

Normers roll i ett mål- och kriterierelaterat bedömningsystem

Jenny Bonnevier, Eric Borgström & Daroon Yassin Falk

RELATING TO NORMS IN A CRITERION-REFERENCED SYSTEM OF EVALUATION. Since the 1990s, the Swedish system of evaluation is criterion-referenced. In this system, one of the main purposes of national tests is to promote equity in assessment. This article explores how teachers in the subject fields of English and Swedish in compulsory school (grades 6 and 9), understand the relationship between the act of grading and the national test results. We find that the teachers we have interviewed typically dismiss quantitative data and comparisons at aggregated levels and we argue that this dismissal has its root cause in an erroneous understanding of criterion-referenced assessment as completely independent of norms. A better, more nuanced understanding of the role of norms in criterion-referenced assessment would, we suggest, allow teachers to make better use of the potential of national tests in promoting equity.

Key words: criterion-referenced assessment, norm-referenced assessment, national tests, equity.

Inledning

Den svenska skolans bedömningsystem är som bekant sedan 1990-talet mål- och kriterierelaterat. Betyg och resultat på centralt utarbetade prov ska därmed inte längre rangordna elevernas prestationer i förhållande till genomsnittsprestationen, så som de skulle i det relativa betygssystemets tidevarv. Istället ska kriterierelaterade

Jenny Bonnevier är lektor i engelska vid Örebro universitet, 702 81 Örebro. E-post: jenny.bonnevier@oru.se

Eric Borgström är lektor i svenska språket vid Örebro universitet, 702 81 Örebro. E-post: eric.borgstrom@oru.se

Daroon Yassin Falk är lektor i svenska språket vid Örebro universitet, 702 81 Örebro. E-post: daroon.yassin@oru.se

resultat bära information om elevens kunskapers art och grad. I detta ligger att betygsfördelningen på nationell nivå inte längre förutsätts följa normalfördelningskurvan. Det finns med andra ord ingen principiell gräns för hur stor andel elever som kan få ett visst betygssteg. I den bästa av världar lämnar ju alla elever skolan med adekvata kunskaper, och därmed med åtminstone betyget E i samtliga ämnen. Det mål- och kriterierelaterade bedömningssystemet skiljer sig med andra ord på fundamentala vis från det normrelaterade.

Alltjämt är emellertid betygens grundläggande funktion urval till bland annat högre studier, en funktion som är central för varje meritokratisk demokrati. För systemets legitimitet fordras likvärdighet, och en av de nationella provens huvuduppgifter i det nuvarande bedömningssystemet är att främja just detta (Skolverket 2017a). Genom proven ska lärare kalibrera sin användning av betygsskalan och kunskapskraven (jfr Wikström 2006).

I den här artikeln fokuserar vi hur några grundskolelärare i ämnena svenska och engelska resonerar om de nationella provens roll i det mål- och kriterierelaterade betygssystemet. Med utgångspunkt i intervjuer visar vi hur dessa lärare är kluvet inställda till de olika resultatunderlag som de nationella proven genererar, och hur dessa kan användas i betygsuppdraget. Som informationskälla på individnivå är proven välkomna. Delprovsresultaten betraktas som meningsfulla och användbara indikationer på hur den individuella eleven presterar i förhållande till de specifika kunskapskrav respektive delprov prövar. Som informationskälla på gruppnivå avfärdas däremot proven mer eller mindre genomgående. Att de meningsfulla delprovsresultaten ska vägas samman i ett så kallat provbetyg för att möjliggöra sådana analyser betraktas som en ovälkommen administrativ exercis. Lärarna ser nämligen inte något värde med de kvantitativa analyser som möjliggörs genom sammanvägningen, det vill säga av hur den egna betygssättningen samvarierar med hur eleverna presterat på de nationella proven. Sådana analyser betraktas, som vi ska se, rent av som oförenliga med betygsuppdraget i det mål- och kriterierelaterade betygssystemet. Det är logiken bakom detta principiella avfärdande av kvantitativa jämförelser mellan provresultat och betyg vi intresserar oss för. Hur leder dessa lärares sätt att resonera om nationella prov och betygssättning till slutsatsen att kvantitativa analyser av provresultat och betyg strider mot idén om kriterierelaterad bedömning?

Lärares betygsuppdrag ställer dem inte sällan inför att hantera motstridiga rationaliteter, och oförenliga förväntningar. Detta är väl belagt i såväl internationell som svensk forskning (Lundahl m.fl. 2015 ger en översikt). Motstridigheterna kan vara av olika slag. Exempelvis uppstår dilemmasituationer för lärare när *accountability*-drivna,

standardiserade tester upplevs kollidera med elevers individuella behov (Pope, Johnson & Mitchell 2009), eller med lärarens egna uppfattningar om vad som är att betrakta som meningsfull bedömning (Suurtam & Koch 2014). I andra fall är det sättet som bedömningsuppdraget är formulerat på, med interna konflikter mellan olika policyskrivningar, som ställer läraren inför att avgöra vilken skrivelse som ska följas och vilken som ska brytas mot (Simon m.fl. 2010). En svensk studie som behandlar just denna typ av dilemma är Rähä (2009), där vi får följa en grupp lärare som är ålagda att skriva lokala betygskriterier. Dilemmat lärarna har att hantera ligger i uppdraget; de lokala kriterierna ska vara konkretiseringar av de nationella kriterierna, samtidigt som de inte får avgränsa deras innebörd. Ytterligare en svensk studie på temat är Mickwitz (2015). Studieobjektet är lärares berättelser om den egna betygssättningspraktiken, vilken konstrueras som full av motsägelser. Skolvardagen blir i lärarnas berättelser en ”där ingen verkar vilja att de ska sätta det rätta betyget” (2015, s 217). Omgivningen – det vill säga elever, föräldrar och rektorer – vill istället att lärarna ska sätta högt betyg.

När vi i denna artikel undersöker lärares sätt att resonera om de nationella provens roll i betygssättningsuppdraget, gör vi detta utifrån argumentet att logiken bakom deras resonemang hänger samman med en specifik rationalitetskonflikt i det svenska betygssystemet. Denna konflikt springer ur det faktum att normer och deras funktion vid mål- och kriterierelaterad bedömning sällan berörs när bedömningsuppdraget beskrivs i styrdokument och stödmaterial från Skolverket (se nedan). Denna avsaknad av en normdiskussion menar vi är problematisk. Även om kriterierelaterad bedömning inte är detsamma som normrelaterad, förlorar kriterierna sin mening om de inte underbyggs av just normer, det vill säga delade föreställningar om vad som är rimligt att förvänta sig av dem som bedöms (jfr Angoff 1974, Kimbell 2007, Koretz 2008, Wiliam 2010). Vi föreslår att en mer nyanserad förståelse av kriterierelaterad bedömning som fenomen skulle ge lärare bättre möjligheter att se de nationella provens potentialer som input i betygsuppdraget.

I avsnittet som följer tar vi utgångspunkt i några exempel på hur idén om kriterierelaterad bedömning som fri från normer kommer till uttryck, och utvecklar sedan idén om normers betydelse vid kriterierelaterad bedömning. Därefter presenterar vi vårt material och vår analysmetod närmare, för att sedan övergå till undersökningens resultat.

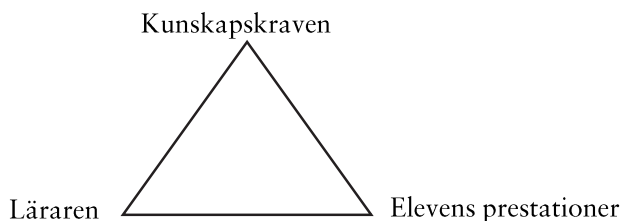
Kriterierelaterad bedömning

Som exemplen nedan kommer att visa, presenteras understundom kriterierelaterad bedömning som väsensskild normrelaterad. Vid kriterierelaterad bedömning bedöms prestationen enbart utifrån aktuella kriterier, medan jämförelser med andra prestationer istället hör normrelaterad bedömning till. Tydligast blir denna idé formulerad i den skolverksutgivna forskningsöversikten *Kunskapsbedömning – vad, hur och varför?* (Korp 2011). Där definieras kriterierelaterad bedömning och vårt nuvarande system på följande vis:

Kriterierelaterad bedömning handlar om att värdera och beskriva en prestation enbart i förhållande till vissa absoluta kriterier – principen är att det omdöme eller det betyg som sätts representerar vissa kunskapskvaliteter eller en viss kunskapsnivå. All bedömning i svenskt skolväsende är kriterierelaterad. (Korp 2011, s 72)

Som synes beskrivs kriterier som absoluta per definition. Det är denna egenskap i kriteriet som driver bort jämförelsebehovet i bedömningsakten. När ett kriterium är absolut kommer en given prestation antingen att nå eller inte nå kriteriets beskrivning – jämförelser med andra prestationer blir överflödiga. I sin renaste form består en kriterierelaterad bedömningsakt därför enbart av tre komponenter: läraren, kunskapskraven och elevens prestationer. Denna idé visualiseras i figur 1:

Figur 1: Idealformen av kriterierelaterad bedömning.



