

Denna text skrevs på uppdrag 2018 för användning i Skolverkets utbildningsverksamhet. Den publiceras 2024 på NAFS-projektets hemsida efter medgivande från Skolverket. Observera att politiska reformer och beslut efter 2018 kan medföra att vissa sakförhållanden nämnda i texten påverkas.

Bedömning och betyg
Avsnitt 2: Likvärdighet genom nationella prov och bedömningsstöd

Extern bedömning av kunskap *eller* Om värdet av att se med andras ögon

Gudrun Erickson, Göteborgs Universitet

Syftet med denna text är att diskutera bedömning av kunskap och kompetens med särskilt fokus på vad utifrån kommande material och vad andra bedömare kan fylla för funktion. Texten börjar med en allmän beskrivning av bedömning som fenomen och dess funktion i utbildningen och fortsätter sedan med en kort historik kring centralt tillhandahållna prov i det svenska skolsystemet. Detta mynnar ut i en beskrivning av nuläget vad gäller det nationella provsystemet – dess syfte, material och användning. Frågor som behandlas är bland annat provens konstruktion, uppgiftstyper och bedömning, resultat per ämne och över tid, lärares synpunkter på proven, relationen mellan provbetyg och slut-/ kursbetyg samt stabilitet över tid. Vidare berörs den nyligen införda regeln om de nationella provens roll vid betygssättning, liksom planerade försöksverksamheter vad gäller avidentifiering av prov samt extern bedömning. Texten avslutas med en kort reflektion kring extern bedömning som sådan – om värdet av att få hjälp att se med andras ögon och att kombinera detta med egna, kontinuerliga iakttagelserna av elevers utveckling av kunskap och kompetens.

Bedömning som fenomen och funktion

Bedömning ingår som en naturlig del i allt lärande. Redan som små barn värderar vi våra färdigheter och framsteg, både på egen hand och påverkade av kommentarer från personer i vår närhet. När så skolan börjar, tas den bedömande och värderande rollen allt mer över av andra och blir successivt också mera formell till sin karaktär, inte minst när betyg börjar sättas. Vid sidan av detta försiggår också kontinuerlig kommunikation om kunskap mellan elever och lärare, där återkoppling ges under lärandets gång. Bedömning av kunskapsutveckling och kunskapsnivåer är därmed en del av skolans kärnuppdrag att stimulera och bidra till så mycket lärande som möjligt – och helst lite till. I detta finns flera aktörer, allt från eleverna själva, ensamma eller i grupp, till lärare och grupper av lärare. Till skaran hör också externa aktörer som bidrar med material och procedurer av skilda slag, i läromedel eller från nationella eller internationella system. Bakom alla dessa bedömningsvarianter finns ett syfte, som bör vara tydligt och giltigt och dessutom präglas av tillförlitlighet och respekt för individer och sammanhang.

Kunskapsbedömning kan sägas ha två grundläggande syften eller, annorlunda uttryckt, en dubbel funktion: den ska stödja både lärande och likvärdighet (Erickson & Gustafsson, 2017). Detta refereras ofta till som formativ respektive summativ bedömning (Scriven, 1967) eller bedömning för respektive av lärande (Black & Wiliam, 1998; Harlen, 2012). Det råder inget motsatsförhållande mellan dessa två funktioner, även om metoder och syften kan variera; båda bör samspela och komplettera varandra i den bedömningspraxis som

tillämpas. Det är också viktigt att tänka på att bedömning fyller en viktig funktion både för elevers lärande och lärares undervisning genom att verka för att (1) kartlägga hur elevers individuella kunskapsprofiler ser ut, med styrkor och svagheter, självförtroende, motivation och preferenser inom olika områden; (2) klargöra så tydligt som möjligt vilka målen är – vart man alltså är på väg, och (3) gemensamt mellan elev och lärare tänka ut vilka de bästa sätten är att ta sig så långt som möjligt i relation till dessa mål (se vidare *Assessment Reform Group/ARG*, 2002). Den mera summerande bedömningen fyller en funktion genom att sammanfatta små eller större delar av den aktuella kunskapen och på så sätt ge underlag både för det fortsatta arbetet och för en samlad bedömning, till exempel i form av betyg. Den kanske mest uppenbara summativa bedömningen är lärares betygssättning efter en termin eller en kurs, då ett antal iakttagelser av skilda slag ska summeras till ett enda omdöme. En genomtänkt och kontinuerlig bedömning i samverkan både med elever och med kolleger ger stöd för detta, liksom de nationella prov och bedömningsstöd som tillhandahålls av Skolverket.

