

Att få sina bedömningar granskade

*Frida Wetterstrand, Marcus Sundhäll &
Christian Lundahl*

TO HAVE YOUR ASSESSMENTS ASSESSED. Since 1940 there has been different types of national tests in Sweden, each with a different aim and purpose. Since 2011, Skolverket/The National Agency for Education is required to review how teachers correct these national tests, and also to publish easily accessible statistics showing possible discrepancies between the results of the students and their final grades. Based on interviews with 17 teachers in mathematics between grades 6-9, the consequences of the review of teachers' assessments have been investigated. The result shows worrying effects. The increased focus on discrepancies has led to an inadequate appraisal of teachers' assessment competence. However, teachers are pleased to have their corrections reviewed in a formative way.

Keywords: Nationella prov, betyg, likvärdig bedömning, matematik.

Inledning

Det svenska systemet för betygssättning är relativt unikt i världen genom att lärares betyg på elevernas kunskaper har en avgörande roll för intagningen till gymnasiet och högskolan (se vidare Lundahl, Hultén & Tveit 2017, Waldow 2013). Tilliten till lärares bedömningar är dock inte självklar och sedan 1940-talet har därför lärares betyg legitimerats genom att de på olika sätt har relaterats till centralt utformade prov. Initialt kallades dessa standardprov i grundskolan och centralprov i gymnasieskolan. Dessa prov genererade en nationell norm

Frida Wetterstrand är adjunkt i matematikdidaktik vid Örebro universitet, 701 82 Örebro. E-post: frida.wetterstrand@oru.se
Marcus Sundhäll är lektor i matematik vid Örebro universitet, 701 82 Örebro. E-post: marcus.sundhall@oru.se
Christian Lundahl är professor i pedagogik vid Örebro universitet, 701 82 Örebro. E-post: christian.lundahl@oru.se

vilken läraren skulle förhålla sig till i sin betygssättning. Gymnasiets lärare var tvungna att motivera inför länskolnämnden om klassens betyg avvek med mer än 0.2 standardavvikelser från denna norm. I grundskolan fanns inga sådana riktlinjer, även om idealet var att inte avvika för mycket från normen. I mitten på 1990-talet, i samband med införande av målrelaterade betyg, fick de centralt utformade proven en annan roll. De skulle vara ett stöd för lärarna i deras betygssättning och fungera som ett bedömningsunderlag bland andra. Det var upp till läraren att avgöra vilken betydelse de skulle ha för betygen. Det var vid mitten av 1990-talet faktiskt inte ens självklart att proven skulle finnas kvar i det nya betygssystemet (se vidare Lundahl 2009). Sedan dess har proven fått en allt mer styrande roll. I 2011 års regleringsbrev ålades Skolverket att granska likvärdigheten i lärares bedömningar genom att årligen med statistik synliggöra diskrepanser mellan resultat på nationella proven och elevers termins- eller slutbetyg. Skolverket skulle även granska lärares rättning av nationella proven (Regeringen 2011). Hur upplevs detta av lärarna? I denna artikel undersöks hur lärare i matematik säger sig ha blivit påverkade av att likvärdigheten i deras bedömningar granskas. Utvecklingen mot mer styrande externa prov är inte något som enbart skett i Sverige, vilket vi kommer att visa, men de nationella provens historia och de funktioner de tidigare haft är relativt unikt. Vi menar att lösningen med prov som ska reglera betygssättningen är problematisk då externa prov och internt satta betyg delvis mäter olika saker och delvis tjänar olika syften. Denna kluvenhet i provens syften är något som lärare har att hantera.

Artikeln har sin grund i ett större projekt med syfte att kartlägga effekterna av de nationella proven i årskurs 6 och 9 (Bonnevier et al. 2016). I den större undersökningen studeras speciellt hur förberedelser inför proven ser ut, hur proven utförs och hur resultaten av nationella prov används utifrån elev- lärar- och skolledarperspektiv i samtliga ämnen som har prov. Frågor kring hur provresultaten används visade sig engagera lärarna i matematik i mycket stor utsträckning, vilket är en anledning till att detta fördjupas i relation till betygssättning i den här artikeln.¹

Bakgrund, syfte och frågeställningar

Bedömning av elevers kunskaper har historiskt sett haft olika funktioner, beroende på skolans roll i samhället och synsätt på utbildning vid olika tidpunkter. Sett till historien har bedömning varit ett verktyg för att mäta elevernas begåvning och intelligens, det vill säga statistiska egenskaper hos elever (Korp 2011). Ett syfte med bedömning har sedan

långt tillbaka varit att rangordna eleverna (Pettersson 2010), men en förändring skedde med start under 1980-talet gällande synen på bedömning, vilket lett fram till att vi idag också ser bedömning som en viktig del i undervisningen och lärandet (Lundahl 2006, Korp 2011).

Från en mättradition där elevers slutprodukt av lärprocessen bedömts och kontrollerats har den svenska skolan gått mot att snarare mäta och jämföra elevernas kunskaper för att främja lärande hos eleverna (Korp 2011, Lindström 2005, Lundahl 2009, Pettersson 2010). Formativ bedömning har lyfts fram som redskap för att forma elevernas pågående lärande och undervisning (Lundahl 2011). Summativ bedömning används för att summera elevernas kunskaper, exempelvis i form av betyg. I båda fallen gör lärarna någon typ av bedömning. Bedömningar är således en stor del av dagens läraryrke och förmågan att precis kunna bedöma elevers prestationer formativt såväl som summativt är en viktig aspekt av lärarens profession (Ready & Wright 2011).

1994 infördes dagens nationella prov i den svenska skolan, och de har sedan dess haft som ett av flera syften att stödja lärare i deras bedömningar av elevers kunskaper (Lundahl 2009). 2011 tillkom ett granskande syfte, då Skolverket ålades av regeringen att granska likvärdigheten i lärarnas summativa bedömningar. Detta sker genom att Skolinspektionen rättar uppgifter på de nationella proven som sedan jämförs med lärarnas rättning, och eventuella avvikelser redovisas av Skolverket (Regeringen 2011). Skolverket ålades ytterligare en granskande funktion, att årligen publicera lättillgänglig statistik kring diskrepanser mellan betyg vid nationella proven och termins-/slutbetyg. De ska även analysera diskrepanserna och redovisa sina analyser till Regeringskansliet (Skolverket 2011).

