

Sambedömning för ökad likvärdighet?

Pia Thornberg & Anders Jönsson

Teachers' assessment and grading of student achievement is often charged for not being reliable. Various suggestions of how to reach higher levels of reliability are advocated in different contexts, and one suggestion is "social moderation". Social moderation involves teachers working together when assessing and grading student performance. Although the claim is often made that moderation practices lead to higher levels of reliability in teachers' assessment and grading, such claims often lack scientific support. In this study, a literature review has been carried out in order to gain knowledge about the potential effects that moderation may have and whether these effects may in fact lead to more reliable assessment and grading. The results indicate that moderation practices do have an effect, which may ultimately lead to higher levels of reliability if these practices involve several levels in the assessment process. Moderation is also a process that can be affected by several factors, such as the organization of the moderation meetings and the selection of student performance.

Keywords: grading, moderation, reliability, teacher assessment

Pia Thornberg, Lecturer in Mathematics Education, Höskolan Kristianstad.

pia.thornberg@hkr.se

Anders Jönsson, Associate Professor in Science Education Research, Höskolan Kristianstad.

anders.jonsson@hkr.se

Inledning

Skolverket och Skolinspektionen har i flera granskningar påvisat att bedömningen och betygssättningen i den svenska skolan brister i likvärdighet (ex Skolinspektionen, 2012, 2013; Skolverket, 2012b, 2012c, 2012d). Att lärares bedömning och betygssättning uppvisar sådana brister är problematiskt bland annat eftersom elevers betyg används som urvalsinstrument och som konkurrensmedel när eleverna senare söker anställning eller till högre utbildningar. Bristande likvärdighet vid bedömning och betygssättning kan därmed få betydande konsekvenser för den enskilde eleven.

Olika förslag att åstadkomma en ökad likvärdighet vid bedömning och betygssättning förs fram i olika sammanhang. Skolverket (2004) föreslår i en handlingsplan att de nationella provens stödjande roll vid betygssättning ska förbättras, att lärar- och rektorsutbildningarna får en stärkt roll samt att stödmaterial som förtydligar hanteringen av styrdokumenterna ska tillhandahållas. Liknande förslag lyfts fram av Gustavsson, Måhl och Sundblad (2012) vilka menar att de som kan vidta åtgärder för att bättre garantera att betygssättningen blir ”rimligt likvärdig”, det vill säga att ”elever med samma betyg har samma slags kunskaper/förmågor på ungefär samma kvalitetsnivåer” (sid. 97) är lärarutbildningen, Skolverket samt rektorer och huvudmän.

På senare tid har sambedömning, det vill säga lärare som tillsammans genomför eller diskuterar bedömning, förts fram som en möjlighet att skapa förutsättningar för en mer likvärdig bedömning och betygssättning. Skolverket (2009a) skriver till exempel i ett missiv till regeringskansliet att ”ökade krav på sambedömning skulle leda till ökad bedömaröverensstämmelse bland lärare och vara ett effektivt sätt att höja kvaliteten på lärares bedömning av nationella prov” (sid. 17). Sambedömning förs även fram i andra rapporter som ett sätt att nå en ökad samsyn och samstämmighet bland lärare (se till exempel Erickson, 2009; Skolinspektionen, 2013; Skolverket, 2009b). Ett exempel är en analys utifrån lärarenkäter och omrättningar av de nationella proven som Skolinspektionen (2012) genomfört som visar att högstadie- och gymnasielärare som ensambedömer elevers prov sätter högre provbetyg relativt externa bedömare jämfört med lärare som sambedömer proven. Den analys som är gjord visar även att flertalet av de systematiska skillnader som återfinns i lärarnas

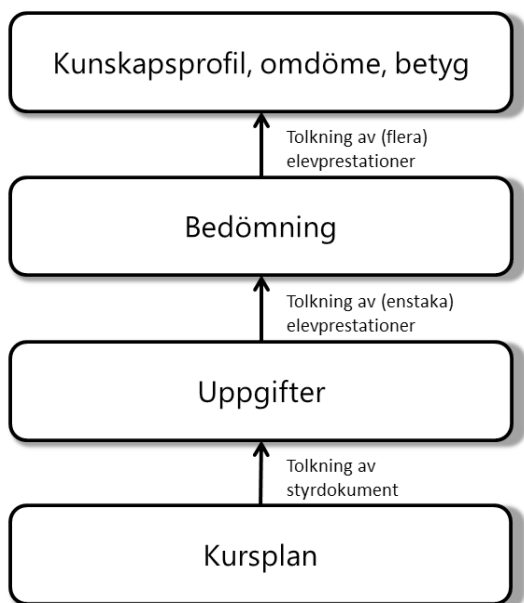
bedömningar är mindre uttalade när proven sambedömts än när de bedömts av enskilda lärare. Även om sambedömning lyfts fram som ett sätt att öka likvärdigheten vid bedömning och betygssättning är detta ofta ett antagande som inte är särskilt väl vetenskapligt förankrat och slutsatser och rekommendationer grundas ofta på intervjuer med lärare om deras upplevelser av sambedömning. Det finns således ett behov av mer kunskap inom området och syftet med denna studie är därför att undersöka och förstå hur sambedömning kan bidra till en mer samstämmig och likvärdig bedömning och betygssättning.

Likvärdighet

Likvärdig bedömning och betygssättning innebär i grunden att samma elevprestation förväntas ge samma bedömning och betyg oberoende av vem som bedömer eller hur prestationen bedöms, samt att det sker i linje med gällande styrdokument. I dagens mål- och kriterierelaterade bedömningssystem jämförs elevers prestationer med på förhand uppställda mål och kriterier, det vill säga kunskapskraven i nuvarande kurs- och ämnesplaner (Skolverket, 2011a, 2011b). De ämnesspecifika långsiktiga mål som formulerats i kurs- och ämnesplaner anger vad undervisningen ska leda fram till och beskriver elevers kunskaper så som de ska komma till uttryck i handling. Kunskapskraven, i förhållande till vilka elevers prestationer ska bedömas och betygssättas, utgår i sin tur ifrån de långsiktiga målen och beskriver hur dessa kan uttryckas på olika kvalitativa nivåer. Det blir därför centralt att undervisning och bedömning verkligen bygger på att eleverna får visa sitt kunnande så som det beskrivs i kurs- och ämnesplaner och att resultaten tolkas och värderas i enlighet med dessa.

En likvärdig bedömning av elevers prestationer ställer därför stora krav på bedömaren eftersom det krävs tolkning i flera olika led. Allal (2013) beskriver professionell bedömning som en process vilken omfattar tre olika tolkningsled, se Figur 1. I ett första led ska kurs- och ämnesplaner tolkas och omsättas i undervisning och bedömningsuppgifter. Elevernas prestationer ska sedan tolkas och värderas utifrån bedömningsanvisningar och/eller kunskapskrav och slutligen ska elevers olika prestationer sammanvägas till ett betyg eller till ett kunskapsomdöme. Utifrån en sådan modell räcker det inte att

beskriva bedömning och betygssättning som en process som startar i och med att elevers prestationer ska bedömas utan bedömningsprocessen behöver ha en tydlig linje som förbinder ett ämnes mål och syfte med undervisning, kunskapskrav och bedömning, så kallad "alignment" (Biggs & Tang, 2007).