Som ett idealiserande sätt att poängtera den kriterierelaterade bedömningens grundvalar fungerar denna beskrivning av bedömningsuppdraget väl. Det stämmer överens med Skolverkets sätt att definiera handlingen att sätta betyg, nämligen ”att bedöma med vilken kvalitet en elev gör och visar det som beskrivs i kunskapskraven” (Skolverket 2017b). Triaden markerar exempelvis att betygssättningen av en viss elevs kunskaper inte ska påverkas av hur trevlig eleven är eller för den delen av föräldrars eller rektorers åsikter om vilket betyg som är lämpligt. Den gör också den viktiga distinktionen mot normrelaterad

bedömning; en elevs betyg ska vid kriterierelaterad bedömning inte bero på hur övriga elever i landet presterat. Dessa poänger är viktiga, inte minst som flera svenska studier pekar på att sådana hot mot lärarbedömningars validitet faktiskt förekommer (se exempelvis Klapp Lekholm 2008, Mickwitz 2011, Selghed 2004).

Men även om skiljelinjen mot normrelaterad bedömning är reell och viktig att betona blir det missvisande att beskriva kriterierelaterad bedömning som fri från normer. I praktiken underbyggs varje kriterierelaterat bedömningssystem av normer, det vill säga gemensamma föreställningar om vad som är rimligt att förvänta sig av elever i en viss årskurs. Kanske fångas detta förhållande allra bäst av William Angoff: "lurking behind the criterion-referenced evaluation, perhaps even responsible for it, is the norm-referenced evaluation" (1974, s 4, i Wiliam 2010, s 254). På systemnivå är normer nödvändiga vid urvalet av vilka kriterier som ska gälla. Även i de fall kriterier är möjliga att formulera i absoluta ordalag, krävs en förståelse för dem som ska bedömas utifrån kriterierna för att avgöra kriteriernas relevans. Sett till mätbarhet är kriteriet "kan simma 200 meter varav 50 meter i ryggläge" inte bättre än "kan simma 20 meter, varav 5 meter i ryggläge". Det är sett till populationen 12-åriga elever i den svenska skolan som det första alternativet framstår som det bästa.

Vidare kan normer även behöva aktiveras i den kriterierelaterade bedömningsakten. Kriterier är nämligen inte per definition absoluta utan kan lika gärna vara diffust formulerade (jfr diskussionen av *sharp vs fuzzy criteria* i Sadler 1989, s 124). Och det är bara i de fall kriterier lever upp till epitetet absoluta som en prestation blir möjlig att bedöma i relation till kriteriet allena. Absolut formulerade kriterier är i själva verket mycket ovanliga i de svenska styrdokumenterna och provmaterialen (vilket förstås hänger samman med de typer av kunskaper som dessa ger uttryck för). Ett exempel på ett absolut kriterium återfinns emellertid i svenskämnet nationella prov i årskurs 3. Där anges att om en elevtext i dessa prov uppvisar tre meningar som börjar med stor bokstav och slutar med stort interpunktionstecken ska texten bedömas nå det mål som gäller interpunktionsfärdighet (Skolverket 2016a, provmaterial åk 3). Med ett sådant absolut formulerat kriterium behövs med andra ord ingen jämförelse med andra prestationer - antingen når prestationen kriteriet, eller så gör den det inte. Om vi istället går till det kunskapskrav i årskurs 3, som de nationella proven är tänkta att indexera, hittar vi inte längre någon skarp gräns för vad det innebär att vara duglig brukare av interpunktionsreglerna i årskurs 3. Här är kunskapskravet ett generellt påstående som bygger på omdöme och erfarenhet i tillämpningen:

I texterna kan eleven använda stor bokstav, punkt och frågetecken. (Skolverket 2015, s 228)

Varje lärare vet att detta kunskapskrav inte nödvändigtvis innebär att eleven i varje mening i varje text måste få till interpunktionen enligt skriftspråkets normer. I så fall skulle inte många elever kunna bedömas nå målet i slutet av årskurs 3. Underförstått i kunskapskravet är att elevens interpunktionskunnande ska ta sig uttryck i elevens texter på den nivå av korrekthet som är rimlig att förvänta sig av en elev i årskurs 3. Kunskapskravets innebörd är med andra ord inte inneboende i orden som sådana, utan genom föreställningar om vad som är att betrakta som årskurstypiska elevprestationer.

Denna typ av värdeordskriterier går som bekant igen i svensk- och engelskämnetts kunskapskrav för hela grundskolan. Vi kan ta ett typiskt exempel på detta från engelskämnetts kunskapskrav, för betyget C i årskurs 9:

I muntliga och skriftliga framställningar i olika genrer kan eleven formulera sig relativt varierat, tydligt och sammanhängande. (Skolverket 2015, s 32)

Den lärare som ska tillämpa denna typ av kriterier behöver göra en rad tolkningar. Ett av de beslut läraren måste fatta för att kunna tillämpa detta kunskapskrav mot en elevs samlade prestationer rör var gränsen går för ”relativt varierat, tydligt och sammanhängande”. Kriteriets innebörd finns inte i orden som sådana, utan i den omdömesgilla och erfarna lärarens uttolkning av orden. I sin tolkning och tillämpning av kriteriet måste läraren ta hänsyn till den grupp som ska bedömas, elever i årskurs 9. Det är just i relation till denna elevkategori som kunskapskravet formulerats och värdeorden får sin innebörd.

Om det är så, som anges i figur 1, att de enda referenspunkterna vid bedömning får vara kunskapskravet och den prestation som läraren för stunden ska bedöma, blir bedömningen starkt beroende av den enskilde läraren. En sådan total villkorlighet undergräver naturligtvis idén om likvärdighet. Med synsättet att kriterierelaterad bedömning är normberoende blir det inte upp till den enskilde läraren att definiera kunskapskraven. Kunskapskravens innebörd behöver vara förankrade i professionsgemensamma normer, och lärarkåren utgöra en tolkningsgemenskap.

Material och urval

Denna artikel bygger på gruppintervjuer med lärare genomförda inom ramen för projektet *Nationella proven i grundskolans årskurs 6 och 9: En uppföljning av lärares, rektorers och elevers uppfattning om proven* (Skolverket 2016b, se även Lundahls inledning till detta nummer). Mer exakt bygger artikeln på 11 gruppintervjuer, med 2-4 lärare, där minst en av informanterna i varje intervju är engelsk- och/eller svensklärare. Totalt ingår 25 engelsk- och/eller svensklärare i materialet. Samtliga intervjuer varade cirka en timme. I sex gruppintervjuer ingår lärare i årskurs 6, och i de resterande fem ingår lärare i årskurs 9. I det utvalda intervjumaterialet finns fyra olika skolor representerade, och från varje skola återfinns minst en intervju med lärare i årskurs 6 och en med lärare i årskurs 9. Det material som blivit föremål för närmare analys är de samtalsavsnitt där de av informanterna som är engelsk- och/eller svensklärare antingen diskuterar vad resultat från de nationella proven står för (det vill säga hur de ska/bör/kan användas), eller betygsuppdraget som sådant och hur detta uppdrag ska/bör/kan utföras. Citaten i texten är dels anpassade till skriftspråksnormer, dels ofta förkortade för att på ett effektivt sätt förmedla den för sammanhanget centrala informationen.