Syftet med artikeln är att undersöka hur matematiklärare och deras arbete, enligt dem själva, påverkas av Skolverkets granskningar som beskrivits ovan. Detta undersöks utifrån frågeställningarna:

- Vilka konsekvenser för matematiklärarens bedömningsarbete har Skolverkets årliga statistik kring diskrepanser mellan resultat på nationella proven och elevers termins-/slutbetyg fått?
- Vilka konsekvenser för matematiklärarens bedömningsarbete har den granskning av lärarens rättning av de nationella proven som Skolverket i samarbete med Skolinspektionen genomfört fått?

Teoretiska utgångspunkter

Internationell kontext: externa bedömningar och/eller interna bedömningar som mått på likvärdighet och kvalitet

Bedömning av elevers kunskaper kan göras på olika sätt. Ett sätt att beskriva olika slags bedömningar är att skilja mellan interna och externa bedömningar (Forsberg & Lundahl 2009, Lundahl 2010). I Sverige har externa test förekommit sedan 1940-talet då centralt utarbetade prov infördes. Dessa har haft olika utformning, men den nuvarande formen infördes 1994 och kallas med ett gemensamt samlingsnamn för nationella prov (Lundahl 2009). De genomförs årligen i den svenska grund- och gymnasieskolan. Proven har i grundskolan oftast varit frivilliga för lärarna att genomföra även om de flesta genomfört dem. Från 2009 och framåt har dock fler och fler prov blivit obligatoriska, både i grundskola och gymnasium. De nationella proven har flera olika syften. Med fokus på hur de ska användas går det dock historiskt sett att tala om två övergripande syften. Det första handlar om att stimulera en ämnesdidaktisk utveckling bland lärarna och på skolorna, medan det andra syftet är att ge svar på frågor av nationellt intresse, som kvalitet och likvärdighet i undervisningen. I samband med den senaste läroplansreformen 2011 tillkom även ett tredje syfte; att granska likvärdigheten i lärares bedömningar (Regeringen 2011). Utvecklingen i Sverige har gått mot att använda de nationella proven för styrning av kvalitet och likvärdighet, vilket följer en internationell trend. Vid mätning och styrning av kvaliteten på utbildningen, har elevers individuella studieresultat kommit att spela en allt viktigare roll under de senaste decennierna (Hopmann 2003, Sahlberg 2016, Lundahl, Hultén & Tveit 2017a). I många länder har policy som följer på denna utveckling förlitat sig på en ökad användning av tester för skolans resultatredovisning (Baker 2016, Brookhart 2015, Linn 2000). Pasi Sahlberg 2016 har kallat utvecklingen mot att använda externa bedömningar för kvalitetsgranskning av utbildningssystem för GERM: ”a Global Educational Reform Movement”, en rörelse där alltså kunskapsmätningar utgör kärnan. Det som gör Sverige intressant i sammanhanget är den långa traditionen av att använda betyg som mått på likvärdighet och kvalitet, det vill säga lärarnas interna bedömningar. De nationella proven har så att säga syftat till att garantera kvaliteten i detta mått. I grunden finns alltså en tillit till lärares bedömningar som kvalitetsmått, men det är villkorad tillit. Således är det intressant att studera hur lärare ställer sig till att deras betyg och bedömningar ytterligare ska knytas till externa mätningar som en följd av 2011 års reform.

Användningen av betyg i resultatbaserade ansvarssystem (accountability systems) har inte dragit till sig särskilt mycket uppmärksamhet i forskning, i stället har fokus legat på prov- och testbaserat ansvarstagande och dess effekter (t.ex. Au 2011, Baker 2016, Dee & Jacob 2011, Harlen & Deakin Crick 2002). Denna forskning kännetecknas ofta av att ha ett elev- eller lärandeperspektiv där lärarnas syn på prov och bedömning negligerats (Mickwitz 2015). Baksidorna med prov- och testbaserade ansvarssystem har lett flera forskare till att argumentera för alternativ. Douglas Harris och Carolyn Herrington (2006) hävdar i en kommentar till den delvis misslyckade amerikanska reformen No Child Left Behind (NCLB), som byggde mycket på användningen av externa test för att utkräva ansvar, att det finns ett behov av att skifta fokus från mätning av resultat tillbaka till frågor om resurser och innehåll. Vi har nyligen också kunna se den typen av policy växa fram i USA (t.ex. Darling-Hammond, Wilhoit & Pittenger 2014).

I argumentationen för alternativ till ett prov- och testbaserat ansvarssystem, har internationella forskare föreslagit att man kan använda de bedömningar lärare gör för att undvika den stress som uppstår i system med mycket tester (Brookhart et al 2016, Willingham, Pollack & Lewis 2002). Problemet är dock att betyg har betraktats som alltför opålitliga för att användas för administrativa ändamål (Brookhart, 2015, Brookhart et al. 2016). Den diskussion som internationellt sett relativt nyligen uppstått kring betygens hållbarhet, liknar den vi haft i Sverige sedan 1940-talet. Det som dock händer i den internationella forskningen är att man ganska snart slår fast att betygsatta prestationen inte ska anses vara mått på samma slags kunskaper som prövas med ett test (Brookhart 2015). Istället bör betyg och test ses som kompletterande mätningar, och de kan därför möjligen fylla kompletterande funktioner i styrning och mätningen av kvalitet och likvärdighet i utbildningen (Brookhart 2015, Brookhart et al. 2016, Südkamp, Kaiser & Möller 2012). Betyg verkar till exempel vara bättre på att förutsäga framtida framgång i utbildning, och är delvis ett mått på så kallade *conative skills* – såsom elevernas intresse, vilja och självreglering (Brookhart 2015, Brookhart et al. 2016). Men även om det kan finnas ett värde i att använda betyg i utvärderingar för administrativa ändamål, är det inte helt oproblematiskt. En del forskare menar att betygens kvaliteter bygger på att de används internt för pedagogiska syften, som ett kontrakt mellan elever och lärare (Willingham et al. 2002). Betyg och externa tester tycks med andra ord mäta olika saker. Är det då rimligt att använda nationella prov för att öka likvärdigheten i lärares betygssättning?