Nationellt tillhandahållna prov – en kort historik

Det har funnits nationellt tillhandahållna prov i den svenska skolan sedan mitten av 1900-talet. Proven har varit, och är även idag, kopplade till rådande styrdokument och betygssystem och har på grund av detta genomgått ett antal förändringar. Ett viktigt dokument, som kan sägas vara startpunkten på det system vi har än idag, var slutbetänkandet från utredningen *Betänkande om betygssättningen i folkskolan*, (SOU 1942:11). Syftet med utredningen var att föreslå ett nytt betygssystem för folkskolan, som skulle ge så stabila och jämförbara betyg att urvalsproven till realskolan kunde tas bort. Utredningens ordförande, Frits Wigforss, hade under lång tid studerat betyg från olika klasser och kunnat se att variationen var stor och jämförbarheten låg. Han konstaterade också att lärare är skickliga på att identifiera och särskilja kvalitativa kunskapsnivåer i den egna klassen, men att förmågan att sätta nationellt likvärdiga betyg är betydligt sämre. Till följd av detta infördes successivt så kallade standardprov, vars syfte var att ge information om den egna klassens kunskapsmässiga nivå i relation till riket. I början av 1960-talet, i samband med införandet av grundskolan, infördes en femgradig betygsskala och en procentuell fördelning för respektive betygssteg baserad på normalfördelningen och avseende hela landet. Motsvarande system infördes i gymnasieskolan i och med 1965 års läroplan. En av nyheterna med den femgradiga skalan var att, de facto, ingen distinktion gjordes mellan godkända och underkända prestationer (Marklund, 1987).

Det relativa systemet kom att utsättas för stark kritik såväl inifrån skolan som utifrån. Systemet ansågs gynna konkurrens snarare än samarbete och den procentuella fördelningen av betygssteg användes felaktigt inom klasser och skolor. Vidare var innehållet i de olika betygsnivåerna inte synligt. Bland annat på grund av kritiken av det senare gav regeringen den dåvarande Skolöverstyrelsen (SÖ) i uppdrag att utreda betygsfrågan. Detta resulterade i ett förslag (Skolöverstyrelsen, 1972) till ett system med så kallade kursrelaterade betyg som skulle vara målrelaterade. Inom projektet *Målbestämning och utvärdering i skolan* (MUT) gjordes försök att utveckla ämnesvisa beskrivningar som var precisa nog att användas för

betygssättning. Kritiken mot förslagen var dock hård och pekade framförallt på att den detaljriktighet som var tydlig i förslagen skulle riskera att bli alltför styrande på undervisningen. SÖ lade därför 1975 ner MUT-projektet (Marklund, 1987). (Det är i detta sammanhang intressant att jämföra med en kommentar ur SOU 1942:11, s. 53, där följande konstaterades angående verbala definitioner av olika betygssteg: ”Om den diskuterade betygsmetoden skall kunna användas måste kunskapsmålen fixeras långt noggrannare än vad som nu är fallet, och det måste tydligt angivas, vilka fordringarna för godkännande skola vara. Det är en svaghet i den diskuterade metoden, att ju effektivare den blir ur betygssynpunkt desto större blir risken för skadlig inverkan på skolarbetet.”) Under åren efter MUT-projektet växte sig emellertid kritiken mot det relativa systemet starkare (se bl.a. Andersson, 1991) och i en departementsrapport från 1990 konstaterades att det relativa betygssystemet spelat ut sin roll och skulle ersättas med ett annat system.

Inom ramen för det relativa betygssystemet tillhandahölls standardprov för grundskolan och centrala prov för gymnasiet. Syftet med dessa var att resultaten skulle ge upplysning om klassers nivå i relation till riksgenomsnittet, medan de för enskilda elever inte skulle betraktas som viktigare än andra provresultat. Proven skulle också på ett rimligt sätt illustrera det ämnesinnehåll som fanns angivet i styrdokumentet, men detta var inte ett uttalat syfte. I början angavs fasta procenttal för respektive betygssteg, från vilket man som lärare inte fick avvika på något betydande sätt utan att motivera varför. Detta kom dock att frångås i grundskolan, där det under de senare åren enbart fastslogs att betygssteget 3 i normalfallet skulle vara det dominerande (Gustafsson & Yang Hansen, 2009; Marklund, 1987).