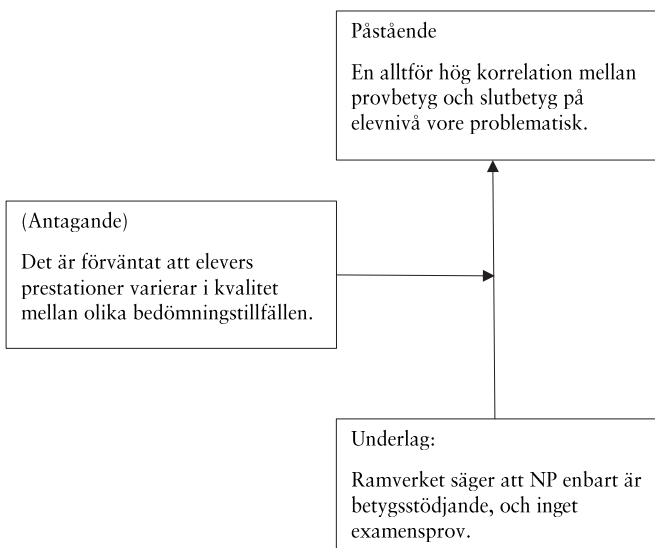
Analysmetod

I artikeln studerar vi vilka påståenden lärare gör om nationella provresultaten och hur provresultat förhåller sig till betygssättningen. Detta gör vi för att få syn på lärares föreställningar om vilka roller de nationella proven kan och inte kan spela vid betygssättning i ett kriterierelaterat bedömningssystem. För att komma åt dessa föreställningar och – viktigare – hur dessa hänger ihop använder vi Toulmins (2003) modell för praktisk slutledning (eng. practical reasoning). Med Toulmin handlar ett resonemang om att göra *påståenden* (claims), samt genom argument stötta trovärdigheten i dessa påståenden. Argumentet bottnar i ett *underlag* (data) av ett eller annat slag, samt *antaganden* (warrants) som gör den specifika slutledningen som påståendet ger uttryck för möjlig.

För att förklara Toulmins tankefigur kan vi ta utgångspunkt i ett resonemang om korrelationen mellan provbetyg och slutbetyg som förs i en rapport från Skolverket (2007). I denna rapport konstaterar författarna i en passage (2007 s 29) att det finns en övre gräns för hur stark korrelationen mellan provbetyg och slutbetyg bör vara på elevnivå. En alltför stark korrelation skulle nämligen kunna indikera att

ämnesproven används som examensprov när de i själva verket enbart ska fungera betygsstödjande. I Toulmins modell är detta resonemang strukturerat som följer. Påståendet författarna gör är att en mycket hög korrelation mellan provbetyg och slutbetyg på elevnivå skulle kunna betyda att ämnesproven fungerar som examensprov. Som *underlag* tjänar det rådande ramverket som säger att nationella prov ska fungera betygsstödjande, men inte vara något examensprov. Det *antagande* som driver argumentet från underlag till påstående är det välkända förhållandet att elever vid enskilda prov inte alltid presterar på just den nivå som är representativ för deras verkliga kunskaper. Mot den bakgrunden blir en fullständig överensstämmelse mellan provbetyg och slutbetyg osannolik – eller rentav icke önskvärd – när proven enbart är betygsstödjande. Resonemanget kan visualiseras som följer:

Figur 2: Strukturen i ett praktiskt resonemang.



På detta vis söker vi i artikeln reda ut hur lärare resonerar om relationen mellan nationella prov-resultat och betygssättning. Vi utgår från frågan vilka återkommande påståenden lärarna gör om relationen, samt vilka underlag de baserar sina påståenden på. Därefter söker vi identifiera och (om-) formulera det eller de antaganden som gör slutledningen från underlag till påstående rimlig. Ibland är antagandena uttalade i intervjuerna, ibland är de underförstådda. Det vi gör är med andra ord att avgränsa, isolera och sammanfatta söka visualisera lärarnas resonemang för att på så vis renodla frågan om normers roll i kriterierelaterad bedömning. Samtidigt innebär detta

snäva angreppssätt att vi inte kan göra rättvisa åt varken de enskilda lärarnas resonemang om bedömning i intervjuerna i sin helhet, eller åt komplexiteten i deras bedömningsuppdrag som helhet.

I avsnitten som följer redovisas vår analys i två steg. Först tar vi upp hur lärarna själva förklarar varför nationella provbetyg och slutbetyg inte alltid överensstämmer. I nästa steg reder vi ut hur dessa förklaringar bildar underlag till ett påstående, nämligen att aggregerade resultat från nationella proven är irrelevanta för betygsuppdraget. I samband med detta diskuterar vi den syn på betygsättningsuppdraget och kriterierrelaterad bedömning som lärarna ger uttryck för, samt föreslår ett alternativt sätt att förstå de nationella provens roll i detta uppdrag.

Lärarnas förståelse av de nationella proven som betygsunderlag

Lärarna vi intervjuat gör tre huvudsakliga påståenden som problematiserar hur de nationella provresultaten förhåller sig till betygssättning: De nationella proven ger otillräcklig information i förhållande till slutbetyg, de nationella proven ger felaktig information i förhållande till slutbetyg och, slutligen, det sammanvägda provbetyget är inte relevant som underlag vid betygssättning. Först ska dock betonas att de lärare vi intervjuat är övervägande positiva till proven som betygsunderlag. De invändningar som undersöks närmare här är dels ofta uttryckta som ambivalens snarare än avfärdande, och dels är det just proven som likvärdighetsfrämjande (normerande) som ger upphov till dessa kluvna och ibland tydligt negativa utsagor.

“De nationella proven ger otillräcklig information i förhållande till slutbetyg”

I lärarnas resonemang om relationen mellan provresultat och betyg beskrivs proven som otillräckliga betygsunderlag då de inte täcker in allt – varken när det gäller innehållsområden eller kunskapskrav. Till exempel konstaterar en lärare att ”vi läser ju grannspråk och språkhistoria, det kommer ju ingenting på det” och hennes kollega instämmer ”nej, allt tas ju inte upp” och konstaterar samtidigt att ”i bedömningen när man sätter betyg” ska allt ”tas upp”. En annan lärare menar också att proven inte prövar läsning av exempelvis faktatexter och förmåga att läsa diagram och liknande och drar den tydliga slutsatsen att ”det är därför jag inte kan ta nationella prov, och säga du får det här betyget. Då skulle de bara gå dit och

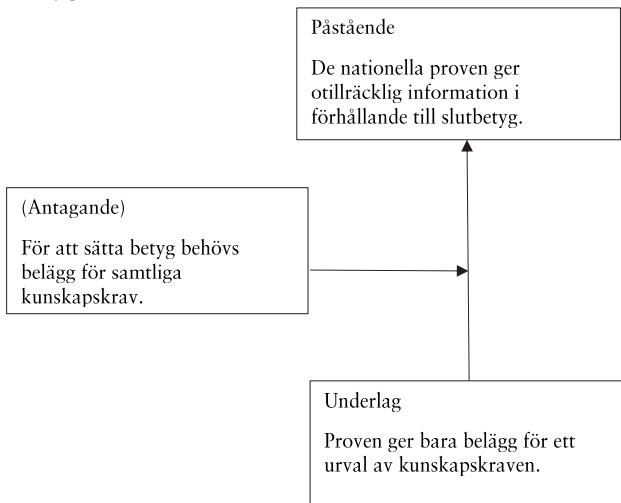
göra nationella prov och sen var det klart.” Med andra ord är man tydliga med att de nationella proven inte är examensprov, vilket är helt i linje med provens syfte. Att proven inte ”täcker allt” resulterar inte enbart i ett enkelt konstaterande att de inte utgör examensprov. Det används också specifikt för att förklara avvikelser mellan enskilda elevers provbetyg och deras slutbetyg. Så här resonerar två lärare:

Lärare 1: Om det är så att jag sätter ett annat betyg som slutbetyg i nian gentemot nationella provet, så skriver vi ner det och lämnar in här på expeditionen. Det finns ju de elever som kanske inte kommer hit och går på lektion och så skriver de ett nationellt prov och så ligger de på jättehöga betyg och sen har de inte presterat någonting annat under terminen.

Lärare 2: Det tycker jag är enklare att motivera nu. Därför att, det är ju så enkelt. Det finns fler kunskapskrav.

Uttryckt enligt Toulmins argumentationsmodell kan det grundläggande argumentet här beskrivas som följer:

Figur 3: De nationella proven ger otillräcklig information i förhållande till slutbetyg.



“De nationella proven ger felaktig information i förhållande till slutbetyg”

När lärare gör påståendet att provresultaten ger felaktig information som betygsunderlag är det två olika underlag som lyfts fram. Det första underlaget rör omständigheterna för genomförandet av proven.