I artikeln *Grades and Test Scores* (Willingham et al. 2002) ställs den enkla frågan: varför skiljer sig lärares interna bedömningar och

resultaten på externa test åt? Författarna listar fem faktorer som bidrar till att test och betyg kan skilja sig åt: *subject covered; grading variations; reliability; student characteristics; teacher ratings*. Det som hade störst betydelse för skillnaderna mellan betyg och testresultat var skillnader i betygspraktik mellan skolor, avvikelser från kursinnehåll samt elevernas engagemang och attityder till skolan. Med andra ord berodde diskrepansen mellan lärares bedömning och provresultat på lokala variationer i värderingar av elevers kunskaper, skillnader mellan vad man undervisat om och vad som provas samt slutligen, vilket är ganska självklart, på elevernas studieinsatser. Resultatet visar att dessa faktorer sammantaget förklarar upp mot 80 procent av variationen mellan elevens betyg och testresultat. Det innebär likväl att det är svårt att använda det ena resultatet för att predicera det andra på individnivå, det vill säga om det är provet som inte mäter det lärarna gjort eller lärarna som inte värderar elevernas kunskaper rätt. Därför drar forskarna slutsatsen att det är viktigt att förstå att testresultat och betyg ömsesidigt bör validera varandra. Skälen till det är för det första att betyg rymmer mycket komplexitet baserat på många observationer, för det andra att elevers prestationer kan variera ganska mycket mellan dessa olika observationstillfällen.

Warren Willingham et al (2002) föreslår att vi behöver mer information om eleverna och den lokala kontexten för att göra korrekta uppskattningar om relationen mellan lärares bedömningar och externa provresultat. När det gäller de svenska nationella proven provas dessa ut av provkonstruktörer tillsammans med lärare och på elever. Utprovning syftar till att så långt som möjligt tillgodose de mättekniska krav som finns på validitet genom att berörda parter (skolaktörer) tillsammans enas om vad som är kompetens inom ämnesområdet på den eller den nivån (det vill säga att de uppgifter som valts ut representerar ett avsett kunnande, så kallat *construct*). Däremot sker ingen lika systematisk kalibrering i efterhand för att se om tolkningarna är valida. I dagsläget kan provresultaten därmed lätt uppfattas som ett facit för vad skolorna borde ha gjort snarare än att bli en del i en intersubjektiv bedömning där resultaten sätts i relation till vad som faktiskt gjorts och under vilka villkor det gjorts.

Sett utifrån tidigare forskning om relationen mellan externa prov och lärares betygssättning, kan en konsekvens av de nya krav som ställs på de nationella proven bli att lärare kan uppleva en spänning kring provens delvis motstridiga syften. Ska proven stödja och kontrollera lärares bedömning? Problemet uppstår när lärares bedömningar skiljer sig från resultaten av externa bedömningar. Vem är det då som har gjort rätt? I länder där betyg och test hålls isär, vilket gäller i majoriteten av Europas länder (Lundahl, Hultén & Tveit 2017b), är detta

inte ett lika stort problem jämfört med i Sverige där proven används för att granska likvärdigheten i lärares betygssättning. Sannolikt är medvetenheten om eventuella diskrepanser mellan lärares betyg och externa bedömning högre i Sverige och eftersom det inte är självklart vad diskrepanserna beror på, var ”felet” ligger, blir det intressant att studera hur lärare ser på det.

Den svenska kontexten: externa bedömningar och/eller interna bedömningar i matematik

Att svenska lärares interna bedömningar skiljer sig från resultaten från externa test är inte ovanligt (Bonnevier et al. 2016). Skolverket (2007) analyserade under åren 1998-2006 om det fanns diskrepanser mellan resultaten på nationella proven och elevers slutbetyg åren 1998-2006 i ämnena svenska, engelska och matematik. Det visade sig att det fanns sådana skillnader och att de var störst i ämnet matematik. Analyser av dessa diskrepanser fortsatte därefter att uppdateras av Skolverket i form av årliga rapporter och sedan 2011 har Skolverket årligen publicerat lättillgänglig statistik över diskrepanser mellan provresultat och elevers termins-/slutbetyg i databasen SiRis.²

I den statistisk och i de analyser som publicerats sedan 1998 (se exempelvis Skolverket 2016a, Skolverket 2016b) är det tydligt att betyg på nationella provet skiljer sig från elevens termins-/slutbetyg i matematik. Sett till riket som helhet får elever generellt högre termins-/slutbetyg än betyg vid nationella provet i ämnet matematik (Skolverket 2016a, Skolverket 2016b).

Proven är avsedda att vara ett stöd för läraren vid betygssättning, men behöver inte vara avgörande för det termins-/eller slutbetyg eleverna erhåller. Diskrepanser dem emellan kan därför finnas på individnivå. Problemet med diskrepanser, som Skolverket (2007) ser det, uppstår om skillnaderna mellan betyg och resultat på nationella prov systematiskt återkommer inom eller mellan skolor från år till år. 2011 inledde Skolverket sina årliga granskningar av både lärares bedömning och diskrepanser, och eftersom dessa har visat sig vara större i matematik än andra ämnen är det intressant att särskilt studera matematiklärares upplevda konsekvenser av granskningen.

Metod

Datainsamling

Materialet till denna artikel är hämtat från en större intervjustudie, gjord av Örebro universitet 2015 på uppdrag av Skolverket (Bonnevier et al. 2016). Elever, lärare och skolledare intervjuades i gruppintervjuer. I den större studien genomfördes totalt 20 gruppintervjuer med 2-4 lärare per grupp, där 5 av grupperna bestod av enbart matematiklärare och där 7 grupper var blandgrupper som delvis bestod av matematiklärare men även andra ämneslärare. I den här artikeln står matematiklärarna i fokus och totalt har intervjutranskriptioner från 17 matematiklärare använts. Intervjuerna var semistrukturerade. Utgångspunkt för diskussionerna var frågor kring förberedelse, genomförande och resultat användning i samband med de nationella proven. Frågorna var av öppen karaktär för att fånga lärarnas berättelser och för att lärarna utifrån erfarenheter av de nationella proven skulle kunna vara med att utveckla diskussionerna. En styrka i semistrukturerade intervjuer är möjligheten för respondenten att utveckla och fördjupa tankar och idéer kring de ämnen som intervjuaren tar upp, så informationen blir djupgående och detaljerad (Denscombe 2013, Kvale & Brinkmann 2009). Då materialet redan var insamlat till den större studien har frågeguiden inte formats i relation till artikelns syfte.

