

# Samsyn eller samstämmighet? En diskussion om sambedömning som redskap för likvärdig bedömning i skolan

ANDERS JÖNSSON & PIA THORNBERG

Avdelningen för naturvetenskap  
Högskolan Kristianstad

På senare tid har *sambedömning*, det vill säga att lärare tillsammans genomför eller diskuterar bedömning och/eller betygssättning, förts fram som en strategi för att få lärarnas bedömning och betygssättning mer likvärdig. Samtidigt har flera kommuner i Sverige gjort omfattande satsningar på olika former av sambedömning, men mer med anledning av sambedömningens förmodade positiva effekter på lärarnas bedömningskompetens och -praktik i riktning mot formativ bedömning än av likvärdighetsskäl. En intressant fråga är därmed vad som händer i mötet mellan kommuners och skolors strävan att förändra klassrumsbedömningen mot formativ bedömning å ena sidan och statsmakternas krav på ökad likvärdighet i bedömning av nationella prov och betygssättning å den andra. Syftet med denna artikel är att, utifrån en systematisk översikt om forskning kring sambedömning, diskutera möjligheten för sambedömning att leva upp till förväntningarna vad gäller ökad likvärdighet samtidigt som man har ett ökat fokus på kompetensutveckling och formativ bedömning.

## SAMBEDÖMNING

På senare tid har *sambedömning*, det vill säga att lärare tillsammans genomför eller diskuterar bedömning och/eller betygssättning, förts fram som en strategi för att få lärarnas bedömning och betygssättning mer likvärdig (bl.a. Erickson, 2009; Skolinspektionen, 2013; Skolverket, 2009b). Skolinspektionen (2012) lyfter till exempel fram sambedömning som ett sätt att uppnå

en större samsyn och samstämmighet kring bedömning och betygssättning, i sina slutsatser från de omrättningar som under tre år gjorts av nationella prov i grundskolan och gymnasieskolan. I en rapport av Skolverket (2009b) från en större undersökning av bedömaröverensstämmigheten i de nationella proven, konstaterar man att bedömaröverensstämmelsen är god eller mycket god i vissa ämnen, men inte i alla. Den rekommendation man ger för att öka likvärdigheten vid bedömning och betygssättning är just sambedömning av de nationella proven. Skolverket (2009a) skriver också 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” (s. 17). Regeringen har med anledning av detta gett Skolverket i uppdrag att sprida modeller för sambedömning som kan tänkas bidra till ökad likvärdighet.

Det finns dock även andra anledningar till att arbeta med sambedömning, som inte nödvändigtvis har sin grund i en övertygelse om ökad likvärdighet, där ett centralt motiv är att förbättra undervisningen och på så sätt öka elevernas måluppfyllelse. Framför allt har Dylan Wiliam (2007) haft stort inflytande, genom att lansera så kallade ”Teacher Learning Communities” (TLCs), det vill säga grupper av lärare som regelbundet träffas för att diskutera just bedömningsfrågor, som ett sätt att implementera formativ bedömning (dvs. bedömning där man utnyttjar information om elevers prestationer för att stödja elevernas fortsatta utveckling, se t.ex. Jönsson, 2014) i undervisningen. Wiliams argument för TLCs utgår dels från att det finns få belegg för att tidigare satsningar på lärarfortbildning gett någon utdelning i termer av förbättrade elevprestationer, dels från att en förändring av lärares bedömningspraktik innebär en omvälvande omställning av djupt rotade sätt att arbeta och tänka, vilket kräver nya former av kompetensutveckling<sup>1</sup>. Flera kommuner i Sverige (t.ex. Borås, Lund och Örebro) har i linje med Wiliams uppfattning gjort omfattande satsningar på TLCs och andra former av sambedömning, men mer med anledning av sambedömningens förmodade positiva effekter på lärarnas bedömningskompetens och -praktik i riktning mot formativ bedömning än av likvärdighetskäl (se t.ex. Tolgfors, 2014).

En intressant fråga är därmed vad som händer i mötet mellan kommuners och skolors strävan att förändra klassrumsbedömningen mot formativ bedömning (och i förlängningen förbättrade elevprestationer) å ena sidan och statsmakternas krav på ökad likvärdighet i bedömning av nationella prov och

---

1 Med kompetensutveckling avses att lärarna genomgår någon form av organiserad utbildning, som ger dem möjlighet att *utvecklas inom sin yrkesroll*. Kompetensutveckling behöver inte avse formell undervisning, till exempel i form av kurser, utan kompetensutveckling kan – som i fallet med TLCs – organiseras inom ramen för den ordinarie yrkesverksamheten.

betygsättning å den andra. Syftet med denna artikel är att diskutera möjligheten för sambedömning att leva upp till förväntningarna vad gäller ökad likvärdighet samtidigt som man har ett ökat fokus på kompetensutveckling och formativ bedömning.