Provens omfattning, både sett till delprov och den sammantagna mängden (del)prov som genomförs, innebär att vissa elever inte kan göra sig själva rättvisa:

[D]et är en specifik dag som man ska genomföra proven. Och svenskan i och med att läsförståelsedelen är ganska mastig för dem. Det är mycket frågor och att man då ska genomföra en skrivdel samma dag. Och det är två hela dagar på en vecka. Det är ju mycket för dem att orka. Och det tycker jag känns lite onödigt, för det är ju inte koncentration och ork vi ska testa. Men det blir ju lite så för en del elever, tycker jag.

Samma lärare berättar också att de gick igenom prov med låga resultat i svenska för att se vad det var ”som brast” för eleverna och ”var det orken i svenskan” så ”fångade vi upp dem på så sätt att de fick göra ett gammalt nationellt prov men med mer stöd, med mer tid”. Ur lärarens perspektiv blir alltså resultaten på provet när det skrivs enligt instruktionerna oanvändbart för vissa elever, det förlorar validitet i förhållande till kunskapskraven och eftersom det är provets funktion som betygsunderlag som ses som meningsfull förlorar också provet sin poäng. Problematiken kring hur olika provformer påverkar resultat och provens validitet (i meningen att proven mäter de kunskaper och förmågor som de avser mäta) är självklart legitim – detta är frågor som provkonstruktörer och Skolverket behöver arbeta med. När det gäller frågan om hur de nationella proven kan fungera likvärdighetsfrämjande inom ramen för betygsuppdraget är det mest intressanta dock att proven så tydligt är meningsfulla endast som betygsunderlag för en enskild elev.

Inte bara provens längd utan också den ovana och, anser många, stressande situationen ses som aspekter av genomförandet som gör resultatet för somliga elever missvisande. Provens genomföranderegler blir även relevanta när det gäller frågan om anpassningar. I sådana resonemang lyfter lärarna det faktum att provet inte ger rätt till samma typ av anpassning som eleven har rätt till under andra bedömningsgrundande situationer. Provresultatet speglar således elevens kunskaper och färdigheter utan eller med begränsade hjälpmedel vilket enligt lärarna gör att provet ger felaktig information i relation till slutbetyget.

Det andra underlaget som ligger till grund för påståendet om att de nationella proven ger felaktig information är att kravnivån, särskilt för betyget E, anses vara för låg i de nationella proven i framförallt engelska. Såhär resonerar exempelvis två lärare i årskurs 9 om nivån i dessa prov:

Lärare 1: E-nivån är väldigt låg.

Lärare 2: När man läser exempeltexterna för uppsatserna. De som ligger på D-C-nivå, det är vår normala E-nivå. För det är så pass mycket inblandad svenska i dom.

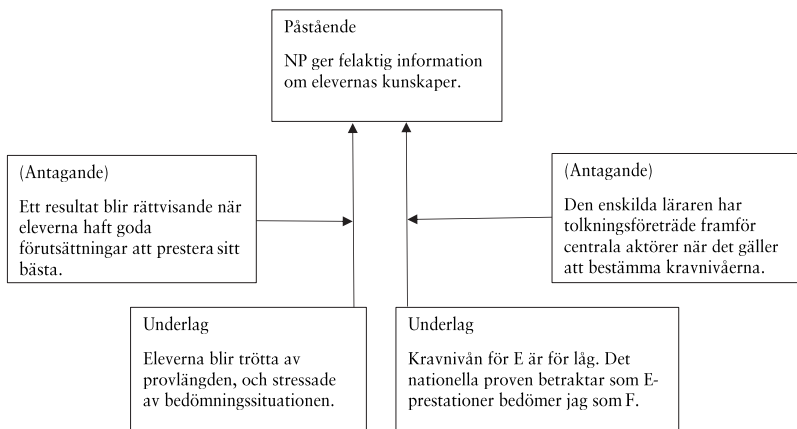
Lärare 1: Och det är ganska stora fel. Jag tror att det är trettio-tre procent man behöver ha för att få ett E [på den receptiva delen av nationella provet i engelska]. Och det är ju ganska lågt. Men det är ju muntliga också. Alltså, överlag alltihop där.

Lärare 2: Och där, där skiljer sig ju svenskan och engelskan. Otroligt mycket. För där tycker jag svenskan är högre på ett bra sätt.

Här anser lärarna att kraven för de olika betygsstegen på det nationella provet i engelska är för låga i förhållande till deras ”normala nivå.” Slutsatsen för dessa lärare blir enkel: eleverna får högre provbetyg än lärarnas bedömning säger att de ska ha i slutbetyg.¹

Sammanfattningsvis kan de olika argument som framförs vad gäller påståendet att proven ger felaktig information i förhållande till slutbetygen visualiseras som följer:

Figur 4: NP ger felaktig information om elevernas kunskaper.



“Det sammanvägda provbetyget är inte relevant som underlag vid betygssättning”

Det tredje och sista påståendet är att provbetyget inte är relevant som underlag vid betygssättning. Detta påstående grundar sig i att de

principer för hur man kommer fram till ett provbetyg och slutbetyg ser olika ut. Enligt Lgr11 innebär betygsättning att läraren gör en samlad bedömning av vilka kunskapskrav en elevs prestationer motsvarar. De bedömningsanvisningar som råder anger att för betygsstegen E, C och A gäller att samtliga kunskapskrav ska vara uppnådda. För att en elev ska få betyget E krävs alltså att samtliga kunskapskrav för betyget E uppnåtts. Om en elevs förmågor inom ett område bedöms motsvara F, medan elevens prestationer gentemot övriga kunskapskrav motsvarar E eller högre, ska eleven få betyget F. När lärarna i de nationella proven ska väga samman delprovresultaten till ett provbetyg ombeds de att göra en sammanvägning som följer andra principer. Här är sammanvägningsprincipen istället kompensatorisk och eleven kan exempelvis få betyget E på provet som helhet även om betyget F har uppnåtts på ett visst delprov. Att principerna skiljer sig åt framställs av flera lärare i våra intervjuer som märkligt, närmast oförklarligt. Konsekvensen blir, enligt lärarna, att det sammanvägda provbetyget blir obrukbart vid betygsättning medan resultaten på de olika delproven bidrar med information om hur eleverna presterar i förhållande till olika kunskapskrav:

Lärare 1: Tittar man bara på resultatet så blir det missvisande.

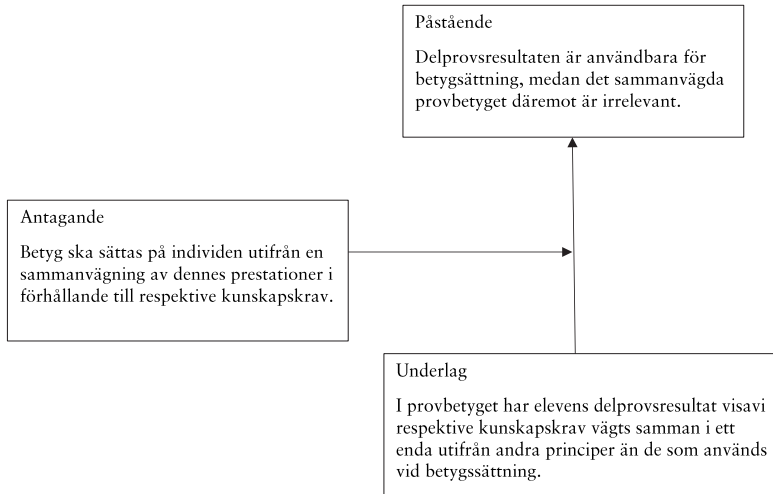
Lärare 2: Man går in och tittar på vad eleven presterar i varje delprov, delar upp det och ser, men den här har styrkor här och den har svagheter där. Okej den svagheten som då drar ner, hur ser det ut, var det just på det här provet, just den här dagen eller är det i vanliga fall att den skriver på den nivån också, både före och efter? För vi har flera texter, det blir givetvis att dom får skriva många olika texter om det nu var texterna som drog ner.

Delprovresultaten blir alltså den enda användbara informationen som proven ger lärarna i deras uppdrag att sätta betyg; vad gäller provens uppgift att vara betygsstödjande i förhållande till den enskilda eleven och dess olika förmågor i relation till kunskapskraven ser lärarna det sammanvägda provbetyg och de jämförelser det möjliggör som fullständigt irrelevant:

Det är olika förmågor. De borde snarare skippa det där [provbetyget]. Det tror jag är mer statistikerna som blir alldeles till sig så: "åh vi får ett provbetyg, vi kan jämföra". Jag vill ju se vad de är duktiga på i de olika delarna. Det är olika förmågor och det är så vi ska sätta betyg sen.

