning/2011/10/25/hattie-ar-du-expert-erfaren-eller-novis

Timperley, H. (2013). *Det professionella lärandets inneboende kraft*. Lund: Studentlitteratur.

Topping, K. J. (2009). Peer Assessment. *Theory Into Practice, 48*, 20–27.

Topping, K. J. (2010). Peers as a Source of Formative Assessment. I H. Andrade, & G. J Cizek (Red), *Handbook of Formative Assessment*, (61-74). New York: Routledge.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice, 14*(3), 281-294.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: a typology. *British Educational Research Journal, 22* (4), 389-404.

Utbildningsdepartementet (2008) *En individuell utvecklingsplan med skriftliga omdömen*. Departementspromemoria publicerad 2008-01-15. http://www.regeringen.se/sb/d/108/a/95933 (hämtad 2008-01-16)

Vallberg Roth, A-C. (2009). Styrning genom bedömning av barn. *Educare – Vetenskapliga skrifter,* nr 2-3, 195-216.

Vallberg Roth, A-C. & Månsson, A. (2008a). Individuella utvecklingsplaner för skolår 3, 6 och 8 i fyra sydsvenska kommuner. Educare.

Vallberg Roth, A-C. & Månsson, A. (2008b). Individuella utvecklingsplaner som uttryck för reglerad barndom: Likriktning med variation. *Pedagogisk forskning i Sverige, 13*(2), 81-102.

Vallberg Roth, A-C. & Månsson, A. (2006). Individuella utvecklingsplaner som fenomen i tiden, samhället och skolan. *Utbildning & Demokrati, Vol. 15*, (3) 31-60.

Wallin, E. (2005). The Rise and Fall of Swedish Educational Technology 1960–1980. *Scandinavian Journal of Educational Research, Vol. 49*, 5 (437-460).

Wang, T. H. (2007). What strategies are effective for formative assessment in an e-learning environment? *Journal of Computer Assisted Learning, 23*, 171-186.

Wedman, I. (2003). Från skallmätning till VG. Om skolans betyg. Ingår i S. Selander (red.), *Kobran, nallen och majjen. Tradition och förnyelse i svensk skola och skolforskning* (s. 307-316). Forskning i fokus nr 12. Stockholm: Myndigheten för skolutveckling.

Wermke, W. (2013). *Development and autonomy. Conceptualizing teachers' continuing professional development in different national contexts*. Doktorsavhandlingar från Institutionen för pedagogik och didaktik 16. Stockholms universitet.

Wernberg, A. (2009). *Lärandets objekt: vad elever förväntas lära sig, vad görs möjligt för dem att lära och vad de faktiskt lär sig under lektionerna*. Institutionen för naturvetenskapernas och matematikens didaktik, Umeå universitet (Doktorsavhandling).

Westlund, B. (2013). *Att bedöma elevers läsförståelse. En jämförelse mellan svenska och kanadensiska bedömningsdiskurser i grundskolans år.* (Doktorsavhandling.). Stockholms universitet. Stockholm: Natur & Kultur.

Wettergren, S. (2013). *SAM-tal om bedömning och matematikkunnighet: en studie av lärares tankestilar*. Licentiatavhandling Stockholms universitet, Institutionen för pedagogik och didaktik.

Whitty, G. (2006). Education(al) research and education policy making: is conflict inevitable? *British Educational research Journal, 32*(2), 159-176.

Wiliam, D. (2011). *Embedded formative assessment.* Bloomington, IN: Solution Tree Press.

Wiliam, D. (2009). *Assessment for learning – why, what and how?* University of London: Institute of Education.

Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. I J. H. McMillan (Red.), *Formative classroom assessment* (s. 29-42). New York: Teacher College Press.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? I C. A. Dwyer (Red.), *The future of assessment: shaping teaching and learning*, (s. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.

Williams, J. A. (2010). 'You know what you've done right and what you've done wrong and what you need to improve on': New Zealand students' perspectives on feedback. *Assessment in Education: Principles, Policy & Practice, 17*(3), 301-314.

Vingsle, C. (2014). *Formative assessment: Teacher knowledge and skills to make it happen*. Licentiand thesis, Educational works no 15. Naturvetenskapernas och matematikens didaktik. Umeå universitet.

Winsløw, C. (2011). A comparative perspective on teacher collaboration: The cases of lesson study in Japan and of multidisciplinary teaching in Denmark. I G. Gueudet, B. Pepin, & L. Trouche (Red.), *Mathematics curriculum material and teacher documentation: From textbooks to shared living resources* (291–304). New York: Springer.

Yin, Y., Tomitia, M. K., & Shavelstone, R. J. (2014). Using Formal Embedded Formative Assessments Aligned with a Short-Term Learning Progression to Promote Conceptual Change and Achievement in Science. *International Journal of Science Education, 36*(4), 531-552.

Svenska tidskrifter - hemsidor

Educare Vetenskapliga tidskrifter: http://www.mah.se/educare

Pedagogisk Forskning i Sverige: http://pedagogiskforskning.se/

Utbildning & Demokrati. Tidskrift för didaktik och utbildningspolitik: http://www.utbildningochdemokrati.se/ud/

Databaser (för sökning av svenska avhandlingar)

LIBRIS: www.libris.kb.se

Skolporten: www.skolporten.se

SwePub: http://swepub.kb.se/form_extended.jsp

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

Vetenskapsrådet har en ledande roll för att utveckla svensk forskning av högsta vetenskapliga kvalitet och bidrar därmed till samhällets utveckling. Utöver finansiering av forskning är myndigheten rådgivare till regeringen i forskningsrelaterade frågor och deltar aktivt i debatten för att skapa förståelse för den långsiktiga nyttan av forskningen.