## LIKVÄRDIG BEDÖMNING OCH BETYGSÄTTNING

Begreppet likvärdighet har använts med skiftande innebörd i olika tider. Tomas Englund och Ann Quennerstedt (2008) beskriver hur innebörden av begreppet har förändrats i den allmänna diskussionen från 1960-talet och fram till idag. De menar att begreppet tidigare hade en betydelse som var starkt kopplad till jämlikhet och enhetlighet, medan innebörden av likvärdighet som den diskuteras idag är snävare och handlar mer om måluppfyllelse och utbildningsresultat. Denna förändring kan ses som en konsekvens av det mål- och resultatstyrda system som infördes under 1990-talet. Likvärdighet i ett sådant perspektiv innebär bland annat att likvärdiga elevprestationer förväntas få samma bedömning och betyg oberoende av *vem* som bedömer eller *hur* prestationen bedöms. Grunden i de bedömningar som utförs förväntas även vara i linje med kurs- och ämnesplaners mål och kunskapskrav.

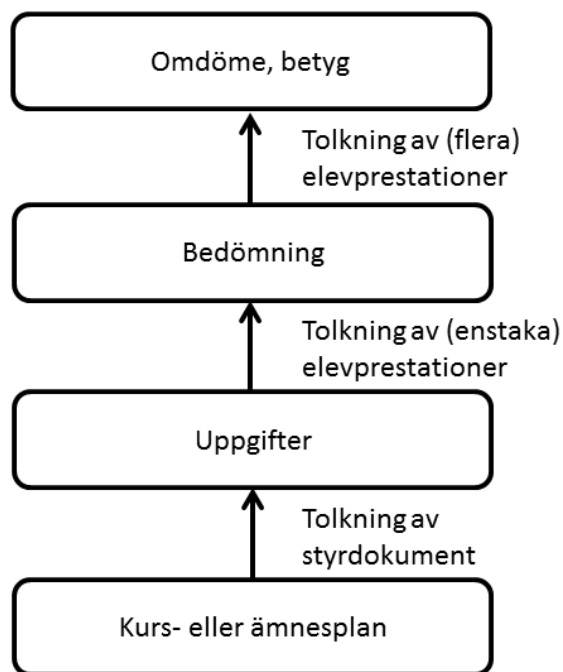
Vad innebär det då, att likvärdiga elevprestationer förväntas få samma bedömning eller betyg? Inte sällan ligger fokus på att själva poängsumman eller betygsbeteckningen överensstämmer, till exempel genom att två olika lärare ger samma betygsbeteckning åt likvärdiga prestationer. Men om lärarna har satt dessa poäng eller betyg på helt olika grunder, är bedömningen ändå att betrakta som likvärdig? Ett sätt att förstå dessa olika aspekter av likvärdig bedömning, är genom begreppen ”samsyn” och ”samstämmighet”.

Royce Sadler (1998) beskriver bedömningen som en process i tre delar, vilket innebär att lärare (1) tolkar elevprestationen, (2) värderar den och (3) avger ett omdöme i form av exempelvis en poäng, en kvalitetsbeskrivning eller ett betyg. En ”samsyn” i bedömningen omfattar de två första punkterna i Sadlers beskrivning, vilket kan innebära att lärarna har en gemensam uppfattning om vad det är som ska bedömas, till exempel vilka kriterier som ska tillämpas. Att lärare har en samsyn innebär dock inte nödvändigtvis att de avger samma omdöme, det vill säga har ”samstämmighet” i sin bedömning. Att samstämmigheten ofta är låg mellan lärares bedömningar, är väl belagt inom bedömningsforskningen, där man konsekvent har visat på stora skillnader mellan lärares bedömningar av komplexa uppgifter, det vill säga uppgifter som kräver tillämpning av flera kriterier (se t.ex. Parkes, 2013, för en översikt), och denna låga samstämmighet kvarstår ofta även om lärarna har tillgång till detaljerade bedömningsanvisningar (Jönsson & Svingby,

2007). En förklaring till detta, är att lärarna visserligen kan vara överens om vilka kriterier som ska ingå i bedömningen och hur de ska tillämpas, men ändå lägger olika vikt vid respektive kriterium vid sammanvägningen till ett omdöme. Även det omvända förhållandet kan råda, det vill säga att lärare ger samma omdöme, men där omdömet baseras på skilda sätt att tolka och värdera elevprestationerna (se t.ex. Davison, 2004). Frågan är dock i vilken av dessa situationer som bedömningen kan anses vara likvärdig: Är det när lärare har en samsyn i bedömningen, när deras omdöme är i överensstämmelse eller både och? För många har säkert denna fråga ett självklart svar, nämligen att det är viktigare att lärare utgår från en likvärdig tolkning av elevprestationer och bedömer enligt samma kriterier, snarare än att överensstämmelsen är exakt. Samtidigt är det naturligtvis svårare att utvärdera huruvida lärare har en samsyn jämfört med en samstämmighet. Det är under vissa omständigheter också lättare att åstadkomma en hög samstämmighet genom träning och detaljerade bedömningsanvisningar, vilket man regelmässigt gör vid bedömning av elevsvar från storskaliga prov och kunskapsmätningar.