Argumentet om sammanvägningsprinciperna kan sammanfattas som följer:

Figur 5: Delprovresultaten är användbara för betygsättning, medan det sammanvägda provbetyget däremot är irrelevant.



Sammanfattningsvis är det alltså viktigt för lärarna att identifiera och hantera anledningar till avvikelser mellan provbetyg och slutbetyg i sin roll som betygsättare. De anledningar till avvikelser som lärarna i intervjumaterialet identifierar står inte på något sätt i strid med Skolverkets beskrivning av hur proven ska förstås och användas², nämligen att provresultaten bara utgör en del av lärarens samlade information om en elevs kunskap (Skolverket 2017a). Det växer också fram en bild av att i lärarnas förståelse är provets enda relevanta funktion att fungera som betygsunderlag för de elever vilkas prestationer lärarna ska betygsätta. Ovan såg vi en lärare uttrycka att det sammanvägda provbetyget endast kan användas till jämförelser som är intressant för ”statistikerna” men som är utan värde för lärarna i deras uppdrag som betygsättare. I nästa avsnitt ska vi utforska föreställningarna som rör hur provresultaten kan fungera i betygssättningsuppdraget närmare.

Lärarnas förståelse av de nationella proven som informationskälla på aggregerad nivå

I detta avsnitt inleder vi med att diskutera lärares föreställningar om vilka konsekvenser avvikelser mellan provbetyg och slutbetyg på individnivå har för resultat på gruppnivå. Vi utforskar sedan de före-

ställningar om betygssättningsuppdraget som lärarna ger uttryck för med fokus på vilka konsekvenser det får för hur de nationella proven kan användas likvärdighetsfrämjande i det uppdraget. Slutligen presenterar vi ett möjligt alternativt sätt att förstå de nationella provens roll i detta avseende, som bygger på en mer rättvisande förståelse av normers roll i ett kriterierelaterat bedömningssystem.

“De nationella provbetygen ska inte samvariera med slutbetygen”

Att lärarna i vårt intervjumaterial ser jämförelser mellan provbetyg och slutbetyg på grupp nivå som missvisande är tydligt; de förefaller anse att dessa två inte ska samvariera. De drar slutsatsen att eftersom välmotiverade avvikelser förekommer på individnivå så är statistik på grupp nivå missvisande – statistiken är i deras ögon ”korrupt”. Annorlunda uttryckt finns ett antagande här om att all avvikelse på grupp nivå mellan provbetyg och slutbetyg är en naturlig konsekvens av de valida avvikelserna på individnivå. Till exempel kan vi se att vad som upplevs som en avsaknad av rimliga anpassningar under de nationella proven innebär att provresultatet därmed inte bara blir oanvändbart som betygsunderlag utan att de prov som genomförs med vad som anses vara fel förutsättningar också i lärarens ögon gör statistiska jämförelser mellan provbetyg och slutbetyg ”skeva”:

Specialpedagogen sa att han [eleven] kan göra provet och sen efteråt så gör han provet liksom igen. Då får han göra om det men med andra förutsättningar. Men det som förs in i statistiken är ju det som skedde då, på riktiga provet. Om man tänker på statistiken, med betyget som ska jämföras med de här nationella proven, så blir ju det inte rätt heller då.

Även när de olika sammanvägningsprinciperna diskuteras uttrycker lärarna att utomståendes jämförelser blir missvisande:

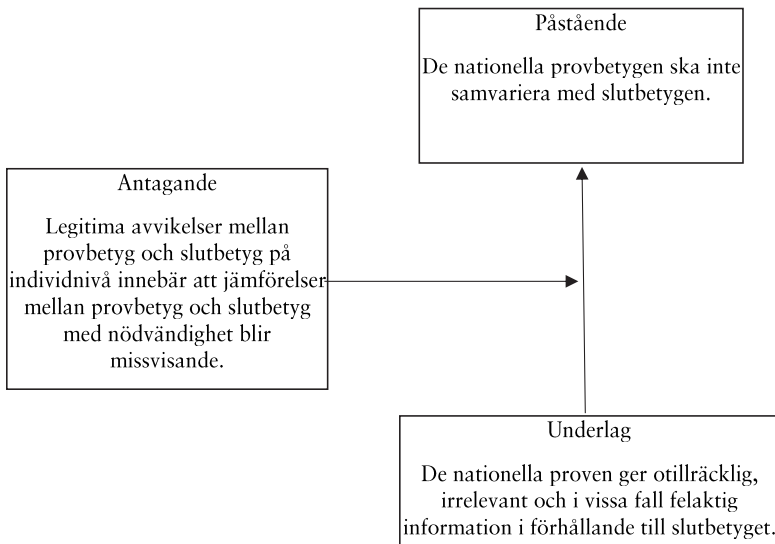
Lärare 1: Man har inte alla förutsättningar för att kunna göra granskningar. Man är inte helt insatt, utan man tittar bara på betyg, meritvärde kontra nationella prov.

Lärare 2: Och det är bara det sammanvägda provresultatet de ser. De ser inte dom olika delarna som vi gör.

Medan den enskilda läraren förstår och kan förklara och motivera eventuella avvikelser har externa aktörer inte förutsättningarna att göra detta. Eftersom varje enskild avvikelse kan förklaras kvalitativt

anses alltså inte variationer i klassers och skolors relation mellan provbetyg och betyg nödvändigtvis tyda på brister i en likvärdig betygssättning. Man kan sammanfatta lärarnas argument så här:

Figur 6: De nationella provbetygen ska inte samvariera med slutbetygen.



Även om de förklaringar till avvikelser som lärarna framför är principiellt helt riktiga är det ett felslut att på dessa grunder avfärda all analys på gruppnivå. Att proven inte prövar alla kunskapskrav gäller ju alla klasser som skriver proven. Detsamma gäller den kompensatoriska sammanvägningsprincipen. Likaså torde det finnas elever som blir stressade eller uttröttade i alla grupper. Att en lärares betygssättning ett givet år avviker mer från provresultaten än genomsnittet kan naturligtvis förklaras att klassen är överrepresenterad vad gäller exempelvis stresskänslighet eller uttröttningsnivå. Över tid kan också den lärare vars undervisning betonar andra kunskaper än de som prövas, tänkas ha en högre grad av avvikelse än de lärare vars undervisning ligger närmre proven. Men detta är snarare anledningar att faktiskt använda jämförelser på gruppnivå på ett omdömesfullt sätt, än motiv att avfärda dem.

Föreställningen att den kompensatoriska sammanvägningsprincipen i proven skapar, snarare än minskar, mätfelet är värd att diskutera närmare. En anledning till motståndet som lärarna uppvisar mot det sammanvägda provbetyget är att det orsakar problem i kommunikation med elever om hur provresultatet förhåller sig till slutbetyget. Så här resonerar två lärare om problemet:

Lärare 1: När du sitter ner och pratar med eleven, och så, ”ja men jag fick provbetyg b, wow jag har ett b”, nej, för ett b när du sätter betyg, betyder då har du nått allting på c och nått det mesta på a, men på provet så fick du ett b därför att du fick b och c. Alltså det väger upp åt andra hållet.

Lärare 2: Det är bättre och prata om, vad man får för betyg på dom olika provdelarna.

Lärare 1: Och sen säga att då måste vi jobba mer med det skriftliga här nu så du får visa mig att du kan komma upp i det högre. Ett klumpbetyg på ett prov blir svårare för eleverna att fatta för den ser bara att blev ett b nu.

Lärare 2: Det ger utrymme för mycket tolkningar.

Lärare 1: Ja

Lärare 2: Och dom är inte inne i det.

Lärare 1: Att ett a, ett b betyder nästan ett a.

Lärare 2: Nej, man får förklara.