Figur 1. Modellen beskriver hur bedömningsprocessen omfattar flera olika tolkningsled. (1) Tolkning av styrdokument för planering av undervisning och för konstruktion av bedömningsuppgifter, (2) tolkning och värdering av enskilda elevprestationer samt (3) sammanvägning av flera elevprestationer till betyg eller kunskapsomdöme. För en likvärdig bedömning krävs samsyn mellan lärare i samtliga dessa led. Modellen är anpassad med utgångspunkt från Allal (2013).

För en likvärdighet i det första tolkningsledet krävs att lärare gör en gemensam tolkning av kurs- och ämnesplaner för sedan kunna planera undervisning och konstruera bedömningsuppgifter i linje med dessa. Harlen (2004) menar att hur konsekventa lärare är vid bedömningen

av elevprestationer hör samman med i vilken utsträckning de har förstått styrdokumentet och i vilken utsträckning detta sedan omsätts i undervisning. Det innebär även att lärare måste göra tolkningar av vad som är utmärkande för prestationer på olika nivåer och vad som skiljer prestationer på olika nivåer ifrån varandra.

I det andra tolkningsledet handlar det om att tolka och värdera elevers prestationer på bedömningsuppgifter. Sadler (1998) beskriver detta som en process i tre steg vilken innebär att lärare (1) riktar uppmärksamhet mot elevprestationen (2) värderar denna mot till exempel ett ramverk och (3) ger ett explicit svar, som till exempel att ge prestationen en poäng, ett omdöme eller ett betyg. Att ha en samsyn kring bedömningen omfattar de två första stegen i Sadlers beskrivning, vilket kan innebära att lärarna har en gemensam uppfattning om vad det är som ska bedömas och vad som kännetecknar olika kvaliteter i elevernas visade kunskaper. Att lärare har en samsyn behöver däremot inte nödvändigtvis innebära att de sedan avger samma omdöme i steg tre, det vill säga att de har en samstämmighet i sin bedömning. Sadler (1998) ger även exempel på faktorer som påverkar lärare vid bedömning. Sådana faktorer kan vara egna ämneskunskaper, kunskap om de kriterier och kravnivåer som används vid bedömningen, kunskap om eleven som bedöms eller erfarenhet av att ha bedömt liknande uppgifter tidigare. Det finns därför goda skäl att anta att det kan finnas en stor variation i hur olika lärare tolkar och värderar elevers prestationer.

I det tredje tolkningsledet ska olika elevprestationer sammanvägas till ett betyg eller till ett kunskapsomdöme utifrån kurs- eller ämnesplanernas kunskapskrav. För att betygssättningen ska vara likvärdig krävs att kunskapskraven tolkas på samma sätt av olika lärare. I kunskapskraven används ett antal värdeord som i viss utsträckning är gemensamma och återfinns i flera ämnen. Skolverket (2012a) menar att värdeorden får ”sin betydelse i de sammanhang de används” och att det därför inte är möjligt att göra generella definitioner av dessa. Ett värdeords innebörd kan dessutom skilja mellan olika situationer och mellan olika uppgifter. Allal (2013) har studerat lärares betygssättningspraktik samt de överväganden som görs vid betygssättning och hon beskriver att denna praktik omfattar två huvudsakliga operationer: (1) insamlande av information utifrån

varierade källor och (2) hur denna information tolkas och sammanvägs till ett betyg. Hur det samlade underlaget används vid sammanvägning till ett betyg beskrivs av Allal som en tolkningsprocess som kan se olika ut för olika elever.

Sammantaget kan man konstatera att de tre tolkningsleden kan utgöra en osäkerhet i bedömningen och betygssättningen, vilket kan beskrivas i termer av bristande likvärdighet. För en ökad likvärdighet krävs därför en samsyn i samtliga tolkningsled och att lärare värderar liknande prestationer på ett likvärdigt sätt.

I Sverige har lärare stort förtroende när det kommer till bedömning och betygssättning. Ofta är det samma lärare som både bedömer, rättar nationella prov och sätter betyg. Fördelarna med ett sådant system där summativa bedömningar utförs av samma lärare som också undervisar eleverna är bland annat att lärare har möjlighet att bedöma utifrån ett bredare underlag och utifrån fler lärandemål än när bedömning sker utifrån enstaka tester utförda vid särskilda tillfällen (Harlen, 2005). Lärares bedömningar anses därför ha hög validitet (Crooks, 1988). Däremot anses de ofta ha lägre reliabilitet, främst eftersom lärare har en förförståelse om enskilda elever samt om den undervisning eleverna fått. Ur ett internationellt perspektiv är det förhållandevis ovanligt att summativa bedömningar med stor betydelse för elever utförs av den undervisande läraren. Eftersom lärare i det svenska skolsystemet har ett stort inflytande över både undervisning, bedömning och betygssättning, trots de brister det kan medföra ur ett likvärdighetsperspektiv, blir det angeläget att finna metoder för att stärka likvärdigheten. Detta görs på olika sätt genom bland annat tillhandahållandet av nationella prov och bedömningsstöd av olika slag.

Sambedömning

En strategi som ofta lyfts fram i forskningslitteratur och i andra sammanhang som ett sätt att få valida och reliabla bedömningar är sambedömning. Sambedömning har till exempel föreslagits av Skolverket (2009a) att ingå som ett sätt att öka likvärdigheten i lärarnas bedömning och betygssättning, tillsammans med nationella prov och övrigt bedömningsstöd. Sambedömning används även i länder som inte har nationella prov då det därmed finns behov av andra sätt att stödja

en likvärdig bedömning. Ett känt exempel är i delstaten Queensland, Australien, där man har mångårig erfarenhet av sambedömning på olika nivåer, både lokalt, regionalt och på delstatsnivå (Klenowski, 2012).

En vetenskaplig definition av sambedömning är inte tydligt etablerad. I forskningsstudier är det vanligt att man undersöker ”inter-rater reliability” där olika bedömare, oberoende av varandra, bedömer samma elevprestationer för att se hur dessa överensstämmer med varandra. Det är även vanligt att bedömare i sådana studier fått speciell träning för uppdraget. Ett annat begrepp är ”co-assessment” som innebär att lärare och elever eller studenter samverkar kring bedömning. Sambedömning, så som begreppet används i denna studie, har en annan innebörd och fokus ligger istället på hur lärare samverkar med varandra kring frågor om bedömning.