Ytterligare en komplikation i diskussionen kring likvärdigheten i lärares bedömningar är att själva bedömningsmomentet endast utgör *en* del i en tolkningskedja från styrdokumentet till betygsättningen. Linda Allal (2013) har skapat en modell som beskriver dessa olika tolkningsled (Figur 1), där första ledet innebär att styrdokumentet tolkas och omsätts i undervisning och bedömningsuppgifter. När eleverna genomfört uppgifterna, behöver deras prestationer tolkas och värderas utifrån bedömningsanvisningar och/eller kunskapskrav (dvs. bedömas), vilket utgör ett andra tolkningsled i modellen. Slutligen ska de olika prestationer, som tillsammans bildar lärares betygsunderlag, sammanvägas i ett betyg eller i någon form av kunskapsomdöme. För att lärares bedömning och betygsättning ska vara likvärdig krävs någon form av samsyn eller samstämmighet i alla dessa tolkningsled.

Nedan presenteras forskning om sambedömning, vilken organiserats utifrån olika syften med sambedömning, som underlag för att diskutera i vilken utsträckning sambedömning förmår att stödja samsyn eller samstämmighet i dessa tolkningsled. Presentationen bygger på en systematisk översikt av forskning om sambedömning (Thornberg, 2014).



**Figur 1.** Vid bedömning och betygssättning måste tolkningar ske i flera led, det vill säga (1) tolkning av styrdokumenterna för undervisning och konstruktion av bedömningsuppgifter, (2) tolkning av elevprestationer samt (3) tolkning av det samlade underlaget inför utvecklingsamtal eller betygssättning. För att bedömning och betygssättning ska bli likvärdiga, behöver alla tre tolkningsled inkluderas. Figuren är anpassad från Allal (2013).

## FORSKNING OM SAMBEDÖMNING

### Sambedömning för ökad samstämmighet

Som nämndes ovan, har samstämmigheten i lärares bedömningar fått mycket uppmärksamhet i forskningslitteraturen. Redan i början på 1900-talet genomfördes studier, som visade på låg överensstämmelse mellan lärares bedömningar. Daniel Starch och Edward Elliott (1912) skickade till exempel ut kopior av två elevsvar skrivna av elever i high school till 200 skolor och bad lärare bedöma prestationerna ”according to the practice and standards of the school” (s. 449). Resultaten visade på en mycket stor spridning mellan olika lärares bedömning av samma elevprestationer (mellan 60-97 och 50-97 på en skala mellan 1-100 för de båda elevsvaren). En mycket likartad studie genomfördes cirka 100 år senare, som gav nästan identiska resultat (Brimi, 2011). Med anledning av den generellt låga samstämmigheten mellan

lärares bedömningar, har man undersökt en rad olika åtgärder för att öka samstämmigheten, som att väga samman bedömningar från flera oberoende bedömare eller att bedöma med hjälp av detaljerade bedömningsanvisningar. Det rekommenderas ofta även att begränsa komplexiteten i uppgifterna, för att på så sätt minska spridningen mellan olika lärares bedömningar, inte minst i storskaliga bedömningar som nationella prov (t.ex. Lie, 2007).

På senare tid har alltså även sambedömning föreslagits som ett sätt att åstadkomma en ökad samstämmighet i lärares bedömningar, till exempel vid rättningen av nationella prov (Erickson, 2009; Skolinspektionen, 2013; Skolverket, 2009b). Pia Thornberg (2014), som har genomfört en systematisk litteraturstudie för att undersöka vilken forskning som bedrivits kring sambedömning och sambedömningens effekter, hävdar emellertid att det inte finns något tydligt stöd för att sambedömning skulle leda till ökad samstämmighet, i betydelsen att lärare tilldelar likvärdiga prestationer samma poäng eller betygsbeteckning. Thornberg skriver vidare att merparten av forskningen om sambedömning grundar sig på intervjuer av lärare före, under och efter sambedömningsinsatsen, men att det är betydligt mer sällsynt med studier där man studerat effekter av sambedömningsinsatser på ett systematiskt sätt. Sammantaget innebär det att forskningen:

*/.../ till betydande del bygger på lärares uppfattningar av sambedömning och i mindre omfattning på en mer objektiv beskrivning av hur det förhåller sig. Till exempel kan lärare uppfatta att likvärdigheten i bedömningen på en skola har ökat utan att man faktiskt har belägg för att det också förhåller sig på detta vis. (Thornberg, 2013, s. 23; vår kursivering)*