Det sammanvägda provresultatet leder med andra ord till att lärarna får göra ett mer omfattande jobb för att förklara och motivera den egna betygsättningen i kommunikation med elever. Och eftersom lärarna ser de nationella provens funktion som betygsunderlag som primär blir dessa kommunikationsproblem viktigare än en mer rättvisande jämförelsepunkt på aggregerad nivå. Sådana jämförelser ser lärarna ändå inte som en del av sitt uppdrag. De görs istället av utomstående som, enligt lärarna, har en stark förväntan om att provbetyg och slutbetyg ska överensstämma: ”man ska ju fylla i betyg och betyg på nationella prov och man ska nästan förklara sig varför man inte håller samma”. När provresultatet vägs samman till ett provbetyg som avviker från slutbetyget upplevs provet alltså som en mätsticka som kan användas för att ifrågasätta lärarens betygsättning. Detsamma gäller när de sammanvägda provresultaten aggregeras, rapporteras in till Skolverket och publiceras i media. Omgivningens förväntningar om överensstämmelse mellan provbetyg och slutbetyg på aggregerad nivå upplevs som ett hot mot eller ifrågasättande av läraren som auktoriserad betygsättare:

Ja det ställer till det mer, för då slås det upp stort: ”oj vilka stora skillnader”. Ja, men ni vet ju inte varför. Jag gick ju direkt till

rektorn när jag hade sammanställt och sa att när tidningarna ringer så vill jag att du talar om det här är förklaringen. Och så sa jag, det här att man buntar ihop så blir det ett högre provbetyg, men man ska kunna alla streck.

Lärarnas kritik handlar om att den annorlunda principen skapar avvikelser som inte hade funnits om proven följt betygssystemets sammanvägningsprinciper. Denna argumentation är på sätt och vis förståelig. Inte desto mindre skjuter kritiken vid sidan av målet, såväl principiellt som praktiskt. Föreställningen att betygssättning och provresultat bör sättas utifrån samma principer är problematisk eftersom förordningens regler för betygssättning gäller just betygssättning, inte den enskilda bedömningen. Rent praktiskt eftersträvar lärarna en sammanvägning i proven som ger god överensstämmelse mellan provresultat och betyg. Detta skulle dock inte åstadkommas med en betygssystemenlig sammanvägning där ett F på ett enskilt delprov resulterar i provbetyget F. Istället skulle avvikelserna bli än större. För att belägga detta kan vi använda oss av svenskämnesprovets egen resultatstatistik, som varje år summerar ca 1500 elevers delprovresultat, provbetyg och höstterminsbetyg. Vi låter höstterminsbetyget vara den variabel som provresultaten ska samvariera med och jämför nuvarande kompensatoriska sammanvägning, med en sammanvägning som innebär att ett F på ett enda delprov ger F i provbetyg:

Tabell 1: Jämförelse mellan provresultat och höstterminsbetyg (Svenska ämnesprov åk 9) utifrån nuvarande respektive alternativ sammanvägningsprincip.

2013-2016	Nuvarande sammanvägning	“Betygssystemets” sammanvägning
Antal elever	Antal ”felaktigt höga” F-elever	Antal ”felaktigt låga” E eller mer- elever
4670	145	645
andel	3,1%	13,8%

Precis som lärarna ger uttryck för visar tabell 1 att nuvarande sammanvägning gör att vissa elever med betyget F de facto får E eller mer i provbetyg trots att de fått ett F på ett delprov. Närmare bestämt händer detta för 145 elever av de 4670 som ingår i statistiken. Med andra ord blir cirka 3,1% elever ”felaktigt” för högt bedömda på grund av den kompensatoriska sammanvägningen. Denna ”felaktiga” kategorisering ska dock ställas mot den som uppstår med en sträng, betygssystemsenlig sammanvägning där ett enstaka F ger F i provbetyg.

En sådan skulle nämligen få konsekvenser för en långt större andel elever. Närmare bestämt skulle 645 elever, eller 13,8% få F på proven trots att de på hösten fått betyget E eller mer. Med andra ord ger en kompensatorisk sammanvägningsprincip på provet resultat som ligger närmare lärarnas egen betygssättning.

Oavsett detta upplever lärarna sammanfattningsvis att de jämförelser som görs mellan provbetyg och slutbetyg används av andra aktörer, mot dem. Det förefaller heller inte finnas någon annan möjlig användning av dessa jämförelser i lärarnas förståelse av de nationella provens roll i betygssättningsuppdraget. Varför är det så?

Den individfokuserade bedömningen

Genomgående i alla intervjuer talar lärarna om betygssättning som något som görs i relation till en enskild elevs prestationer. Betygsättningsakten förstås precis på det sätt som visades i figur 1 – läraren värderar elevens prestationer utifrån kursplanens kunskapskrav, vilket resulterar i ett rättvist betyg. Betygsättningsuppdraget som helhet blir lika med summan av de individuella betygssättningsakterna och de nationella provens funktion blir att utgöra ett underlag för hur enskilda elever ligger till i relation till olika förmågor. I detta sätt att förstå provens betygsstödjande funktion blir det sammanvägda provbetyget, som följer de förmågespecifika prestationerna, ett störningsmoment. Jämförelser mellan provbetyg och betyg tillhör som vi såg tidigare ”statistikernas” domän, inte lärarnas, och deras egentliga drivkrafter tycks outgrundliga ur lärarperspektiv. Det sammanvägda provbetygets syfte att göra proven jämförelse- och analyserbara på gruppnivå är ett syfte som inte anses vara till nytta för den individuella läraren i betygssättningsuppdraget.

Till grund för den individfokuserade bedömningssynen ligger en föreställning om betygskriterier som oberoende av normer. Draget till sin spets blir betygskriterier fenomen som har en given inneboende innebörd. Att söka analysera och alls diskutera huruvida lärare i sina bedömningar kan tänkas ligga närmare eller längre ifrån en förväntad förståelse av betygskriteriernas innebörd blir därmed inte relevant. Understundom i intervjuerna framträder ytterligare föreställningar om kunskapskravens natur. En sådan är att kunskapskraven utgör en checklista som kan ”bockas av” gentemot enskilda prestationer. Denna avbockningslogik finns till exempel där som ett skuggargument när lärare framhåller att det faktum att de nationella proven inte täcker samtliga kunskapskrav är ett generellt gångbart sätt att förklara och motivera varför betyget inte blir detsamma som provresultatet. I detta

resonemangs logik ligger nämligen att avvikelser skulle bli svårare att förklara om proven verkligen prövade samtliga kunskapskrav. En annan föreställning som underbygger den individfokuserade bedömningssynen kommer till uttryck när lärarna pratar om bedömningsakten inom proven. Varje individs prestation ska ju värderas mot kunskapskraven och inte rangordnas i relation till andra elevers prestationer som i det tidigare relativa, eller grupprelaterade, betygssystemet. Att jämföra en prestation med en annan blir något som lärarna understundom känner sig nödgade till, samtidigt som det är något man nog borde avhålla sig ifrån:

[Det nationella provet i svenska] är svårbedömt, framför allt uppsatserna. På något sätt jämför man ju dom med varandra ändå, för det är svårt och bara utgå från exempeltexterna. Och det blir ju lite dumt, men det är ju så det blir.

Den gängse bedömningsdiskursen, som bland annat kommer till uttryck i Skolverkets råd kring betygssättning (Skolverket 2011), betonar bedömning som en kvalitativ handling där en individs prestationer värderas visavi kunskapskraven. Detta är, som framgått, en syn som lärarna i vårt material förefaller dela. Enbart vid ett enstaka tillfälle under intervjuerna tangerar samtalet frågan om normering som något meningsfullt också i ett mål- och kriterierelaterat system. Detta görs när en lärare resonerar om att det faktum att proven är utprovade möjligen gör att de ligger på ”rätt” nivå:

Men det är helt klart en vägledning, för på något vis så känns det som det är inget prov jag har gjort, utan det kommer från centralt prov, det här ska du testa, på det här sättet. Och då känns det på något vis att jag blir frikopplad. Vi gör det provet med alla, det är någon som har tänkt till, det är någon som har testat det här provet på barn förut, och hela den här biten, för det vet vi ju att dom har gjort. Så det känns ändå som en liten säkerhetsgrej på något vis. Jag vet inte hur jag ska uttrycka mig.

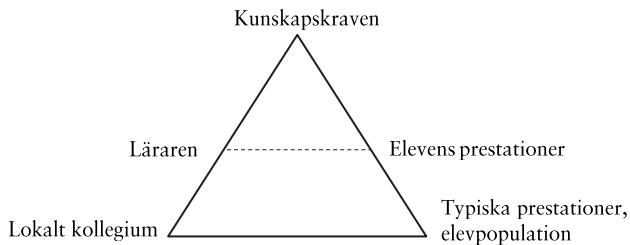
Att tanken rinner ut i en oförmåga att hitta ett sätt att formulera detta kan förstås som att vare sig normer eller mätsäkerhet tillhör diskursen om betyg i dagens skola. Detta, menar vi, är djupt problematiskt och värt att utforska närmare.