Ett begrepp som återkommer i forskningslitteraturen är ”moderation”. ”Moderation” kan innebära en extern inspektion där särskilt utvalda inspektörer kontrollerar hur bedömningen genomförs (Harlen, 1994). En annan form kan vara att lärares bedömning jämförs med något referenstest eller kalibreras mot något annat bedömningsinstrument (Gipps, 1994) genom så kallad ”statistical moderation”. Det begrepp som bäst fångar sambedömning i detta sammanhang är den beskrivning av ”moderation” som benämns ”social moderation” eller ”consensus moderation” (Gipps, 1994; Linn, 1993) och innebär att olika bedömare försöker nå konsensus i hur bedömning och betygssättning ska genomföras (Gipps, 1994). Begreppet ”moderation” kommer av ”att reglera” eller ”att plana ut” (Sadler, 2013). En sådan reglering kan åstadkommas genom att medelvärden av flera olika bedömares bedömningar beräknas, men då kommer man inte åt de underliggande orsakerna till olikheterna. Syftet med ”social moderation” är att få högre reliabilitet och validitet i lärares bedömningar så att de både stämmer överens med intentioner i styrdokumentet och att de är jämförbara på olika nivåer (Wilson, 2004). Eftersom sambedömning ger förutsättningar för diskussioner, som kan leda till en ökad samsyn kring styrdokument, samtidigt som det ger möjligheter för lärare att kalibrera sin bedömning gentemot andra lärares bedömningar, förs sambedömning ofta fram som en väg för att nå ökad samstämmighet vid bedömning.

Begreppet ”moderation” används i vid bemärkelse i forskningslitteraturen och kan innebära att lärare träffas vid särskilda möten, antingen inom den egna skolan eller mellan olika skolor och har med sig exempel på elevarbeten på olika nivåer. ”Moderation” kan även genomföras ”on-line” vilket gör att lärare från geografiskt skilda skolor kan komma i kontakt med varandra (se till exempel Adie, 2010). ”Moderation” kan utföras före, under och efter själva bedömningstillfället och är även genomförbart som en fortlöpande process (Maxwell, 2002). Bedömningen kan ske med eller utan explicita bedömningsanvisningar eller direkt utifrån kriterier och kravnivåer. Vid sådana möten ges lärare tillfälle att diskutera olika exempel på elevarbeten och gå igenom hur de motsvarar specificerade kravnivåer. Genom sådan diskussion och klassificering ges förutsättningar för lärare att kalibrera sin bedömning mot andra vilket kan leda till en högre reliabilitet (Klenowski & Wyatt-Smith, 2010).

Klenowski och Adie (2009) beskriver tre olika typer av ”social moderation”:

- ”The calibration model”. Denna modell innebär att lärare först bedömer ett antal elevarbeten från olika klasser eller skolor individuellt. Bedömningen diskuteras sedan gemensamt i syfte att nå konsensus och gemensam förståelse för kravnivåer. Därefter bedömer varje lärare sina egna elevers arbeten.
- ”The conferencing model”. Denna modell innebär att lärare först bedömer sina egna elevers prestationer. Därefter diskuteras ett representativt urval gemensamt med målet att nå konsensus och gemensam förståelse. Eventuell görs sedan en omvärdering av den ursprungliga bedömningen.
- ”The expert model”. Modellen innebär att lärare bedömer elevarbeten som sedan lämnas vidare till en expert. Lärarna får sedan tillbaka feedback på sin bedömning kring hur konsekvent de använt sig av angivna kravnivåer.

Begreppet sambedömning så som det används och definieras i den här beskrivna studien avser lärare som tillsammans genomför eller diskuterar bedömning och/eller betygssättning och kan sägas omfatta de två första beskrivningarna av ”social moderation”. I båda dessa modeller är målet att nå konsensus kring bedömningen. I definitionen av sambedömning som avses i denna studie behöver målet nödvändigtvis inte vara konsensus utan kan även omfatta att synliggöra skilda synsätt vid bedömning och betygssättning av elevers prestationer.

Studiens upplägg och genomförande

Då likvärdighet vid bedömning och betygssättning är en process som kan påverkas i flera led blir det viktigt att förstå var i denna process som sambedömning kan ge effekter och om det finns andra faktorer som kan påverka, så kallade modererande faktorer. För att kunna diskutera på vilket sätt sambedömning kan bidra till en mer samstämmig och likvärdig bedömning och betygssättning blir de frågor som denna studie söker svar på därför:

- Vilka effekter kan sambedömning ge?
- Vilka faktorer kan påverka sambedömningsprocessen?

För att få en översikt över det aktuella kunskapsläget kring sambedömning och dess eventuella effekter har en systematisk litteraturstudie valts som vetenskaplig metod (jfr Forsberg & Wengström, 2008), vilket innebär att sökningar efter kvalitativa och kvantitativa studier genomförs. Dessa studier granskas sedan och resultaten sammanförs till en ny helhet. Genom att använda resultaten från vetenskapliga studier kan en djupare förståelse och en bredare bild skapas samtidigt som det ger möjlighet att identifiera eventuella kunskapsluckor (ibid). Då de vetenskapliga studier och resultat som ligger till grund för denna litteraturstudie är rapporterade på olika sätt samt i stort sett saknar rapportering av kvantitativa data, har vi valt att inte göra en meta-studie utan istället en narrativ beskrivning av resultaten.

Litteratursökning

Studier söktes primärt under september 2012 i olika databaser. Följande sökord användes och kombinerades på olika sätt: teacher, assessment, teacher-based assessment, national test, moderation, marking, reliability, consensus, validity. Exempel på sökordskombinationer var teacher, assessment, moderation och teacher-based assessment, reliability. Den primära sökningen resulterade i ett stort antal sökträffar vilka utgjorde ett första urval. Efter genomläsning och katalogisering gjordes ett andra urval. Ett fåtal studier gick av olika anledningar inte att få fram eller var bortplockade ifrån de adresser som angavs. Även manuella sökningar utfördes utifrån referenslistor i syfte att utvidga det möjliga resultatet. Under arbetets gång gav tidskriften "Assessment in Education: Principles, Policy & Practice" ut ett temanummer om "Moderation practice and teacher judgement". Även dessa artiklar lästes igenom och relevanta artiklar inkluderades i denna studie.

Urvalskriterier

För att en studie skulle anses relevant och inkluderas i det andra urvalet sattes följande urvalskriterier upp:

1. Studien skulle vara empirisk, det vill säga den skulle ha studerat lärare som sambedömer, antingen som studiens huvudfokus eller som en delstudie.
2. Studien skulle avse sambedömning i grundskola, gymnasium eller vuxenutbildning (på grundskole- eller gymnasienivå) eller motsvarande. Inte förskola eller högre utbildning. Vilket skolämne sambedömningen utfördes i var inte begränsande.
3. Studien skulle vara publicerad och vetenskapligt granskad.

Urval

Urvalet av studier genomfördes i olika steg. I det första urvalet ingick 62 studier. Efter genomläsning föll ett antal studier ifrån, främst eftersom det inte handlade om sambedömning i bemärkelsen att lärarna utförde bedömningen tillsammans eller för att de inte var empiriska.

Efter en slutlig genomläsning återstod 23 studier vilka ligger till grund för denna litteraturstudie.