I några få studier menar man att samstämmigheten i lärarnas bedömningar ökat efter att lärarna deltagit i sambedömning, men hur dessa effekter dokumenterats är ofta oklart (de Eca & Torres, 2005; Falk & Ort, 1998; Syverson, 2009) och det finns dessutom studier där man inte kunnat påvisa några effekter alls på överensstämmelsen (Baird, Greatorex & Bell, 2004; Morgan & Watson, 2002). Att ökad samstämmighet är en *trolig* effekt av sambedömning, är däremot något som lyfts fram i flera olika studier (t.ex. Adie, Klenowski & Wyatt-Smith, 2012; Clarke & Gipps, 2000; se också Skolverket, 2009b), eftersom lärarna ofta – och i hög utsträckning – *upplever* att överensstämmelsen i bedömningen ökar (Black et al., 2011; Bolt, 2011; Connolly, Klenowski & Wyatt-Smith, 2012). Adie et al. (2012) menar därför att lärarnas upplevelser måste verifieras med andra typer av studier än de som är genomförda.

Även en svensk studie, genomförd av Gunilla Olofsson (2006), visar på svårigheten att med hjälp av sambedömning åstadkomma högre samstämmighet. I Olofssons studie bedömde fyra lärare elevprestationer från det nationella ämnesprovet i matematik, först enskilt och därefter ännu en gång efter att ha genomfört sambedömning där emellan. Resultatet visar att lärarna efter sambedömning visserligen ändrade sina bedömningar i viss utsträckning, så att de blev mer samstämmiga, men också att lärarna var olika benägna att revidera sina bedömningar, bland annat beroende på vilka kriterier de hade skilda uppfattningar kring.

Sammantaget kan man alltså se att det finns en generell problematik i att lärares bedömningar av komplexa uppgifter ofta skiljer sig åt, det vill säga de har en låg grad av samstämmighet, vilken ofta kvarstår även om lärare har tillgång till detaljerade bedömningsanvisningar. Huruvida sambedömning kan bidra till att öka samstämmigheten är inte klarlagt i forskningen, men forskningen tyder på att problemet dels kan bero på att lärare saknar gemensam tolkning av styrdokumentet och hur elevprestationer ska bedömas, dels på att lärare kan sätta olika omdömen även om de är överens om exempelvis vilka kriterier som ska tillämpas i bedömningen. För att samstämmigheten inte enbart ska bli ”likvärdig på ytan”, det vill säga att elever kan få samma poäng eller betyg men på olika grunder, krävs att lärares bedömningar baseras i en samsyn kring styrdokument och bedömning.

### **Sambedömning för samsyn**

Även om det är tveksamt i vilken utsträckning sambedömning förmår bidra till ökad samstämmighet, finns det tydligare belägg för att sambedömning kan bidra till ökad samsyn. I forskningen om sambedömning framgår att lärare, genom att analysera och diskutera konkreta elevprestationer tillsammans, kan nå en gemensam förståelse för hur skolans styrdokument kan tolkas och tillämpas (se t.ex. Connolly et al., 2012; Falk & Ort, 1998). Detta tycks bland annat bero på att lärares olika sätt att uppfatta kriterier och kravnivåer blir synliggjorda i sådana diskussioner. I studier genomförda i Queensland, Australien, har det till exempel visat sig att när lärare har olika uppfattningar om på vilken nivå en elevprestation ska placeras, uppstår det diskussioner som ger goda möjligheter till lärande för gruppen som sambedömer. Det tycks nämligen vara just i förhållande till dessa skilda uppfattningar, som förhandlingar måste ske och resonemang fördjupas, vilket resulterar i att kvalitetsskillnader mellan elevprestationer uppmärksammas och en gemensam förståelse kan komma till stånd (Adie et al., 2012; Klenowski & Adie, 2009). Ett sätt att locka fram skilda uppfattningar, är genom att välja elevprestationer på ett sådant sätt att de representerar en stor bredd av prestationer och elever med olika bakgrund (Falk & Ort, 1998).



Lenore Adie och hennes kollegor (2012), som bedrivit omfattande forskning kring sambedömning, drar därför slutsatsen att sambedömningen är viktig, eftersom den synliggör skillnader och tvingar deltagarna att förklara och rättfärdiga sina bedömningar. Liknande slutsatser dras även i andra studier och Paul Black, Christine Harrison, Jeremy Hodgen, Bethan Marshall och Natasha Serret (2011) menar att det är de utmanande diskussionerna som utgör fundamentet i sambedömningen, genom att lyfta fram och synliggöra frågor kring kravnivåer och kvaliteter i elevers prestationer.