Databearbetning

Utsagorna i den ursprungliga studien transkriberades och kategoriserades med hjälp av programmet NVivo i de på förhand bestämda kategorierna *från förberedelser till bedömning, provens utformning, bedömningsanvisningar, resultat användning och effekter* samt *formativ användning och dess effekter* (Bonnevier et al. 2016).

I den större studien gjordes en jämförelse mellan utsagor från lärare i olika ämnen, med fokus på ämnesvisa likheter och skillnader. Denna jämförelse visade att bedömningsanvisningar, återkoppling på den gjorda rättningen samt diskrepanser mellan resultat på nationella proven och elevernas termins-/slutbetyg fick särskilt stort utrymme i intervjuerna med matematiklärare. Matematiklärarna valde själva att diskutera detta även vid intervjufrågor som behandlade andra teman. Detta var alltså diskussioner som återkom genom alla intervjuer med matematiklärare, dels på direkta frågor från intervjuaren, men också på eget initiativ från lärarna. Ämnena engagerade således lärarna i matematik i särskilt stor utsträckning jämfört med andra ämneslärare vilket är anledningen till artikelns inriktning. Datamaterialet

bearbetades inledningsvis med öppet sinne utifrån frågan ”varför engagerar provanvändningen matematiklärarna i så hög utsträckning?”, och utsagorna lästes och tolkades i flera omgångar för att förstå varför lärarna ständigt valde att återkomma till ämnet. Då tolkningar av lärarnas utsagor gjorts för att förstå vad lärarna menar, exempelvis genom att läsa vad som föregått och vad som följde uttalandet, kan analysen sägas ha hermeneutiska inslag.

I en hermeneutisk tolkningsprocess tolkas och omtolkas förståelsen av en utsaga, då forskaren ständigt försöker förstå texternas delar i relation till dess helhet (Westlund 2009). Inledningsvis lästes och tolkades lärarnas utsagor individuellt av artikelns författare och därefter sammanställdes de individuella tolkningarna gemensamt. I de fall där tolkningarna inte stämt överens lästes materialet på nytt och diskuterades för att hitta en gemensam förståelse för vad det var lärarna gett uttryck för.

För att finna svar på den initiala frågan grupperades utsagor med liknande teman och på så vis kunde mönster, likheter och kontraster identifieras. En tydlig likhet var frustration över provresultatens användning vid betygssättning samt att lärarna på olika sätt upplevde sig granskade i sitt bedömningsarbete. Det blev tydligt att lärarna återkommande diskuterade *konsekvenser* av olika sätt att använda provresultatet som måttstock på elevernas kunskaper. Studiens forskningsfrågor formulerades därefter och återigen granskades datamaterialet för att samla utsagor kring konsekvenser av användandet av provresultatet som måttstock på elevernas kunskaper. Analysprocessen innefattade att gruppera likheter och skillnader i relation till forskningsfrågorna och att sätta lärarnas utsagor i relation till aktuell forskning om intern och extern bedömning. Den senare delen av analysprocessen syftade till att stärka trovärdigheten i de tolkningar som gjordes. Att datamaterialet genomlästes vid flera tillfällen och att artikelns författare skapade förståelse för materialet både individuellt och genom gemensamma diskussioner kan också antas stärka trovärdigheten i tolkningarna. Ytterligare ett led i att stärka studiens trovärdighet är att genomskinlighet eftersträvas genom användande av citat från intervjumaterialet som stöd för tolkningarna.

I resultatdelen återfinns direkta citat från datamaterialet för att förtydliga och ge belägg för att de tolkningar som gjorts och de resultat som framkommit är trovärdiga. Citaten är skrivna med skriftspråk och inte talspråk för att underlätta läsningen. Lärarna har tilldelats fingerade namn. Vi har använt samma namn som i den större studien (Bonnevier et al. 2016). Deltagarna har gett sin tillåtelse till att materialet får användas i forskningssyfte. Sju lärare är citerade i texten. Elisa,

Ellen och Anders har erfarenhet från nationella proven i årskurs 9, Emilia och Edvin har erfarenhet från nationella proven i årskurs 6 medan Bengt och Berit är har erfarenhet från nationella proven både i årskurs 6 och 9. De citat som används ska ses som exempel som återspeglar den allmänna diskursen bland studiens respondenter. I lärarnas citat har talspråk uttryckts med skriftspråk och ord som Öh, Äh och Mmm har tagits bort från ursprungscitaten.

Resultat

Redovisningen av studiens resultat struktureras i relation till respektive forskningsfråga. Detta innebär att vi först, under rubriken ”diskrepanser mellan resultat på nationella proven och termins-/slutbetyg”, kommer att återge lärares uppfattningar om den lättillgängliga statistik om diskrepanser som Skolverket ålagts att redovisa. Därefter, under rubriken ”Skolinspektionens och Skolverkets granskning av lärares rättning”, presenteras hur lärare upplever den granskning av deras rättning som Skolinspektionen är ålagda att göra i samband med nationella proven. Därefter sätts lärarnas utsagor i relation till forskning om relationen mellan interna och externa bedömningar.

Diskrepanser mellan resultat på nationella proven och termins-/slutbetyg

I matematiklärarnas utsagor är det tydligt att de påverkats av att diskrepanser mellan resultat på nationella proven och elevernas termins-/slutbetyg synliggjorts och granskats allt mer de senaste åren. Matematiklärarna i studien är medvetna om, och poängterar återkommande, att en avsikt med nationella proven är att använda resultaten som ett stöd vid betygssättning, men att de nationella provens resultat inte är ensamt betygsgrundande.

Elisa: Det ska räknas som ett prov som alla andra.