Analys

I studien har en kvalitativ innehållsanalys (se till exempel Graneheim & Lundman, 2004) genomförts. Innehållsanalysen inleddes genom att de studier som föll inom ramen för urvalskriterierna lästes igenom ett flertal gånger och därefter valdes meningsbärande enheter, *meaning units*, ut och fördes in i en tabell. Meningsbärande enheter är ord, meningar eller delar av texter som innehåller aspekter vilka relaterar till varandra genom innehåll och kontext (Graneheim & Lundman, 2004). Detta urval är en kritisk process (ibid.) och utfördes i två steg. Inledningsvis markerades delar av texterna och först efter ytterligare genomläsningar valdes enheterna ut och fördes in i en tabell. Kategorier skapades genom att de meningsbärande enheterna kodades och grupperades i gemensamma teman. I den genomförda studien skapades preliminära kategorier som efter diskussioner reviderades. Krippendorff (1980) menar att de kategorier som skapas måste vara både uttömmande och ömsesidigt uteslutande för att skapa tillförlitlighet. I något fall innebär kategoriseringen att data passade in i mer än en kategori. Graneheim och Lundman (2004) menar dock att det inte alltid är möjligt att skapa sådana ömsesidigt uteslutande kategorier när de texter studien baseras på innefattar människors erfarenheter, vilket är fallet i denna studie.

Resultat

Kontext och forskningsdesign

Av de 23 studierna har sju genomförts i Australien, främst i delstaten Queensland, sju i Storbritannien, sex i USA och en studie vardera från Portugal och Nya Zeeland. En av studierna har genomförts som en komparativ studie mellan Australien (Melbourne) och Hong Kong. De flesta studierna (17) innefattar lärare från olika nivåer i skolsystemet med en tonvikt mot grundskolan. Tre studier har omfattat lärare från motsvarande vårt gymnasium, två från högstadiet och en studie från lågstadiet. Merparten av studierna har varit huvudsakligen kvalitativa och inneburit att lärare som deltagit i sambedömning blivit intervjuade

före, under och efter sambedömningsinsatsen. Observationer av sambedömningsmöten har ibland genomförts och ibland har olika typer av dokument och andra underlag samlats in. En del studier har kompletterat sådana data med enkäter. Lärare och rektorer är de huvudsakliga informanterna, men även personer med särskild kunskap eller ansvar för bedömningsfrågor har ingått som informanter. Få studier är experimentella, där effekter av sambedömningsinsatser studerats och mätts på ett systematiskt sätt. Tre studier har varit kvantitativa i bemärkelsen att de använt data för att avgöra i vilken omfattning sambedömningen leder till en ökad bedömaröverensstämmelse. Resultaten i denna studie baseras därmed till betydande del på lärares uppfattningar av sambedömning. Lärare kan till exempel uppfatta att likvärdigheten i bedömningen på en skola har ökat utan att man faktiskt har belegg för att det också förhåller sig på detta vis.

Vilka effekter kan sambedömning ge?

Analysen av den forskning som ligger till grund för litteraturöversikten resulterade i att fyra kategorier av effekter kunde urskiljas:

- Effekter på lärares förståelse för kriterier och kvaliteter i elevers prestationer
- Effekter på samstämmighet vid bedömning
- Effekter på lärares kompetens
- Effekter på de kollegiala relationerna

Gränserna mellan dessa kategorier är inte skarpa vilket medför att resultat som kan sorteras in under en kategori eventuellt även kan höra hemma i en annan. Till exempel kan man anta att om en grupp lärare får en djupare förståelse för vad bedömningskriterierna innebär så ökar även kompetensen hos den enskilde läraren att bedöma sina elevers prestationer i förhållande till dessa. Resultaten i varje kategori kommer att redovisas genom att först ge en beskrivning av vad som är utmärkande för respektive kategori åtföljt av resultaten och exempel på någon eller några studier.

Effekter på lärares förståelse för kriterier och kvaliteter i elevers prestationer

Resultat inom denna kategori innebär att lärares förståelse för och uppfattning av de kriterier och kravnivåer som elevers prestationer ska bedömas utifrån vid sambedömning påverkas. Sådana uppfattningar blir synliga vid diskussioner utifrån konkreta elevprestationer varför även lärares förståelse för och uppfattning av olika kvaliteter i elevers prestationer hamnar inom denna kategori. Att lärare tillsammans diskuterar bedömning av elevers prestationer i förhållande till kriterier och kravnivåer verkar resultera i en ökad samsyn kring vad dessa kriterier och kravnivåer innebär och hur de kan användas (Connolly, Klenowski & Wyatt-Smith, 2012; Hall & Harding, 2002; Reid, 2007; Roberts, Wilson & Draney, 1997). Davison (2004) visar till exempel i en studie att lärare med erfarenhet av sambedömning har en större samsyn vid bedömning än lärare utan sådana erfarenheter. Trots det finns det variationer i hur lärare uppfattar och använder bedömningskriterier vilket kan tolkas som att även om bedömningar på ett ytligt plan är att betrakta som samstämmiga så kan de underliggande avvägningarna som ligger bakom ändå skilja sig åt (ibid.).

Effekter på samstämmighet vid bedömning

Till denna kategori räknas studier som ger belägg för att lärares samstämmighet vid bedömning av elevers prestationer påverkas vid sambedömning. I de studier som litteraturöversikten omfattar har detta uttryckts på olika sätt, till exempel i termer av reliabilitet, överensstämmelse och likvärdighet. Att lärares bedömning av elevers prestationer blir mer samstämmig som en direkt följd av sambedömning finns inte tydligt dokumenterat. I tre studier (de Eca & Torres, 2005; Falk & Ort, 1998; Syverson, 2009) rapporteras att lärares bedömningar överensstämmer till mellan 80 och 90 procent efter att de deltagit i sambedömning. Däremot är resultaten i samtliga dessa studier inte redovisade på ett sådant sätt att data kunnat verifieras. Det finns även studier som pekar på att sambedömning inte har någon större påverkan på lärares samstämmighet vid bedömning (Baird, Greatorex & Bell, 2004) och att lärare utan erfarenhet av sambedömning gör bedömningar som är jämförbara med bedömningar utförda av lärare med erfarenhet av sambedömning (Connolly et al., 2012; Morgan & Watson, 2002).

Att samstämmigheten vid bedömning ökar lyfts däremot fram i några studier som en *trolig* effekt av sambedömning (Adie, Klenowski & Wyatt-Smith, 2012; Clarke & Gipps, 2000), men det poängteras samtidigt att sådana antaganden behöver styrkas genom ytterligare studier. Det finns alltså inga tydliga resultat, varken för eller emot, som pekar på att sambedömning skulle leda till en mer samstämmig bedömning av elevers prestationer.