### **Sambedömning som kompetensutveckling**

Sambedömning används inte enbart som ett sätt att öka likvärdigheten i bedömning och betygsättning, utan även som ett redskap för kompetensutveckling. Wiliam (2007) argumenterar för att lärare behöver förändra sitt sätt att arbeta med bedömning, från en traditionell och summativt orienterad bedömningspraktik till en formativ dito, eftersom det finns belägg för att en formativ bedömningspraktik kan ge stora positiva effekter på elevernas prestationer i skolan. En sådan förändring av lärarnas bedömningspraktik kan förväntas vara svår att åstadkomma, då det handlar om att ändra starkt etablerade sätt att tänka och arbeta. Wiliam menar dock att TLCs är särskilt lämpade för just denna typ av professionell utveckling. Antagandet får stöd bland annat från en forskningsöversikt om metoder för lärares kompetensutveckling, där Hilda Borko (2004) drar slutsatsen att så kallade ”strong professional learning communities” kan ge goda effekter på lärares undervisning samt bidra till reformarbete. Hon uppmärksammar dock även att en avgörande förutsättning för att den här typen av professionella grupper ska fungera, är att man lyckas skapa ett klimat i gruppen som är såväl stödjande som utmanande. För även om lärare vanligen välkomnar möjligheten att byta erfarenheter med kollegor, är kritiskt ifrågasättande diskussioner inte alltid lika populära.

Även om han inte använder samma terminologi, påminner Wiliams (2007) TLCs mycket om sambedömningsgrupper, genom att lärare i båda fallen träffas regelbundet för att diskutera bedömning tillsammans (se också Smith, 2012). Som beskrivs ovan, finns det också överlappning i slutsatserna kring hur sambedömning kan stödja ökad samsyn genom att olika åsikter kommer fram i diskussioner om elevprestationer och i hur avgörande kritiskt ifrågasättande diskussioner är för fungerande TLCs. Ett antal studier om sambedömning har också dokumenterat positiva effekter på lärarnas bedömningskompetens (t.ex. Black et al., 2010; 2011; Falk & Ort, 1998).



### **Sambedömning för formativ bedömning**

Att lärarnas bedömningskompetens ökar till följd av medverkan i sambedömning är inte nödvändigtvis samma sak som att de faktiskt förändrar sin undervisningspraktik. Det finns dock ett antal studier som pekar på att ett ökat fokus på formativ bedömning i de professionella diskussionerna faktiskt kan få ett genomslag i lärarnas undervisning, så att de i högre grad arbetar med formativ bedömning i klassrummet (Black, Harrison, Lee, Marshall & Wiliam, 2003, 2004; Kirton, Hallam, Peffers, Robertson & Stobart, 2007; Smith, 2011). Frågan är därför hur man kan få ett ökat fokus på formativ bedömning i sambedömningen.

När lärare sambedömer ges möjlighet till diskussioner kring frågor kopplade till såväl formativ som summativ bedömning, men fokus kan förskjutas i endera riktningen beroende på vad som bedöms och vad man lägger vikt vid i diskussionen (Black et al., 2011; Reid, 2007). Om det till exempel är slutprodukter som bedöms, tenderar bedömningen och samtalen att få en mer summativ och bakåtblickande karaktär. Om det däremot är ”work in progress”, till exempel ett första utkast till en text, kan diskussionen få en mer formativ och framåtsyftande prägel (Reid, 2007). Beverly Falk och Suzanna Ort (1998) menar dessutom att när elevarbetena är valda på ett sätt så att de representerar en stor bredd av prestationer och elever med olika bakgrund, är det lättare för lärarna att skapa sig en nyanserad uppfattning om kvaliteten i elevernas prestationer, vilket är en grundläggande förutsättning för formativ bedömning och återkoppling.

Ett annat sätt att öka fokus på den formativa bedömningen är genom att i diskussionen lägga vikt vid hur elevens prestationer svarar mot undervisningens mål och hur denna information kan användas för att planera och genomföra undervisning på ett sätt som bättre svarar mot elevernas behov (Bolt, 2011; Falk & Ort, 1998; Malone et al., 2004).

### **DISKUSSION**

Som beskrevs inledningsvis, har ett antal skolhuvudmän introducerat sambedömning som ett sätt att öka lärarnas bedömningskompetens och elevernas måluppfyllelse. Parallellt med detta har sambedömning förts fram som en strategi för att få lärarnas bedömning av nationella prov och betygssättning mer likvärdig. Frågan som ställts i denna artikel är i vilken utsträckning dessa skilda syften kan rymmas inom ramen för samma arbetssätt eller vilka konflikter detta kan innebära.

Om vi börjar med idén om att med hjälp av sambedömning öka samstämmigheten i lärarnas bedömning, kan man se att det i stor utsträckning saknas forskning som stödjer detta antagande. Det behövs således uppenbart fler studier, där man följer upp lärarnas bedömningar efter att de medverkat i sambedömning. Som påtalades ovan, finns det emellertid anledning att misstänka att man med ett fokus på att lärares bedömningar överensstämmer (motsvarande de studier som genomförts av interbedömaröverensstämmelse mellan lärares bedömningar av elevprestationer på nationella prov) riskerar att missa på vilka grunder som bedömningen gjorts.