Detta är ett återkommande inslag i lärarnas utsagor även när frågorna inte fokuserat på resultat användningen. Lärarna lyfter dock att de kan bli ifrågasatta av olika aktörer som media, skolhuvudmän, politiker och Skolverket när de efterlever Skolverkets intentioner med att använda provresultatet endast som en del av betygssättningen om lärarens betyg skiljer sig från provresultatet. I lärarnas diskussioner om diskrepanser problematiseras och ifrågasätts att dessa synliggörs offentligt och lättillgängligt. En följd av den statistik som publiceras av Skolverket

är enligt lärarna exempelvis medias ibland felaktiga rapportering och granskning av diskrepanser:

Emilia: De haussar ju upp det här och det enda när man ser i tidningar, då är det nationella provets resultat. [...] De tittar inte på kunskapskraven för elever. Överhuvudtaget. Utan man ser bara i tidningar och media att de tar fram nationella proven och deras resultat. Inte vad de kan egentligen.

Edvin: Nej, media har inte förstått att nationella provet inte är betygsgrundande i den meningen. Utan nationella provet är ju prov som alla andra prov. Det står ju klart och tydligt i instruktionerna också.

Det är tydligt att medias granskning problematiseras och ifrågasätts av lärarna. Men även den granskning av diskrepanser som skolhuvudmän, politiker och Skolverket gör problematiseras, eftersom den ibland försvårar för lärarna att använda provresultatet endast som en del i betygssättningen. Om diskrepanser uppstår upplever lärarna att de tydligt måste kunna motivera avvikelser mellan provbetyget och termins-/slutbetyget:

Berit: Ja. Och det liksom petas i. Jaha, varför stämmer inte betygen, slutbetyget med provbetyget?

Bengt: Och diffar det för mycket, ja, då måste man kunna motivera väldigt tydligt med varför det diffar. Och det stämmer ju inte överens med riktlinjerna för nationella provet. Det står att det ska vara en del i bedömningen som, som helhet.

Emilia: Det som jag har varit med om då, det är ju att jag har bedömt en elev under de år som jag har haft, att den har nått de kunskaper. Och sedan kan de dock missa och inte klara nationella provet. Alltså de kommer inte upp i de, det slår på någon poäng. Då har ju vi i vårt uppdrag att vi måste skriva och förklara så skolledningen har varför jag trots allt ger dem ett godkänt betyg när de går ur nian. Trots att de inte klarar nationella provet.

Intervjuare: Är det praxis, så fort provbetyg avviker från slutbetyg? Att man måste skriva?

Edvin: Men vi har ju blivit ifrågasatta från Skolverket

Lärarna tycks inte kollegialt diskutera eventuella diskrepanser för att kunna motivera, utan det förefaller vara upp till den enskilda

läraren att göra så. Lärarna diskuterar återkommande diskrepanser på individnivå, men inte på skol- eller huvudmannanivå.

Lärarna menar att granskningen från olika håll riskerar att ge oönskade konsekvenser för lärarens arbete med rättning och betygsättning. En konsekvens som lyfts fram handlar om att provet får oproportionellt stort utrymme i förhållande till övrig undervisning och att elevens provresultat får större roll för betyget än Skolverket och läraren själv tänkt vid betygsättning. Lärarna poängterar risker med att lägga för stor, eller all, vikt på provens resultat. Hur väl elever lyckas på ett prov, menar lärarna, kan påverkas av många faktorer som inte nödvändigtvis har med deras matematikkunskaper att göra. Lärarna lyfter fram yttre faktorer, som t.ex. vad som hänt hemma på morgonen eller på rasten innan provet, som faktorer som kan påverka elevernas provsituation och provresultat:

Anders: Det kanske har varit så att han har bråkat stenhårt med sina föräldrar på morgonen och sitter där och är skitförbannad och han vill helst knyckla ihop papperet och gå därifrån. Fast han har som mål att gå ut med ett A i slutbetyg i matematik

Betyget riskerar att inte korrelera med elevens faktiska kunskaper om provresultatet vägleder i allt för stor utsträckning vid betygssättningen. Lärarna är samstämmiga kring att det därför måste få finnas diskrepanser mellan provbetyg och termins-/slutbetyg på individnivå. Detta är något som lärarna återkommer till gång på gång.

En annan konsekvens som lärarna lyfter är att diskrepanser tenderar att bli en måttstock på lärarnas förmåga att sätta betyg. Enligt lärarna förefaller det vara önskvärt att på individnivå sätta samma betyg som provresultatet visar, och ju mindre diskrepanser desto bättre skulle lärarna anses vara på att sätta betyg, som är en sammanvägning av elevernas tidigare kunskaper. Om lärarens sammanvägning av elevernas kunskaper skiljer sig från resultatet på nationella proven kan lärarnas bedömningsförmåga ifrågasättas av skolhuvudmän:

Anders: Det är ju det som blir märkligt. Tycker jag. Och sedan att man får den här statistiken kastad i ansiktet sedan, hur kommer det sig att du har gett eleven det här betyget när den skrev det här resultatet på nationella provet? [...] Det blir ju att ifrågasätta mig. Det är ju så det blir. Alltså att om jag... Eleven har skrivit ett E på nationella provet. Men får ett C i slutbetyg. Då är det ju också sådär att ja, men hur kan eleven få ett C i slutbetyg?

Sammanfattningsvis ger lärarna en bild av att vilja använda proven som en del av betygssättningen vilket skulle kunna generera vissa skillnader mellan provbetyg och termins-/slutbetyg. Fokus ligger på diskrepanser på individnivå snarare än skolnivå. De diskuterar och är kritiska till att diskrepanser offentliggörs genom lättillgänglig statistik, framför allt eftersom statistiken, enligt lärarna, felaktigt ger proven en förstärkt roll i förhållande till annan undervisning. Lärarna ger också uttryck för att granskningen leder till ett ifrågasättande av deras bedömarkompetens.

Skolinspektionens och Skolverkets granskning av lärarnas rättning

Matematiklärarna är i jämförelse med andra ämneslärare mer positiva till bedömningsverktyg som tillhandahålls i samband med nationella proven. Genomgående är de positiva till de bedömningsmallar som finns och ger uttryck för att de är bra stöd vid rättning och betygssättning. När det gäller rättning görs detta i stor utsträckning enskilt men vid tveksamheter tar de stöd av andra kollegor. Flera av lärarna uttrycker dock en önskan om att, utöver kollegialt stöd, få återkoppling på sin rättning från central instans.