Effekter på lärares kompetens

När lärare sambedömer elevers prestationer kan det ge effekter som påverkar den enskilde lärarens kompetens i dennes yrkesutövning. Sådana effekter som kunde urskiljas i de genomgångna studierna var: en ökad kompetens hos lärare i att tolka styrdokument och att omsätta dessa i undervisning, ökad kompetens att urskilja kvaliteter i elevers prestationer, ökad kompetens att på ett tydligt sätt kommunicera kriterier och kravnivåer till elever samt ökad ämnes- och ämnesdidaktisk kunskap hos lärarna. I en studie av Bolt (2011) uppgav till exempel lärare att de genom sambedömning blivit bättre på att finna belägg för elevers kunnande och att sedan använda dessa för att förbättra undervisningen.

Effekter på de kollegiala relationerna

Att samarbeta med andra lärare genom sambedömning kan också ge effekter som påverkar de kollegiala relationerna på olika sätt. Till denna kategori räknas därför resultat som pekar på att till exempel känslan av gemenskap och tillhörighet ökar eller att lärare får stöd att hantera olika situationer i det dagliga arbetet. Även om detta inte är en faktor som har en direkt påverkan på bedömningens likvärdighet så har det betydelse för enskilda lärares upplevelse av att sambedöma och därmed för på vilket sätt dessa lärare deltar i sambedömningen. Roberts et al. (1997) visar till exempel på att en stark kollegialitet bidrar till framgång vid sambedömning medan Adie (2012) varnar för att framför allt oerfarna lärare kan vara särskilt känsliga för meningsskiljaktigheter och för att uttrycka en avvikande åsikt.

Vilka faktorer kan påverka sambedömningsprocessen?

Det finns även en rad olika faktorer som kan påverka sambedömningsprocessen och därmed även indirekt bedömningens och betygssättningens likvärdighet. I analysen identifierades ett antal sådana faktorer vilka delades in i tre olika kategorier: (1) faktorer som kopplas till struktur och organisation; (2) faktorer kopplade till sambedömningsituationen och (3) faktorer kopplade till deltagarna i sambedömningsprocessen.

Struktur och organisation

Att ha en tydlig och noga genomtänkt struktur och organisation kring sambedömningen tycks vara en avgörande faktor för utfallet och i denna kategori finns tre faktorer som är särskilt viktiga att beakta. Den första är att det finns en god struktur och ett tydligt ledarskap eftersom detta bidrar till att hålla fokus och att fördjupa diskussionerna på ett utvecklande sätt (Little, Gearhart, Curry & Kafka, 2003). För det andra är det av vikt att ta ställning till om syftet med sambedömningen är att nå konsensus eller inte. Det verkar som att det är när lärare har olika uppfattningar kring hur en prestation ska bedömas som det ges störst möjlighet till lärande (Adie, 2012). Den tredje faktorn kopplad till denna kategori berör hur gruppen som ska sambedöma är sammansatt. I studier där lärare intervjuats om sin uppfattning om sambedömning värderar de ofta sambedömning tillsammans med kollegor på den egna skolan eller i det egna skolområdet högst (se exempelvis Bolt, 2011). I andra studier argumenteras för värdet av att sambedömning sker mellan lärare från olika skolor eller distrikt eftersom det innebär ökade möjligheter att tillföra nya perspektiv i diskussionerna (se exempelvis Reid, 2007).

Bedömningssituationen

När lärare sambedömer så kan bedömningssituationen utformas på olika sätt. Till denna kategori räknas därför faktorer som kan påverkas av hur bedömningssituationen utformas som till exempel vilket bedömningsunderlag som används, hur underlaget är utvalt samt vilka eventuella bedömningsanvisningar som används. Sådana val kan få betydelse för utfallet av sambedömningen. Till exempel så tenderar

bedömningen att få en mer formativ karaktär om man utgår ifrån en icke-färdig prestation eftersom man då fokuserar på vad som kan bli nästa steg i elevens lärande (Reid, 2007). En viktig faktor för att sambedömningen ska bli framgångsrik är att det finns en variation och bredd bland de elevprestationer som väljs ut som underlag. En stor bredd av prestationer från elever med olika bakgrund är centralt för att lärare ska kunna göra mer nyanserade bedömningar av olika kvaliteter i elevernas prestationer och för att kunna ta med sig sina erfarenheter i nya situationer (Falk & Ort, 1998). Att kunna använda tillgängligt bedömningsmaterial, så som bedömningsanvisningar och kravnivåer, på ett flexibelt sätt är ytterligare en central faktor eftersom det hjälper lärare att hålla fokus på elevprestationerna och att fördjupa diskussionerna (Little et al., 2003).

Bedömnarna

Till denna kategori räknas faktorer som kan knytas till deltagarna och det sätt på vilket de deltar. Det kan till exempel handla om den påverkan den ursprungliga bedömaren eller den undervisande läraren har om denne deltar i sambedömningen. Ett annat exempel är deltagarnas olika erfarenheter. Oerfarna lärare verkar ha särskilt stort utbyte av sambedömning (Baird et al., 2004; Crisp, 2013) eftersom det kan vara ett sätt att få stöd i att utföra bedömningar och att få ta del av mer erfarna kollegors kunskaper. Det kan emellertid vara svårt för en oerfaren lärare som har en avvikande uppfattning att göra sin röst hörd i förhållande till mer erfarna kollegor (Adie, 2012), vilket i sin tur kan påverka både diskussionerna och hur man ställer sig till att fortsätta delta i sambedömning (Adie, 2012; Adie et al., 2012; Black, Harrison, Hodgen, Marshall & Serret, 2010).

Resultatdiskussion

Studien har syftat till att förstå på vilket sätt sambedömning kan bidra till en likvärdig bedömning och betygssättning av elevers prestationer genom att söka svar på vilka dokumenterade effekter sambedömning kan ha samt hur denna process kan påverkas av olika faktorer. Resultaten pekar på att sambedömning kan bidra till en ökad likvärdighet främst eftersom det kan medföra en samsyn kring hur styrdokumentet tolkas

och kring hur elevers prestationer bedöms, men även eftersom lärare kan få en ökad kompetens att undervisa och bedöma.

Som tidigare beskrivits kräver en likvärdig bedömning att lärare är samstämmiga i flera tolkningsled. I detta avsnitt kommer frågan om huruvida sambedömning har potential att leda till en sådan samstämmighet att diskuteras i förhållande till de tre tolkningsleden. Inledningsvis lyfts argument som talar för detta fram (dvs. ökad bedömarkompetens och ökad samsyn) och därefter argument som talar emot (dvs. lärares olika tolkning, att samsyn inte är samma sak som samstämmighet och sammanvägningen till omdömen).