Diskussion – normers roll i ett kriterierelaterat bedömningssystem

Det är den auktoriserade läraren som har i uppdrag att värdera den individuella elevens olika prestationer visavi kunskapskravens skrivelser. Samtidigt bygger kunskapskraven i ämnen som svenska och engelska på värdeord, vilka, som nämnt i inledningen, får mening i relation till föreställningar om vad man kan förvänta sig av elever i en viss årskurs. Likvärdigheten i betygssättning av elever bygger på att det inte är den enskilda läraren som ensam har att bestämma vilka föreställningar och förväntningar på elevernas prestationer som de olika kriterierna är tänkta invokera. Istället bygger det kriterierelaterade betygssystemet på förekomsten av kollektiva och gemensamma uppfattningar om kunskapskravens faktiska innebörd i förhållande till den elevgrupp vars prestationer ska bedömas. Eller annorlunda uttryckt: kriterierelaterad betygssättning blir likvärdig då det råder en rimlig samsyn i lärarnas tolkningsgemenskap i termer av delade förväntningsnormer.

När lärarna i vårt material resonerar om behovet av tolkningsgemenskap för likvärdiga betyg görs detta nästan uteslutande i termer av vikten av sambedömning. Detta lyfts särskilt vad gäller delproven i skriftlig förmåga, men även de andra delproven upplevs av lärarna vara betjänta av sambedömning i olika grad. De ger tydligt uttryck för ett behov av att tillhöra en tolkningsgemenskap och en lokal sådan anses legitim och väl i linje med professionens autonomi och betygssättningsuppdraget. Med den lokala tolkningsgemenskapen i ryggen underlättas betygssättandet. Sambedömningens vikt i lärarnas föreställning om tolkningsgemenskap är självfallet inte begränsad till de nationella proven, utan en del av deras förståelse av uppdraget generellt. Man kan visualisera denna idé genom att låta triaden i figur 1 underbyggas av denna lokala tolkningskontext:

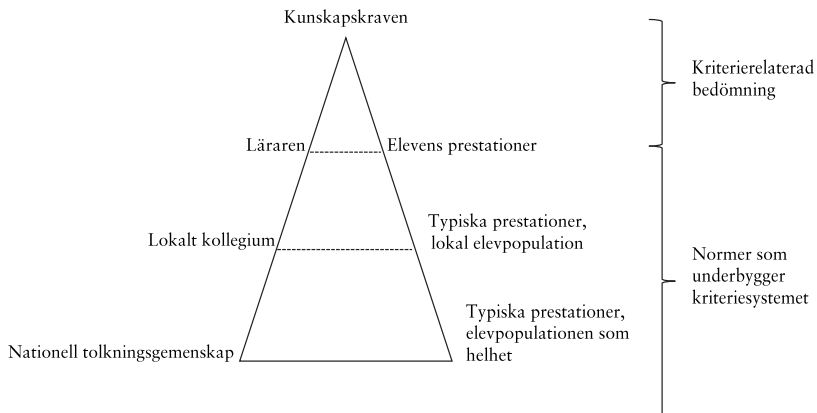
Figur 7: Sambedömningsidén och dess relation till bedömningsakten.



Idén om sambedömning understryker att den kriterierelaterade bedömningen är mer komplex än idealfallet och att den inte får bli godtycklig. Kollegor bör så långt som möjligt vara överens om kunskapskravens innebörder. Det är dessa kollegialt framförhandlade innebörder som blir kunskapskravens faktiska referenspunkter. Det som sällan uttrycks explicit i vare sig litteraturen eller lärarnas utsagor är vad sambedömning egentligen handlar om, nämligen att etablera de professionsgemensamma normer som behövs för att kriterierelaterad bedömning ska bli meningsfull. Istället förpassas normbegreppet till det tankegods som hör det normrelaterade betygssystemet till och som hotar likvärdigheten i det kriterierelaterade systemet.

Att lärarna söker utveckla tolkningsgemenskaper är självfallet positivt och en omistlig del i att skapa en professionell förståelse av vad som ska bedömas och av de normer som underbygger kunskapskraven (jfr Jönsson & Thornberg 2014). Dock kan det, menar vi, uppstå en likvärdighetsproblematik om dessa lokala tolkningsgemenskaper inte kalibreras mot varandra på nationell nivå. Likvärdighetsproblematiken riskerar att bli särskilt uttalad när dessa tolkningsgemenskaper formas i skolor med markant olika elevpopulationer. Ur ett likvärdighetsperspektiv är det alltså nödvändigt att vidga tolkningsgemenskaper så att de normerande föreställningar som möjliggör den kriterierelaterade bedömningen skapas i en kontext som motsvarar den i vilken bedömningssystemet är giltigt, med andra ord den nationella kontexten. Figur 8 visar hur bedömningsakten också inbegrips i en sådan vidare kontext:

Figur 8: Kriterierelaterad bedömning i ett sammanhållet bedömningssystem.



Idealt görs inte den enskilda bedömningsakten utifrån personliga eller privata preferenser, eller enbart som uttryck för vad som råkar

vara det lokala kollegiets samsyn. I likvärdighetsidén ligger att läraren idealt gör bedömningen i egenskap av att vara representant för ämneslärarkåren som helhet – det som i figuren kallas för den nationella tolkningsgemenskapen. Det är denna förståelse av den kriterierelaterade bedömningens vidare normerande underbyggnad som lyser med sin frånvaro både i de lärarintervjuer vi genomfört och i den vidare kursen om bedömning och betygssättning som lärarna är en del av.

När betygssättningsuppdraget ses i ett större perspektiv, där den enskilda läraren bedömer som representant för en nationell tolkningsgemenskap blir de kvantitativa analyser som möjliggörs av det sammanvägda provresultatet en viktig informationskälla. Att den enskilda bedömningsakten är kvalitativ till sin natur, innebär inte att kvantitativa underlag måste avfärdas. Snarare är dessa nödvändiga för att kunna få syn på sig själv som betygsättare. Genom en longitudinell analys över flera prov- och betygssättningscykler kan lärare identifiera mönster i termer av hur stor andel elever de brukar ge högre respektive lägre betyg än de fick på nationella provet i jämförelse med kollegor.

Att resultat på nationella prov och betyg vare sig ska eller bör överensstämma fullständigt betyder nämligen inte att relationen mellan måtten är ointressant. Med hjälp av Skolverkets resultatstatistik kan den enskilda läraren jämföra hur den egna samvariationen mellan provresultat och betyg ser ut över tid, och jämföra detta med genomsnittslärares dito i ämnet. På så vis kan läraren med hjälp av provresultat – som mätsticka i den egna handen – reflektera över hur hen positionerar sig i förhållande till den nationella tolkningsgemenskapen. Om det exempelvis skulle visa sig att läraren sätter påtagligt fler betyg som är högre än provresultaten än vad genomsnittsläraren gör, så är detta värt att fundera över. Naturligtvis kan en sådan avvikelse vara giltig – provkonstruktioners validitet kan vara olika gentemot olika elevgrupper, vår lärare kan vara extra skicklig att undervisa i andra områden i ämnet än de som provas etc. Men en tänkbar förklaring – som vore bekymmersam ur ett likvärdighetsperspektiv – är att avvikelsen beror på att vår lärare har en mer generös förståelse av kunskapskraven än lärarkåren i allmänhet i sin betygssättning. Det är naturligtvis viktigt för den enskilda läraren att bli medveten om hur hen positionerar sig som normbärare, och den som principiellt avfärdar kvantitativa analyser går miste om ett av få underlag som kan ge sådan information.

I vårt material kan vi se att lärarna uppvisar olika grad av villighet att justera sina betygsnivåer i förhållande till den information proven ger dem. Vissa av lärarna ger uttryck för värdet med en extern norm som nationella prov utgör. Samtidigt är det tydligt att man inte fullt ut givits förutsättningar att förstå det inomprofessionella värdet av avvikelseanalyser på klassnivå. Den omgivande kursen om det

kriterierelaterade systemets normlöshet gör att denna likvärdighetsfrämjande funktion med proven inte accepteras.