Edvin: Man har rättat och sedan har man då ett datum, ja den här elev ska jag skicka in. Och det skulle ju vara väldigt trevligt och veta någon gång, har jag gjort rätt? Har jag gjort fel? Är det någonting som ni har och anmärka på min rättning eller så vidare och så vidare. Och då kan det ju vara för hela kollegiet, för vi kanske har sambedömt vissa uppgifter. Och då får vi, jaha, men där tänkte vi inte riktigt som de tänkte.

Erik: [...] Jag har skickat in en hel årskull en gång. Jag har inte hört ett ljud om jag har gjort ett bra jobb eller ett dåligt jobb. Så jag tycker dom missköter den punkten, eller också så är det jag som inte förstår att jag måste ta kontakt med dom eller någonting sådant. Men jag förväntar mig att få reda på det här och inte läsa det i tidningen att eleverna är upprättade eller har fått för bra betyg. Utan att dom verkligen tar kontakt med oss då, talar om vad gör vi för fel. Vad är problemet? Eller ska man tolka det som att är det tyst är det bra, eller?

En övervägande del av lärarna är positiva till att få externa granskningar med återkoppling på den egna rättningen. Lärarna önskar ytterligare återkoppling på sin rättning, vid de fall den skickas in till Skolinspektionen för granskning.

Elisa: Det skulle vara min önskan. Skulle jag bara titta på mig själv som lärare så skulle jag med glädje skicka in dem här för rättning. [...] Jag skulle självklart vilja titta på dem, få tillbaks dem och titta på dem, absolut.

Edvin: För man vill ju liksom, när man ändå skickar in saker och ting. För vi får ju respons av varandra när vi sitter och diskuterar det här när vi är färdiga. Medan då Skolverket, den stora draken, borde ju bemöda sig om att skicka ut ett svar. Ja, men det här har ni gjort väldigt bra eller det här, du får tänka på det här och det här till, till en annan gång. Så man får något gensvar på det jobb man lägger ner.

Rättningsgranskning ses som en möjlighet att få bekräftat i vilken utsträckning de rättat en uppgift på provet såsom Skolverket avsett, om återkoppling sker till lärarna. Återkopplingen ses framförallt som fortbildande.

Ellen: Ja. Det skulle ju vara jättekul. Ja. Och att man gör en liten anteckning själv, på ett ungefär om man tittar igenom dem. Och sedan få tillbaks dem och sedan se hur de har rättat. Det skulle vara intressant faktiskt.

Anders: Förklara hur du tänkte när du gjorde den här bedömningen. För jag menar, det skulle ju vara en utveckling för mig som lärare också. Alltså att om jag får en respons på att jag har gjort någonting.

Sammanfattningsvis är lärarna positiva till att få sina rättningar granskade av Skolverket och Skolinspektionen, om det sker i ett fortbildande syfte. Lärarnas vill få återkoppling och en chans att motivera och diskutera rättningen av specifika elevexempel, men också kalibrera sina rättningar med Skolinspektionens rättning. Eftersom de upplever brist på återkoppling uppfattas dock rättningsgranskningen som en kontroll från ”den stora draken” Skolverket istället för som ett mer positivt stöd.

Diskussion

I den här artikeln har undersökts vilka konsekvenser matematiklärare uttrycker att de nationella provens syfte att granska likvärdigheten i lärares bedömning fått.

En första del av det senast tillkomna syftet med nationella proven är att genom statistik förstå och förklara avvikelser mellan provresultat och termins-/slutbetyg. Vi betraktar detta som en del av en internationell trend mot ökad kontroll av skolans resultat genom resultatpublicering, den trend som exempelvis Sahlberg (2016) kallar GERM. I forskning om effekter av den här typen av ansvarsutkrävande är lärares perspektiv sällan representerat (Mickwitz 2015) och vi har därför frågat oss om statistiken är till för att belysa och ge förklaringar till de diskrepanser som finns mellan resultaten på nationella proven och elevers termins-/slutbetyg, samt hur lärare ser på detta.

Lärarnas utsagor i denna artikel indikerar att diskrepanser belyses i större utsträckning nu än tidigare, och att lärare i hög grad är medvetna om de diskrepanser som finns. Detta har medfört olika konsekvenser som lärarna diskuterar. Kraven på att kunna motivera diskrepanser på individnivå menar lärarna har ökat, vilket lärarna tycker är frustrerande. Alla intervjuade lärare i studien är överens om att det måste få finnas avvikelser mellan provresultatet och elevens betyg. Detta eftersom de menar att resultaten på proven kan innefatta mindre information än elevernas betyg. Lärarnas synpunkt stöds av Willingham et al. (2002) eftersom de menar att externa prov och betyg kan bygga på delvis olika *constructs*, vilket behöver tas i beaktande vid betygssättning. Lärarna uttrycker en tydlig oro att provresultaten tenderar att få en allt större roll vid betygssättning. Skolverkets intention, att de nationella provens resultat endast ska utgöra en del vid betygssättning, är svår att leva upp till. I förlängningen finns det en risk att undervisningen blir mer instrumentellt och reduktionistiskt inriktat på det enkelt mätbara (Kelly et al. 2017).

Att lärarna förväntas motivera individuella avvikelser mellan provresultat och betyg skulle å ena sidan kunna bidra till en ökad förståelse för uppkomsten av avvikelser mellan betyg och provresultat vilket är en av intentionerna med att diskrepanser ska synliggöras. Detta eftersom lärarna förväntas ge förklaringar till de diskrepanser som finns, vilka Skolverket sedan skulle kunna använda för att analysera orsaker till uppkomna diskrepanser. Å andra sidan är det skillnader på individnivå som lärarna ger uttryck för att de måste motivera, medan Skolverket (2007) och Regeringen (2011) avser att förklara diskrepanser på skol- och huvudmannanivå. Att lärarna enskilt snarare än kollegialt motiverar diskrepanser stärker bilden av att det är individuella diskrepanser som motiveras. Sett ur detta perspektiv ger motiveringar på individnivå ingen ny information om diskrepanser på skol- och huvudmannanivå. Att förtydliga vikten av att göra systematiska resultatanalyser och analysera skillnader på skol- och huvudmannanivå snarare

än på individnivå skulle kunna vara en möjlighet att använda Skolverkets lättillgängliga statistik som diskussionsunderlag för skolutveckling.