Ökad bedömarkompetens

Ett av argumenten för att likvärdigheten i lärares bedömning och betygssättning ökar som en följd av sambedömning är att resultaten pekar på att det leder till ökad bedömarkompetens hos lärarna. Sambedömning ger lärare möjlighet till diskussion och reflektion kring frågor kopplade till undervisning och bedömning utifrån konkreta elevprestationer. Vid diskussioner kring kvaliteter i elevers prestationer tvingas lärare att sätta ord på vad som skiljer olika prestationer från varandra och vad som är prestationens styrkor respektive svagheter. Därmed måste lärare även identifiera vad som är nästa steg i elevernas lärande, vilket i sin tur kan innebära att lärare blir bättre på att finna belägg för elevers kunnande och att använda kunskapen som utgångspunkt vid planering av undervisning (Bolt, 2011; Falk & Ort, 1998; Malone, De Lucchi & Long, 2004). Ökad bedömarkompetens i kombination med utvecklade insikter som stödjer undervisningen (Bolt, 2011; Clarke & Gipps, 2000; Falk & Ort, 1998) innebär därmed att både det första och det andra tolkningsledet stärks, och därmed även likvärdigheten. Resultaten av denna forskningsöversikt ger däremot inga tydliga belägg för att lärares bedömningskompetens beträffande sammanvägning av olika prestationer till ett kunskapsomdöme eller betyg, det vill säga tolkningsled tre, ökar.

Ökad samsyn

Ett annat resultat som talar för att sambedömning kan leda till en ökad likvärdighet vid bedömning och betygssättning är att det skapar en

samsyn mellan lärare kring vad mål, kriterier och kravnivåer innebär (Connolly et al., 2012; Hall & Harding, 2002; Reid, 2007; Roberts et al., 1997). Att lärare får en samsyn kring styrdokumentet beror bland annat på att olika sätt att uppfatta dessa blir synliggjorda och föremål för diskussion utifrån konkreta elevprestationer. Adie et al. (2012) menar att en gemensam förståelse bland lärare för de begrepp som används är en förutsättning för att kunna diskutera och sambedöma. Att utgå ifrån konkreta elevprestationer tvingar lärare att sätta ord på sina uppfattningar så att dessa blir tydliga och medvetandegjorda, både för den enskilde läraren men även för kollegor, vilket kan leda till att lärare omförhandlar sin bedömning (ibid.). Enligt Sadler (2009) är sådana kravnivåer som används i styrdokumentet abstrakta och för att utveckla en förståelse för dessa behövs just konkreta exempel för att illustrera olika kvaliteter. Diskussioner utifrån konkreta exempel blir då betydelsefulla för att utveckla en gemensam förståelse för innebörden av kravnivåerna och för utvecklandet av en känsla för vad som utmärker olika kvaliteter. Harlen (2004) menar att hur pass konsekventa lärare är i vid bedömningen av elevprestationer hör ihop med hur man tolkat styrdokumentet och hur dessa omsatts i undervisning. Att därför låta sambedömningen omfatta även det första tolkningsledet blir då betydelsefullt för en likvärdig bedömning och betygssättning eftersom samsyn påverkar både reliabilitet och validitet vid bedömning (Harlen, 2004; Holroyd, 2000).

Lärares olika tolkning

Ökad bedömarkompetens och samsyn ger goda förutsättningar för en likvärdig bedömning och betygssättning, men det finns å andra sidan sådant som talar emot detta. Ett problem kan vara att även om lärare enas om bedömningen av en elevprestation eller kring tolkningen av styrdokumentet så är det inte självklart att en sådan tolkning är korrekt eller i linje med hur andra lärare gör motsvarande tolkningar. En konsekvens kan då bli att lärares bedömningar blir samstämmiga utan att vara i linje med styrdokumentet. Ett sätt att komma ifrån att lärares bedömningar sker utifrån en felaktig tolkning av styrdokumentet skulle kunna vara att enbart utgå från till exempel nationella prov vid bedömning och betygssättning, eftersom olika provkonstruktörer här gjort tolkningen av styrdokumentet åt lärarna. Å andra sidan ger

det inte lärare den kunskap och kompetens som krävs för att själva kunna planera och genomföra undervisning och bedömning i linje med styrdokumentet. Genom att låta sambedömningen omfatta det första tolkningsledet kan man däremot förvänta sig att lärare utvecklar en samsyn kring styrdokumentet som i sin tur kan ge förutsättningar för en mer samstämmig bedömning också i andra bedömningssituationer än vid bedömningen av de nationella proven. En annan konsekvens kan vara att likvärdigheten tenderar att bli tämligen lokal när sambedömning sker mellan lärare på enskilda eller närliggande skolor. Det finns studier där lärare träffats över skolgränser och ibland också över internet, vilket kan skapa förutsättningar för nationell likvärdighet. Ett utvecklat exempel med goda erfarenheter från detta är i Queensland, Australien. Här har man inga nationella prov utan istället ett väl utbyggt system för sambedömning i flera steg, på lokal, regional och nationell nivå (Maxwell, 2007). Detta kräver dock en organisation och struktur som i dagsläget inte finns i Sverige.

Samsyn är inte samstämmighet

Ökad samsyn betyder inte heller nödvändigtvis att likvärdigheten vid bedömning och betygssättning ökar, eftersom samsyn och samstämmighet inte är samma sak. Det finns få studier som styrker att sambedömning ger ökad samstämmighet vid bedömning och de studier som återfunnits ger motsägelsefulla resultat. Sadler (1998) beskriver den del av bedömningsprocessen där lärare tolkar och värderar elevers prestationer som ett förlopp i tre steg: (1) uppmärksamhet riktas mot elevprestationen; (2) elevprestationen värderas mot till exempel ett ramverk och (3) poäng, betygsbelägg eller dylikt utdelas. Utifrån en sådan beskrivning av lärares bedömning kan samsyn innebära att lärare är överens i steg 2 om hur en elevprestation ska tolkas och värderas, det vill säga vad som utifrån bedömningskriterierna är elevprestationens styrkor och vad i denna som är mindre bra. En sådan samsyn betyder inte nödvändigtvis att lärare i nästa steg (3) är överens och ger elevprestationen exakt samma poäng eller betygsbelägg. Även det omvända förhållandet kan råda, det vill säga att lärare i steg 2 värderar olika kriterier på olika sätt i förhållande till varandra men ändå gör en likartad helhetsbedömning i steg 3. En svensk studie genomförd av Olofsson (2006) pekar på detta. I studien har fyra lärare bedömt

elevarbeten från det nationella provet i matematik, först enskilt och därefter ännu en gång efter att ha sambedömt där emellan. Resultatet visar att lärarna efter sambedömning ändrade sina bedömningar i viss utsträckning så att de blev mer samstämmiga men även att lärare var olika benägna att revidera sina bedömningar, bland annat beroende på vilka kriterier de hade olika uppfattningar kring. Detta kan få till följd att olika elevers prestationer, även om man sambedömer, kan få en likartad bedömning men att denna baseras på olika kriterier. Om detta är att betrakta som godtagbart eller inte beror på i vilken utsträckning man kräver samstämmighet och på vilka grunder. Måste lärares bedömning och betygssättning vara exakt lika eller är det tillräckligt att vara överens om på vilka grunder elevprestationerna bedöms? Ett sätt att få till stånd samstämmighet är förstås att kräva konsensus. En nackdel med detta, förutom att det är väldigt tidskrävande att bedöma samtliga elevprestationer tillsammans, är dock att det finns en risk att olika uppfattningar blir svåra att hantera och att målet blir att komma överens snarare än att skapa en samsyn (Davison, 2004). Moss och Schutz (2001) varnar för att en strävan efter konsensus kan leda till en falsk trygghet där viktiga argument utesluts, vilket pekar på betydelsen av att skapa goda förutsättningar för ett öppet och tillåtande diskussionsklimat om sambedömningen ska ha möjlighet att leda till samsyn kring styrdokument och bedömning av elevprestationer.