Slutsatser och implikationer

I den här artikeln har vi visat hur de intervjuade lärarna, utifrån det faktum att betygssystemet är mål- och kriterierelaterat, når slutsatsen att de inte är betjänta av aggregerade resultat från nationella proven i betygssättningsuppdraget. De antaganden som ligger till grund för denna slutsats är att betygsuppdraget i sin helhet är detsamma som summan av de enskilda, kvalitativa bedömningsakterna, samt att kunskapskravens innebörder antingen är inneboende i kravformuleringarna som sådana alternativt är upp till det lokala kollegiet att uttolka.

Det är inte märkligt att lärarna gör dessa antaganden – de har genomsyrt den svenska bedömningsdiskussionen alltsedan det mål- och kriterierelaterade systemet infördes. Inte desto mindre är antagandena och deras konsekvenser olyckliga. Värdeordsbaserade kriterier har ingen inneboende innebörd, och likvärdighetsidén omfattar skolsystemet som helhet. Mot denna bakgrund finns, menar vi, god anledning att tro att lärare skulle kunna använda kvantitativa jämförelser mellan sin egen betygssättning och de nationella provresultaten i kalibreringen mot den nationella norm som systemet förutsätter finns. Betygssättningsakten är tvivelsutan kvalitativ till sin natur, men att för den skull avfärda kvantitativa underlag som meningslösa är problematiskt ur likvärdighetssynpunkt. Betygssättningsuppdraget bör betraktas som större än summan av de enskilda bedömningsakterna och även inkludera en utvärdering av den egna betygssättningen. För detta ändamål är lärarna betjänta av nationella provresultat på aggregerad nivå. Man skulle kunna uttrycka det så att den mätsticka lärare upplever att utomstående granskare tar ifrån dem och använder mot dem, istället kan tas emot som ett inspel i lärarens betygssättningsuppdrag. Genom mätstickan får lärare enskilt och i grupp möjlighet att tala om, tolka och vid behov kalibrera sina egna normer och förväntningar. När resultaten inte utnyttjas är det bara exempellösningarna i provmaterialen som ger information om den förväntade kravnivån i den nationella tolkningsgemenskap som betygssystemet bygger på finns där. I ett omdömesgillt jämförande av aggregerade provresultat och betyg finns därför, menar vi, en underutnyttjad potential i de nationella provens likvärdighetsfrämjande funktion. Inte minst kan sådana analyser vara värdefulla för lärare i en skolverklighet där andra aktörer sätter press på dem att sänka

sina normer (jfr Mickwitz 2015). På samma vis som Skolverket (2014) uppmanar till och presenterar olika sätt att arbeta med sambedömning, bör de uppmanar till, och presentera, hur lärare kan arbeta med kvantitativa resultatmått.

För att detta ska åstadkommas behöver vi göra upp med några till synes djupt rotade föreställningar. En är att kriterier är absoluta per definition och att kriterierelaterad bedömning är fri från normer, en annan är att enbart kvalitativa informationskällor är relevanta för bedömningsuppdraget som helhet. I skiftet från ett normrelaterat till ett mål- och kriterierelaterat system och från en psykometrisk testkultur till en bedömningskultur kan dessa idéer ha fyllt sin funktion. Att omsätta en sådan systemförändring i praktiken är nämligen naturligtvis inget som görs över en natt. Det tar tid att göra upp med föreställningar om bedömningsakt och bedömningsuppdrag. I en övergångsperiod kan därför idealiserande beskrivningar av det nya systemet vara rimliga ur ett pragmatiskt perspektiv. Men när det av idealbeskrivningen följer att betygssättning enbart är den enskilda lärarens, eller det lokala kollegiets angelägenhet blir den kontraproduktiv. Tiden är därför inne att explicit prata om kriterierelaterad bedömning såsom den de facto fungerar, nämligen som beroende av normer. Det är först när vi ser detta som de nationella provens potentialer som likvärdighetsfrämjande kan utnyttjas full ut. I detta ligger, som vi visat, ingen jakt efter någon exakt överensstämmelse mellan proven och betygen. Istället bör de aggregerade provresultaten tas emot och tolkas för vad de är: som en av flera för betygsuppdraget meningsfulla informationskällor.

Noter

1. Men det förekommer också att lärarna berättar att de, i olika grad, justerar sina egna nivåer utifrån proven och att proven därmed till viss del fyller sin likvärdighetsfrämjande funktion.
2. Undantaget är de lärare som uttrycker att kravnivån för betyget E ligger för lågt.

Referenser

- Angoff, William (1974): Criterion-referencing, norm-referencing and the SAT. *College Board Review*, 92, 2-5.
- Jönsson, Anders & Thornberg, Pia (2014): Samsyn eller samstämmighet? En diskussion om sambedömning som redskap för likvärdig bedömning i skolan. *Pedagogisk Forskning i Sverige*, 19(4-5), 386-402.
- Kimbell, Richard (2007): E-assessment in project e-scape. *Design and Technology Education: An International Journal*, 12(2), 66-76.
- Klapp Lekholm, Alli (2008): *Grades and Grade Assignment: Effects of Student and School Characteristics*. Diss. Göteborgs Universitet.
- Koretz, Daniel (2008): *Measuring Up. What Educational Testing Really Tells Us*. Cambridge, Massachusetts: Harvard University Press.
- Korp, Helena (2011): *Kunskapsbedömning – Vad, hur och varför?* Stockholm: Skolverket.
- Lundahl, Christian; Hultén, Magnus; Klapp, Alli & Mickwitz, Larissa (2015): *Betygens geografi – Forskning om summativa bedömningar i Sverige och internationellt*. Delrapport från SKOLFORSK-projektet. Vetenskapsrådets rapporter. Stockholm: Vetenskapsrådet.
- Mickwitz, Larissa (2011): *Rätt betyg för vem? Betygsättning som institutionaliserad praktik*. Lic. Stockholms universitet.
- Mickwitz, Larissa (2015): *En reformerad lärare. Konstruktionen av en professionell och betygssättande lärare i skolpolitik och skolpraktik*. Diss. Stockholms universitet.
- Pope, Nakia; Green, Susan K.; Johnson, Robert L. & Mitchell, Mark (2009): Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education*, 25(5), 778-782.
- Räihä, Helge (2009): *Lärares dilemman*. Diss. Örebro universitet.
- Sadler, Royce D. (1989): Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Selghed, Bengt (2004): *Ännu icke godkänt. Lärares sätt att erfara betygssystemet och dess tillämpning i yrkesutövningen*. Diss. Malmö högskola.
- Simon, Marielle; Tierney, Robin D; Forgette-Giroux, Renée & Charland, Julie (2010): A secondary school teacher's description of the process of determining report card grades. *McGill Journal of Education*, 45(3), 535-554.

- Skolverket (2007): *Provbetyg – Slutbetyg – Likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolans årskurs 9, 1998–2006*. Stockholm: Skolverket.
- Skolverket (2011): *Allmänna råd för planering och genomförande av undervisningen*. Stockholm: Skolverket.
- Skolverket (2014): *Sambedömning i skolan. Exempel och forskning*. Stockholm: Skolverket.
- Skolverket (2015): *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011, reviderad 2015*. Stockholm: Skolverket.
- Skolverket (2016b): *Nationella proven i grundskolans årskurs 6 och 9: En uppföljning av lärares, rektorers och elevers uppfattning om proven*. Stockholm: Skolverket (Skolverkets rapport 447).
- Suurtam, Christine & Koch, Martha J. (2014): Navigating dilemmas in transforming assessment practices: Experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation and Accountability*, 26(3), 263-287.
- Toulmin, Stephen E. (2003): *The Uses of Arguments*. Cambridge: Cambridge University Press.
- Wikström, Christina (2006): Education and assessment in Sweden. *Assessment in Education: Principles, Policy & Practice*, 1(13), 113-128.
- William, Dylan (2010): What Counts as Evidence of Educational Achievement? The Role of Constructs in the Pursuit of Equity in Assessment. *Review of Research in Education*, 34, 254-284.

Elektroniska källor

- Skolverket (2016a): Provmaterial svenska åk 3. Hämtat 2016-11-03 <http://www.skolverket.se/bedomning/betyg/provbetyg/relationen-mellan-provbetyg-och-betyg-1.224495>
- Skolverket (2017a): *Nationella prov*. Hämtat 2017-09-13. <https://www.skolverket.se/bedomning/nationella-prov>
- Skolverket (2017b): *Betygsättning*. Hämtat 2017-09-21. <https://www.skolverket.se/bedomning/betyg>