En annan konsekvens som följer av att statistiken publiceras är, enligt lärarnas utsagor, att det blir lärarens förmåga att bedöma och att sätta betyg som i sig värderas och bedöms, snarare än att det leder till en diskussion kring varför diskrepanser uppkommit. Lärarnas bild är att detta allt större fokus på diskrepanser mellan resultaten på nationella proven och termins-/slutbetygen gjort att deras förmåga att bedöma och betygssätta ifrågasätts från olika aktörer som Skolverket, Skolinspektionen, skolhuvudmän, media och även elever och deras målsmän. Att bedöma lärares bedömningskompetens utifrån samstämmigheten mellan betyg och resultatet på ett externt prov är problematiskt eftersom mycket mer information om eleven och den lokala kontexten behöver vägas in för att en sådan bedömning ska vara legitim (Willingham et al. 2002). Ett ifrågasättande av lärares bedömningsförmåga kan också påverka deras professionella självbild (Mickwits 2015).

En andra del av det senast tillkomna syftet med de nationella proven är att Skolverket tillsammans med Skolinspektionen ska granska lärares rättningar och uppmärksamma eventuella avvikelser. Hur upplever matematiklärare att de påverkas av denna granskning? Lärarna har skickat in sina rättningar för granskning, men ingen av lärare i vår studie har fått återkoppling på rättningen eller uppmärksamhets på avvikelser mellan sin rättning och Skolverkets rättning. Flera av lärarna ger uttryck för att de gärna ser att deras rättning av de nationella proven kontrolleras och sedan återkopplas till dem. De uttrycker tydligt en önskan om att Skolverket och Skolinspektionen bedömer och ger återkoppling på deras rättning i fortbildande syfte, något som de i sin tur tror kan leda till en mer likvärdig bedömning.

Å ena sidan framför således lärarna en önskan om att granskas, medan de å andra sidan ifrågasätter att diskrepanser mellan provbetyg och termins-/slutbetyg granskas på det sätt som gjorts, där olika aktörer som Skolverket, huvudman och Skolinspektionen lagt stort fokus vid att utreda varför diskrepanserna förekommer. Vid en första anblick kan det upplevas paradoxalt. Dock är det tydligt att det är platsen och formen för granskningen, snarare än granskningen i sig, som lärarna vänder sig emot. De menar att en granskning av diskrepanser i offentlig miljö har snedvridit de nationella provens roll, vilket gör att lärarna exempelvis oroas över att resultaten på nationella proven får en alltför stor roll vid den slutliga betygssättningen.

I ett större sammanhang framskymtar ett dilemma med två lager. I ett första lager förutsätts att tillit finns till lärarnas rättning. Utifrån exempelvis Susan Brookhart (2015) och Susan Brookhart

et al (2016) ligger problematiken i detta lager i att provresultat blir dominerande då lärarna väger samman elevernas kunskaper. Flera nödvändiga aspekter beaktas då inte vid betygssättning och betyget riskerar att inte återspegla elevernas faktiska kunskaper. Skolverkets och Skolinspektionens uppdrag från Regeringen, att granska lärarnas rättningar av uppgifter på nationella proven (Regeringen 2011), indikerar dock att lärarnas rättningsförmåga *måste* granskas. Även lärarna uttrycker att den granskning som de utsätts för till viss del ifrågasätter deras förmåga att rätta och sätta betyg. Saknas tilltro till lärares rättningsförmåga finns det i ett andra lager dessutom en problematik med provresultatets validitet, och betyget riskerar att i ännu mindre utsträckning återspegla elevernas faktiska kunskaper. Dilemmat uppstår om provresultatet styr mer än det borde för att tilltron till lärarnas förmåga att göra bedömningar inför betygssättning är ifrågasatt, samtidigt som denna brist på tilltro till lärarna också gör att provresultatets validitet är hotad.

Resultaten i denna artikel visar vissa oroande effekter av det senast tillkomna syftet med nationella proven. Effekterna handlar om de nationella provens allt mer styrande roll vid betygssättning och ifrågasättande av lärares professionalism vid rättning. Dessa effekter skulle behöva undersökas systematiskt utifrån ett större datamaterial än vad vi haft tillgång till för denna studie, med fördel då också mer koncentrerat till matematik där diskrepanser mellan resultat på nationella proven och termins- och slutbetyg är större än i engelska och svenska (Skolverket 2007).

Noter

1. I artikeln används återkommande begreppet rättning, betygssättning och bedömning. Med *rättning* avses processen som sker när läraren granskar en elevs lösning för att tilldela poäng och/eller betyg på en uppgift. Detta kan ske med eller utan överväganden baserat på givna anvisningar för hur rättningen ska gå till. I artikeln används begreppet rättning främst i relation till uppgifter på nationella proven. Med *betygssättning* avses den process som sker när läraren samlar in och värderar information om elevers kunskaper för att tilldela ett betyg, vilket i den meningen är detsamma som summativ bedömning. När vi i artikeln använder begreppet *bedömning* är det just med syftningen i en summativ bedömning, det vill säga i meningen att värdera elevernas befintliga kunskaper för att sätta betyg. Begreppet kan således sägas innefatta både rättning och betygssättning. Vidare inkluderas i begreppet de överväganden som lärare behöver göra för att i vissa fall kunna rätta en uppgift
2. SiRiS är en förkortning för: Skolverkets Internetbaserade Resultat- och kvalitetsInformationsSystem. Se vidare: <http://sir.is.skolverket.se/siris/f?p=SIRIS:11:0>