Sammanvägning

Ett annat problem med en likvärdig bedömning är kopplat till det tredje och sista tolkningsledet när elevers olika prestationer ska sammanvägas till ett kunskapsomdöme eller ett betyg utifrån olika kravnivåer. För en likvärdig bedömning i detta led krävs att kunskapskraven tolkas på samma sätt av olika lärare och att sammanvägningen till omdöme och betyg sker enligt styrdokumentens föreskrifter. I en studie av Allal (2013) uppgav lärare att de ofta samarbetade i det första tolkningsledet, med att planera undervisning och att konstruera bedömningsuppgifter medan det var mindre vanligt med samarbete i de övriga tolkningsleden. Dessutom visade det sig att lärarna vid sammanvägning av elevers resultat till ett betyg gav olika tyngd åt olika informationskällor och att det skiljde sig åt hur man hanterade olika elevers resultat (ibid). Även Korp (2006) visar utifrån svenska förhållanden att lärare vid

betygssättning använder olika metoder för att sammanväga olika elevprestationer till ett betyg. Konsekvensen kan bli att även om lärare är överens om vad som ska bedömas och att enskilda bedömningar är både reliabla och valida så blir sammanvägningen av olika resultat en process som hanteras på skilda sätt. Även om det utifrån denna studie inte finns tydliga belägg för att sambedömning leder till ökad likvärdighet kring det tredje tolkningsledet finns det andra erfarenheter som talar för att sådan kan stödjas genom sambedömning, exempelvis erfarenheter ifrån Queensland. I en rapport från Queensland Studies Authority (QSA) där man granskat ett stort urval elevportföljer, fann man en överensstämmelse mellan skolornas och granskningsgruppens bedömning i 92 procent av fallen (QSA, 2012). Detta tyder på att sambedömning kan stödja även betygssättningen.

Slutsatser

Det visar sig således att den befintliga forskningen inte ger ett entydigt svar på frågan om sambedömning leder till en ökad likvärdighet vid bedömning och betygssättning eftersom utfallet av sambedömningsprocessen kan påverkas av flera olika faktorer. Den slutsats som ändå kan dras är att sambedömning har potential att stödja en likvärdig bedömning och betygssättning eftersom det finns belägg för att sambedömning kan bidra till en samsyn och till en ökad bedömarkompetens bland lärare. Man bör däremot inte förvänta sig att en samsyn är det samma som att lärare ger samma poäng eller betygsbelägg för samma elevprestation utan snarare att de grunder utifrån vilka sådana omdömen ges är likartade. Vidare bör sambedömning ses som en process som omfattar och inkluderar samtliga tolkningsled i bedömningsprocessen om sambedömning ska kunna leda till en ökad likvärdighet. Ju fler av tolkningsleden som sambedömningen omfattar desto större är förutsättningarna för ökad likvärdighet vid bedömning och betygssättningen.

En ökad likvärdighet till följd av sambedömning bör även ses som en lokal process som sker på lång sikt och med behov av en tydlig struktur och ledning. Att likvärdigheten ökar på lokal nivå är däremot inte det samma som att likvärdigheten ökar på regional eller nationell nivå, om inte dessa nivåer kopplas samman på något sätt. En sammankoppling

kräver en välutbyggd struktur och omfattande satsningar. Erfarenheter från Queensland, Australien, har visat att detta är möjligt och dessutom med hög likvärdighet som följd.

Att identifiera andra enskilda faktorer som är av särskild betydelse för att sambedömning ska leda till en ökad likvärdighet är svårare. Något som kan ses som centralt för att sambedömning ska vara framgångsrikt är variation. Variation blir då en indirekt faktor som påverkar likvärdigheten genom att: (1) olika åsikter och perspektiv beaktas för att skapa förutsättningar för fördjupade diskussioner; (2) att urvalet av elevprestationer för bedömning speglar en bredd och variation av prestationer samt (3) att lärare kommer ifrån olika kontexter och ges möjlighet till sambedömning i olika konstellationer.

Att befintlig forskning om sambedömning och dess effekter till betydande del grundas på intervjuer med lärare eller på observationer av lärare när de sambedömer får till följd att forskningen i stor utsträckning baseras på lärares *uppfattningar* om sambedömning och dess effekter. Till exempel kan lärare uppfatta att deras bedömningar blir mer samstämmiga utan att ha belägg för att det verkligen förhåller sig på detta sätt. Därför finns ett behov av ytterligare studier som på ett systematiskt sätt undersöker om sambedömning verkligen leder till ökad samstämmighet i lärares bedömning. Det finns även en avsaknad av forskning som studerat sambedömning i det tredje tolkningsledet, när elevers prestationer ska sammanvägas till ett kunskapsomdöme eller till ett betyg. Allal (2013) ser detta tolkningsled som centralt för att stödja en likvärdig bedömning och betygssättning eftersom detta är en process som hanteras på olika sätt av olika lärare.

Referenslista

- Adie, Leonore E. (2010). *Developing shared understandings of standards-based assessment: On-line moderation practices across geographically diverse contexts*. Doktorsavhandling, Queensland University of Technology, Australien.
- Adie, Leonore (2012). Learning as identity and practice through involvement in online moderation. *Educational Assessment, Evaluation and Accountability*, 24, 43-56.
- Adie, Leonore E., Klenowski, Valentina & Wyatt-Smith, Claire (2012). Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*, 64, 223-240.
- Allal, Linda (2013). Teachers' professional judgement in assessment: a cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20, 20-34.
- Baird, Jo-Anne, Greator, Jackie & Bell, John F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in Education: Principles, Policy & Practice*, 11, 331-348.
- Biggs, John B. & Tang, Catherine S. (2007). *Teaching for quality learning at university: what the student does*. (3. ed.) Maidenhead: McGraw-Hill/Society for Research into Higher Education.
- Black, Paul, Harrison, Christine, Hodgen, Jeremy, Marshall, Bethan & Serret, Natasha (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17, 215-232.
- Bolt, Susan (2011). Making consistent judgments: Assessing student attainment of systemic achievement targets. *Educational Forum*, 75, 157-172.
- Clarke, Shirley & Gipps, Caroline (2000). The role of teachers in teacher assessment in England 1996-1998. *Evaluation and Research in Education*, 14, 38-52.
- Connolly, Stephen, Klenowski, Valentina & Wyatt-Smith, Claire M. (2012). Moderation and consistency of teacher judgement: teachers' views. *British Educational Research Journal*, 38, 593-614.