Ett sätt att åstadkomma en samstämmighet i lärares bedömningar, vore att låta ett antal lärare bedöma alla elevprestationer oberoende av varandra, varefter deras bedömningar antingen sammanvägdes mekaniskt (t.ex. genom att beräkna ett medelvärde) eller genom att lärarna fick resonera sig fram till ett konsensusbeslut. Förutom att denna lösning vore ytterst tids- och kostnadskrävande, finns det även andra nackdelar. Att beräkna ett medelvärde innebär till exempel i princip att man accepterar att lärarna bedömer på olika grunder och att bedömningen utgör ett konglomerat av olika tolkningar och värderingar, där ytterligheterna förhoppningsvis tar ut varandra. Även konsensusbeslut kan vara problematiska, eftersom flera studier pekar på att det är de situationer när lärarna är oeniga kring hur en prestation ska bedömas som erbjuder de största möjligheterna till lärande för de som medverkar. Om målet med sambedömningen är att komma överens, snarare än att fritt lyfta fram olika uppfattningar, riskerar således spridningen i uppfattningar att bli ett hinder istället för en tillgång. Enskilda lärares status i förhållande till varandra kan också få stor inverkan på processen. Pamela Moss och Aaron Schutz (2001) hävdar att:

*/.../ dissensus is an essential natural resource that should be acknowledged and nurtured along-side the search for consensus or agreement. Although it does not ensure a fair and inclusive dialogue, the representation and exploration of dissensus helps to protect us from the false assurance of an articulated consensus that may underrepresent, misrepresent, or exclude groups of voices within the community (s. 65)*

Ett annat, men mer långsiktigt, sätt att uppnå ökad samstämmighet, vore att gå via samsyn. Genom att med stöd av sambedömning stärka lärarnas gemensamma tolkning av styrdokument och syn på bedömning kan eventuellt samstämmigheten öka på sikt. Även om det saknas tydliga belägg för detta antagande i forskningslitteraturen, finns det rapporter från betygssystem som i hög grad utnyttjar sambedömning som kvalitetssäkring, där

samstämmigheten i lärarnas betygsättning är förvånansvärt hög. I Queensland, Australien, rapporterar man till exempel en överensstämmelse på 92 procent mellan lärare och granskare vid en granskning av slumpmässigt urval av 2649 elevportföljer (QSA, 2012).

Enligt ovanstående resonemang, skulle det inte nödvändigtvis innebära en konflikt mellan olika syften, om arbete med sambedömning för att uppnå samsyn på sikt förväntades leda även till ökad samstämmighet. Ytterligare en faktor att ta i beaktande, är emellertid vad som bedöms i sambedömningen. Till exempel utgår propåerna om sambedömning inte sällan från att likvärdigheten i bedömningen av de nationella proven och i betygen ska öka. Men är det tänkbart att sambedömningen utgår från elevers prestationer på nationella proven, eller att inkludera betygsättningen i sambedömningen, utan att detta kommer i konflikt med sambedömningens andra syften?

### **Sambedömning av nationella prov**

De nationella proven i Sverige är standardiserade prov som för närvarande ges i 11 ämnen i grundskolans årskurs 6 och 9<sup>2</sup>. Årskurs 3 har något färre prov och gymnasieskolan har ett antal olika kursprov i svenska, engelska och matematik. Att proven är standardiserade innebär att de ser likadana ut och ska genomföras på samma sätt av alla elever, för att på så sätt ge förutsättningar att jämföra resultaten. De flesta prov ges en gång om året och elevernas svar bedöms oftast av den lärare som undervisat eleverna, även om det förekommer att man exempelvis byter provsvar med varandra inom skolan, så att lärarna bedömer svar från andra klasser än sin egen. Som stöd för bedömningen finns facit och bedömningsanvisningar, vilka kompletteras med autentiska elevsvar för att exemplifiera nivåerna i bedömningen av komplexa uppgifter, som uppsatser eller uppgifter där eleven förväntas föra utvecklade resonemang. Anvisningarna kan därmed vara mycket omfattande, ifall proven innehåller flera komplexa uppgifter. Elevernas resultat sammanställs i de flesta fall i form av delprovs- eller provbetyg.

Att alla elever i en viss årskurs ska skriva samma prov vid samma tillfälle, samt att lärarna förväntas bedöma provsvaren på ett likvärdigt sätt, ställer vissa krav på de nationella proven, vilket bland annat innebär att proven huvudsakligen är skriftliga. Det förekommer visserligen muntliga och laborativa uppgifter, men det rör sig i de flesta fall om enstaka uppgifter. I de flesta fall är också andelen korta, lättbedömda uppgifter relativt stor (dvs. 50% eller mer i förhållande till det totala antalet uppgifter). Skäl till att ha många sådana uppgifter är exempelvis för att hinna pröva ett bredare utsnitt av ämnes-

---

2 Eleverna skriver dock ”bara” prov i fem ämnen, eftersom de endast skriver i ett NO-ämne, ett SO-ämne samt antingen svenska eller svenska som andraspråk.

innehåll under den begränsade tiden, för att lätta lärarnas bedömningsbörda samt ibland även för att göra provet som helhet mer mätsäkert.