Referenser

- Au, Wayne (2011): Teaching under the new Taylorism: high-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, 43(1), 25-45.
- Baker, Eva (2016): Research to controversy in 10 decades. *Educational Researcher*, 45(2), 122–133.
- Bonnevier, Jenny; Arensmeier, Cecilia; Borgström, Eric; Lennqvist Lindén, Ann-Sofie; Lundahl, Christian; Nilsson, Per; Sundberg, Bodil; Sundhäll, Marcus; Yassin, Daroon & Wetterstrand, Frida (2016): *De nationella proven och deras effekter i årskurs 6 och 9 – en intervjustudie med elever, lärare och skolledare*. Stockholm: Skolverket.
- Brookhart, Susan (2015): Graded achievement, tested achievement, and validity. *Educational Assessment*, 20(4), 268–296
- Brookhart, Susan; Guskey, Thomas; Bowers, Alex; McMillan, James; Smith, Jeffrey; Smith, Lisa; Stevens, Michael & Welsh, Megan (2016): A century of grading research: meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848.
- Darling-Hammond, Linda; Wilhoit, Gene & Pittenger, Linda (2014): Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, 22(86), 1-38.
- Dee, Thomas & Jacob, Brian (2011): The impact of No child left behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- Denscombe, Martyn (2013): *Forskningshandboken – för småskaliga forskningsprojekt inom samhällsvetenskaperna*. Lund: Studentlitteratur.
- Harlen, Wynne & Deakin Crick, Ruth (2002): A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1*). In: *Research evidence in educational library*. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harris, Douglas & Herrington, Carolyn (2006): Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education*, 112(2), 209–238.
- Hopmann, Stefan Thomas (2003): On the evaluation of curriculum reforms. *Journal of Curriculum Studies*, 35(4), 459-478.

- Forsberg, Eva & Lundahl, Christian (2009): Skolans interna och externa kunskapsbedömningar. I Rita Forss-Lindblad & Rolf Lander, red: *Att säkra det osäkra – reflektion och makt i skolans utvärdering*, s. 61 – 84. Lund: Studentlitteratur.
- Kelly, Peter; Andreassen, Karen Egedal; Kousholt, Kristine; McNess, Elizabeth & Ydesen, Christian (2017): Education governance and standardised tests in Denmark and England. *Journal of Education Policy*. A head of print. doi:10.1080/02680939.2017.1360518
- Korp, Helena (2011): *Kunskapsbedömning - vad, hur och varför?* Stockholm: Natur & Kultur.
- Lindström, Lars (2005): Pedagogisk bedömning. I Lars Lindström & Viveca Lindberg red: *Pedagogisk bedömning. Om att dokumentera, bedöma och utveckla kunskap*, s. 9-30. Stockholm: HLS förlag.
- Linn, Robert (2000): Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- Lundahl, Christian (2006): *Viljan att veta vad andra vet. Kunskapsbedömning i tidigmodern, modern och senmodern skola*. Arbetsliv i omvandling 2006:8. Akademisk avhandling vid Uppsala universitet. Stockholm: Arbetslivsinstitutet.
- Lundahl, Christian (2009): *Varför nationella prov? - framväxt, dilemman, utmaningar*. Lund: Studentlitteratur.
- Lundahl, Christian (2010): Skolbedömningens pedagogiska och administrativa dimensioner. I Christian Lundahl & Maria Folke-Fichtelius, red: *Bedömning i och av skolan - praktik, principer, politik*, s. 299 – 314. Lund: Studentlitteratur.
- Lundahl, Christian (2011): *Bedömning för lärande*. Stockholm: Norstedts.
- Lundahl, Christian; Hultén, Magnus & Tveit, Sverre (2017a): The power of teacher-assigned grades in outcome-based education. *Nordic Journal of Studies in Educational Policy* 3(1), 56-66.
- Lundahl, Christian; Hultén, Magnus & Tveit, Sverre (2017b): Betygssystem i internationell belysning. Underlagsrapport till Skolverkets utvärdering av betyg. Stockholm: Skolverket.
- Mickwitz, Larissa (2015): *En reformerad lärare. Konstruktionen av en professionell och betygssättande lärare i skolpolitik och skolpraktik*. Doktorsavhandling, Stockholms universitet.
- Pettersson, Astrid (2010): Bedömning av kunskap för lärande och undervisning. *Forskning om undervisning och lärande* 2010(3): s. 7-19.

- Ready, Douglas & Wright, David (2011): Accuracy and inaccuracy in teacher's perception of young children's cognitive abilities: the role of child background and classroom context. *American Educational Research Association*, 48(2), 335-360.
- Regeringen (2011): *Regleringsbrev för budgetåret 2011 avseende Statens Skolverk*. (Regeringsbeslut I:12). Hämtad 2016-09-06 från <http://www.esv.se/Verktyg--stod/Statsliggaren/Regleringsbrev/?RBID=14188>.
- Sahlberg, Pasi (2016): The global educational reform movement and its impact on schooling. I Karen Mundy; Andy Green; Bob Lingard & Antoni Verger, red: *The Handbook of Global Education Policy*. New York: John Wiley & Sons, 128-144.
- Skolverket (2007): *Provbetyg - slutbetyg - likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolans årskurs 9, 1998-2006*. Hämtad 2016-09-06 från: <http://kvutis.se/wp-content/uploads/2014/05/pdf1725.pdf>
- Skolverket (2011): *Kunskapsbedömning i skolan - praxis, begrepp, problem och möjligheter*. Stockholm: Fritzes.
- Skolverket (2016a): *Statistik över relationen provbetyg - terminsbetyg*. Hämtad 2016-09-06 från: <http://siris.skolverket.se/siris/f?p=SIRIS:149:0::NO::>
- Skolverket (2016b): *Statistik över relationen provbetyg-slutbetyg*. Hämtad 2016-09-06 från: <http://siris.skolverket.se/siris/f?p=SIRIS:150:0::NO::>
- Skolverket (2017): *Vilka är syftena med bedömning och betyg?* Hämtad 2017-02-23 från: <http://www.skolverket.se/skolutveckling/forskning/bedomning/tema-bedomning/vilka-arsyftena-med-bedomning-och-betyg-1.157699>
- Südkamp, Anna; Kaiser, Johanna & Möller, Jens (2012): Accuracy of teachers' judgment of students' academic achievement: a meta-analysis. *American Psychological Association*, 104(3), s. 743-762.
- Waldow, Florian (2013). Conceptions of justice in the examination systems of England, Germany, and Sweden: a look at safeguards of fair procedure and possibilities of appeal. *Comparative Education Review*, 58(2), 322-343.
- Westlund, Ingrid (2009): Hermeneutik. I Andreas Fejes & Robert Thornberg, red: *Handbok i kvalitativ analys*, s 62-80. Stockholm: Liber.
- Willingham, Warren; Pollack Judith & Lewis, Charles (2002): Grades and test scores: accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1-37.