- Crisp, Victoria (2013). Criteria, comparison and past experiences: how do teachers make judgements when marking coursework? *Assessment in Education: Principles, Policy & Practice*, 20, 127-144.
- Crooks, Terence J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Davison, Chris (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- de Eca, Pereira & Torres, Maria T. (2005). Using portfolios for external assessment: An experiment in Portugal. *International Journal of Art & Design Education*, 24, 209-218.
- Erickson, Gudrun (2009). *Nationella prov i engelska – en studie av bedömsamstämmighet*. Hämtad 24 februari 2014, från http://www.nafs.gu.se/digitalAssets/1319/1319572_bed.np-eng-g.erickson-2009.pdf
- Falk, Beverly & Ort, Suzanna (1998). Sitting down to score. Teacher learning through assessment. *Phi Delta Kappan*, 80, 59-64.
- Forsberg, Christina & Wengström, Yvonne (2008). *Att göra systematiska litteraturstudier: värdering, analys och presentation av omvårdnadsforskning*. (2., [uppdaterade] utg.) Stockholm: Natur & Kultur.
- Gipps, Caroline V. (1994). *Beyond testing: towards a theory of educational assessment*. London: Falmer.
- Graneheim, Ulla H. & Lundman, Berit (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, 24(2), 105-112.
- Gustavsson, Anders, Måhl, Per & Sundblad, Bo (2012). *Betygsättning: en handbok*. (1. uppl.) Stockholm: Liber.
- Hall, Kathy & Harding, Austin (2002). Level descriptions and teacher assessment in England: towards a community of assessment practice. *Educational Research*, 44(1), 1-16.
- Harlen, Wynne (Ed.) (1994) *Enhancing Quality in Assessment*. BERA Policy Task Group on Assessment, Paul Chapman Publishers.

- Harlen, Wynne (2004). A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, Wynne (2005). Teachers' summative practices and assessment for learning - tensions and synergies. *Curriculum Journal*, 16(2), 207-223.
- Holroyd, Colin (2000). Are assessors professional? *Active Learning in Higher Education 1* (1), 28-44.
- Klenowski, Val & Adie, Leonore (2009). Moderation as judgement practice: Reconciling system level accountability and local level practice. *Curriculum Perspectives*, 29, 10-28.
- Klenowski, Valentina & Wyatt-Smith, Claire (2010). Standards-Driven Reform Years 1-10: Moderation an Optional Extra? *Australian Educational Researcher*, 37(2), 21-39.
- Klenowski, Valentina (2012). Raising the stakes: the challenges for teacher assessment. *The Australian Educational Researcher*, 39(2), 173-192.
- Korp, Helena (2006). *Lika chanser i gymnasiet? En studie om betyg, nationella prov och social reproduktion*. Doktorsavhandling, Malmö högskola.
- Krippendorff, Klaus (1980). *Content Analysis. An Introduction to its Methodology*. The Sage Commtext Series, Sage Publications Ltd., London.
- Linn, Robert L. (1993). Linking Results of Distinct Assessments. *Applied Measurement in Education*, 6(1), 83-102.
- Little, Judith W., Gearhart, Maryl, Curry, Marnie & Kafka, Judith (2003). Looking at student work for teacher learning, teacher community, and school reform. *Phi Delta Kappan*, 85, 185-192.
- Malone, Larry, De Lucchi, Linda, & Long, Kathy (2004). All Things in Moderation. *Science and Children*, 41(5), 30-34.

- Maxwell, Graham S. (2002). Progressive assessment: Synthesising formative and summative purposes of assessment. In Frank Ventura and Grace. Grima (eds), *Contemporary issues in educational assessment: Proceedings of the ACEAB Second International Conference, Malta 2002*, pp. 220–234. Msida, Malta: MATSEC Examinations Board.
- Maxwell, Graham (2007). *Implications for moderation of proposed changes to senior secondary school syllabuses*. Paper commissioned by the Queensland Studies Authority. Brisbane: Queensland Studies Authority.
- Morgan, Candia & Watson, Anne (2002). The interpretative nature of teachers' assessments of students' mathematics: Issues for equity. *Journal for Research in Mathematics Education*, 33, 78-110.
- Moss, Pamela A. & Schutz, Aaron (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38, 37-70.
- Olofsson, Gunilla (2006). *En studie av lärares bedömning av eleverbeten på ett nationellt prov i matematik kurs A*. Stockholms universitet, PRIM-gruppen.
- Queensland Studies Authority. (2012). *Random sampling project 2012. Report on random sampling of assessment in Authority subjects*. Brisbane: Queensland Studies Authority.
- Reid, Lesley (2007). Teachers talking about writing assessment: valuable professional learning. *Improving Schools*, 10, 132-149.
- Roberts, Lily, Wilson, Mark & Draney, Karen (1997). The SEPUP assessment system: An overview *BEAR Report Series* (Vol. SA-97-1). Berkeley: University of California.
- Sadler, D. Royce (1998). Formative Assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84.
- Sadler, D. Royce (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34(7), 807-82.
- Sadler, D. Royce (2013). Assuring academic achievement standards: from moderation to calibration, *Assessment in Education: Principles, Policy & Practice*, 20 (1), 5-19.

- Skolinspektionen (2012). *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*. Stockholm: Skolinspektionen.
- Skolinspektionen (2013). *Tillsyn av bedömning och betygssättning*. Stockholm: Skolinspektionen.
- Skolverket (2004). *Handlingsplan för en rättssäker och likvärdig betygssättning*. Stockholm: Skolverket.
- Skolverket (2009a). *Redovisning av regeringsuppdrag att ge förslag på hur det nationella provsystemet bör utvecklas och utformas*. Stockholm: Skolverket.
- Skolverket (2009b). *Bedömaröverensstämmelse vid bedömning av nationella prov*. Dnr 2008:286. Stockholm: Skolverket.
- Skolverket (2011a). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011*. Stockholm: Skolverket.
- Skolverket (2011b). *Läroplan, examensmål och gymnasiegemensamma ämnen för gymnasieskola 2011*. (2011). Stockholm: Skolverket.
- Skolverket (2012a). *Betygsskalan A-E*. Hämtad 24 februari 2014, från: <http://www.skolverket.se/bedomning/betyg>
- Skolverket (2012b). *Likvärdig utbildning i svensk grundskola? En kvantitativ analys av likvärdighet över tid. Rapport nr. 374*. Stockholm: Skolverket.
- Skolverket (2012c). *Redovisning av uppdrag om avvikelser mellan provresultat och betyg i grundskolan årskurs 9*. Dnr 75-2012:311. Stockholm: Skolverket.
- Skolverket (2012d). *Redovisning av uppdrag om avvikelser mellan provresultat och kursbetyg i gymnasieskolan*. Dnr 75-2012:311. Skolverket: Stockholm.
- Syverson, Margaret A. (2009). Social justice and evidence-based assessment with the Learning Record. *Forum on Public Policy*, 1-27.
- Wilson, Mark (Ed.). (2004). *Towards coherence between classroom assessment and accountability*. Chicago: University of Chicago Press.