Dessa egenskaper hos de nationella proven påverkar även sambedömningen, om sambedömningen skulle utgå från elevernas svar på de nationella proven. Till exempel genomförs de nationella proven endast en gång om året, vilket innebär att sambedömningen skulle bli mer sporadisk än kontinuerlig om man enbart använde nationella prov som underlag för sambedömningen. Att sådan sporadisk sambedömning förekommer, framgår till exempel av Skolverkets stödmaterial för sambedömning (Skolverket, 2014), som bygger på intervjuer med lärare och skolledare på skolor som arbetar aktivt med sambedömning.

Men det finns även andra, mindre uppenbara, svårigheter med att utgå från de nationella proven, däribland det begränsade uppgiftsformatet. Som framgick i forskningssammanfattningen om sambedömning ovan, är ett av sambedömningens fundament att skilda uppfattningar behöver komma fram i diskussionen, vilket underlättas av en bredd i uppgiftsformat. De nationella provens begränsade format kan därför motverka lärandepotentialen i sambedömningen, samtidigt som det är just sådana uppgifter som *inte* ryms i provformatet som är svårast att bedöma och därmed bäst skulle behöva diskuteras.

Ytterligare en viktig aspekt i relationen mellan sambedömning och nationella prov, är att de nationella proven företrädesvis har en summativ funktion. Provens huvudsyfte är att stödja en likvärdig och rättvis bedömning och betygssättning samt att ge underlag för analys av i vilken utsträckning kunskapskraven uppfylls på olika nivåer i skolsystemet. Proven är således inte primärt konstruerade för att ge den nyanserade information om elevernas kunskaper, som behövs för att exempelvis ge konstruktiv återkoppling till eleverna. Att utgå från de nationella proven kan därför göra det svårare för lärarna att inkludera formativ bedömning i diskussionen. Att inkludera ännu icke färdiga uppgifter i underlaget, vilket enligt forskningssammanställningen ovan är ett sätt att föra in formativ bedömning i sambedömningen, är inte heller möjligt med nationella prov.

### **Sambedömning för bedömning eller betygsättning?**

I de nationella proven finns det oftast tydliga regler för hur poäng eller delprovsresultat ska sammanvägas till ett provbetyg. Däremot finns det inga särskilda riktlinjer för hur provbetyget ska relateras till termins-, slut- eller kursbetyg, mer än att provresultatet ska vara rådgivande snarare än styrande. På liknande sätt finns det få riktlinjer om hur lärarnas betygsättning förväntas gå till och bland annat Helena Korp (2006) har identifierat flera olika strategier för lärares betygsättning. Betygsättningen skulle därför, på motsvarande

sätt som bedömningen av prestationer på enskilda uppgifter, kunna utgöra grund för arbete med sambedömning. Även i förhållande till Allals (2013) modell, som beskrevs tidigare (Figur 1), skulle betygsättningen behöva inkluderas i arbetet med sambedömningen om man vill uppnå ökad likvärdighet i betygen.

I likhet med de nationella proven, har betygen en företrädesvis summativ funktion, även om deras motiverande funktion lyfts fram i vissa (framför allt skolpolitiska) sammanhang. Även i relation till betygen är det därför av intresse att ställa sambedömningens olika syften mot varandra. Det man då kan konstatera är att betygen, också i likhet med de nationella proven, ges sporadiskt, vilket därmed inte ger underlag för en kontinuerlig sambedömning om sambedömningen enbart utgår från betygsättningen. Däremot kan betygsättningen lättare än bedömningen av de nationella proven knytas till formativ bedömning, till exempel genom att diskutera elevers samlade prestationer i en sambedömningsgrupp inför utvecklingssamtal eller motsvarande.

#### SLUTSATSER

Det finns en stark tilltro till sambedömningens förmåga att bidra till ökad likvärdighet i lärarnas bedömning och betygsättning, inte minst genom att lärarna bedömer elevernas prestationer från de nationella proven tillsammans. Som denna artikel visat, riskerar detta synsätt att vilseleda såväl elever som professionella i skolan genom en överdriven tilltro till vad som egentligen kan åstadkommas med sådan sambedömning, till exempel i termer av exakt överensstämmelse mellan olika bedömare. Genom att inskränka sambedömningen till att endast omfatta nationella prov, riskerar man även att begränsa lärarnas möjligheter att utnyttja sambedömningens pedagogiska potential. Om man däremot låter sambedömningen bli en kontinuerlig verksamhet, som omfattar såväl tolkning av styrdokument, bedömning av de egna elevers prestationer (både färdiga och ännu inte färdiga produkter) som nationella prov och betygsättningen, tycks det enligt ovanstående diskussion finnas goda förutsättningar för att sambedömningen kan bidra till att ge stöd och styrka åt lärarnas bedömningspraktik.

## REFERENSER

- Adie, L. E. Klenowski, V. & Wyatt-Smith, C. (2012). Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*, 64(2), 223–240.
- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20–34.
- Baird, J.-A., Greatorex, J. & Bell, J. F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in Education: Principles, Policy & Practice*, 11(3), 331–348.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215–232.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451–469.
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Berkshire, UK: Open University Press.
- Black, P. J., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9–21.
- Bolt, S. (2011). Making consistent judgments: Assessing student attainment of systemic achievement targets. *Educational Forum*, 75(2), 157–172.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15.
- Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research & Evaluation*, 16(17), 1–12.
- Clarke, S. & Gipps, C. (2000). The role of teachers in teacher assessment in England 1996–1998. *Evaluation and Research in Education*, 14(1), 38–52.
- Connolly, S., Klenowski, V. & Wyatt-Smith, C. M. (2012). Moderation and consistency of teacher judgement: Teachers' views. *British Educational Research Journal*, 38(4), 593–614.

- Davison, C. (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, P. & Torres, M. T. (2005). Using portfolios for external assessment: An experiment in Portugal. *International Journal of Art & Design Education*, 24(2), 209–218.
- Englund, T. & Quennerstedt, A. (2008). *Vadå likvärdighet? Studier i utbildningspolitisk språkanvändning*. Göteborg: Daidalos.
- Erickson, G. (2009). *Nationella prov i engelska – en studie av bedömersamstämmighet*. Hämtad 10 april 2013, från <http://www.nafs.gu.se/publikationer/>
- Falk, B. & Ort, S. (1998). Sitting down to score: Teacher learning through assessment. *Phi Delta Kappan*, 80(1), 59–64.
- Jönsson, A. (2014). *Lärande bedömning* (3:e uppl.). Malmö: Gleerups.
- Jönsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- 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.
- Klenowski, V. & Adie, L. (2009). Moderation as judgement practice: Reconciling system level accountability and local level practice. *Curriculum Perspectives*, 29(1), 10–28.
- Korp, H. (2006). *Lika chanser i gymnasiet? En studie om betyg, nationella prov och social reproduktion*. Doktorsavhandling, Malmö högskola.
- Lie, S. (2007). Evalueringen av de nasjonale prøvene og hva vi kan lære av dem. I S. Tveit (red.), *Elevvurdering i skolen. Grunnlag for kulturendring* (ss. 85–92). Oslo: Universitetsforlaget.
- Malone, L., De Lucchi, L. & Long, K. (2004). All things in moderation. *Science and Children*, 41(5), 30–34.
- Morgan, C. & Watson, A. (2002). The interpretative nature of teachers' assessments of students' mathematics: Issues for equity. *Journal for Research in Mathematics Education*, 33(2), 78–110.



- Moss, P. A. & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38(1), 37–70.
- Olofsson, G. (2006). *En studie av lärares bedömning av elevarbeten på ett nationellt prov i matematik kurs A*. Stockholm: Stockholms universitet, PRIM-gruppen.
- Parkes, J. (2013). Reliability in classroom assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 107–123). Thousand Oaks, CA: SAGE Publications.
- QSA, Queensland Studies Authority. (2012). *Random sampling project 2012. Report on random sampling of assessment in Authority subjects*. Brisbane: Queensland Studies Authority.
- Reid, L. (2007). Teachers talking about writing assessment: valuable professional learning. *Improving Schools*, 10(2), 132–149.
- Sadler, D. R. (1998). Formative Assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77–84.
- 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 (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 (2014). *Sambedömning i skolan – exempel och forskning*. Stockholm: Skolverket.
- Smith, K. (2011). Professional development of teachers – A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation*, 37(1), 55–61.
- Smith, C. (2012). Why should we bother with assessment moderation? *Nurse Education Today*, 32(6), 45–48.
- Starch, D. & Elliot, E. (1912). Reliability in grading high school work in English. *School Review*, 20(7), 442–457.

- Syverson, M. A. (2009). Social justice and evidence-based assessment with the Learning Record. *Forum on Public Policy*, 1, 1–27.
- Thornberg, P. (2014). *Kan sambedömning leda till en mer likvärdig bedömning och betygssättning?* Masteruppsats, Högskolan Kristianstad.
- Tolgfors, B. (2014). *Bedömning för lärande i idrott och hälsa. Didaktiska konsekvenser av styrning.* Licentiatuppsats, Örebro Universitet.
- William, D. (2007). Content then process: Teacher learning communities in the service of formative assessment. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 182–204). Bloomington, IN: Solution Tree.