Sammanfattningsvis kan konstateras att det relativa betygssystemet under sin ca 30-åriga historia var mycket kritiserat, dels på grund av vad det stod för, dels - och inte minst - för det sätt på vilket det användes. Vad som dock också bör nämnas är att systemet trots detta i vissa avseenden hade positiva effekter, främst på grund av att betygen, stödda av de nationellt tillhandahållna proven, motsvarade högt ställda krav på likvärdighet mellan lärare, skolor och kommuner (Skolverket, 2009b). Betygen visade sig också ha god prognosförmåga vad gällde framgång i högre utbildning, såväl i övergången mellan grundskola och gymnasium som mellan gymnasiet och högre utbildning (Gustafsson, Cliffordson & Erickson, 2014).

Det nuvarande nationella provsystemet

Den svenska skolan förändrades på ett genomgripande sätt under det tidiga 1990-talet. Förändringarna berörde bland annat skolans organisation och styrning genom en omfattande decentralisering av ansvaret för genomförande av utbildning till kommuner och fristående huvudmän samt införande av en ny statlig skoladministration. Exempelen på förändringar är många, men som särskilt relevanta för den här aktuella texten kan nämnas införandet av nya läroplaner med en låg grad av central styrning (LpO 94 och LpF 94) och nya kursplaner för ämnen, ett mål- och kunskapsrelaterat betygssystem samt ett nytt nationellt provsystem.

I det följande kommer några centrala aspekter av det nationella provsystemet i det nuvarande, målrelaterade systemet att beröras. Utifrån syftet att de nationella proven ska ge det avsedda, externa stödet för betygssättning med så god relevans och precision som möjligt och därmed bidra till rättvisande och likvärdiga betyg, är områden som material och användning fundamentala ur kvalitetssynpunkt. Dessa områden, liksom provens syften, är också sådana som vållat och vållar mycket diskussion.

Syften med nationella prov

Proven i det normrelaterade, relativa systemet hade ett enda syfte, nämligen att fungera som riktninggivare för betygssättning på gruppnivå. De hade i denna bemärkelse en styrande funktion. Vid sidan av detta förutsattes de också vara innehållsligt rimliga i relation till styrdokumentet, men detta var inget officiellt syfte. I samband med övergången till det mål- och kunskapsrelaterade systemet i mitten av 1990-talet beslutades också om ett nytt nationellt provsystem, som för det första innehöll ett bredare spektrum av material, för det andra fick helt andra syften än föregångaren. I detta spelade den implementerande funktionen en viktig roll, dvs. att proven skulle konkretisera styrdokumentet inte minst genom att operationalisera kursplanernas ämnesbeskrivningar. Vidare skulle de hjälpa lärare att diagnostisera elevers svårigheter. Proven funktion var därmed avsedd att vara stödjande, inte styrande för betygssättningen.

Denna utveckling av proven var starkt präglad av den vidgning av bedömningsbegreppet som var tydlig internationellt och som kom att bli omtalad som en strävan att använda bedömning för att främja lärande, inte enbart för att fastställa nivåer (Black & Wiliam, 1998; Gardner, 2012; Gipps, 1994). Den mätande funktionen var däremot påtagligt nedtonad i de första direktiven för det nya systemet, i synnerhet vad gällde grundskolan. Det fanns också en påtaglig oro att proven skulle komma att styra lärares bedömning och betygssättning för mycket, varför betygsgränserna i proven togs bort inför läsåret 2001/2002. Protesterna mot detta blev dock kraftiga, både från politiskt håll och från lärare, varför gränserna kom tillbaka tämligen snabbt. I provuppdraget från Utbildningsdepartementet till Skolverket 2004 fanns även en viss markering av likvärdighetssyftet; fem syften för de nationella proven definierades, nämligen att de skulle:

- bidra till ökad måluppfyllelse för eleverna
 - förtydliga målen och visa på elevers starka och svaga sidor
 - konkretisera kursmål och betygskriterier
 - stödja en likvärdig och rättvis bedömning och betygssättning
- samt
- ge underlag för en analys av i vilken utsträckning kunskapsmålen nås på skolnivå, på huvudmannanivå och på nationell nivå.

Frågan om hur många syften ett prov eller provsystem kan eller bör ha är ofta och mycket diskuterad, såväl från teoretiska som praktiska utgångspunkter, och något entydigt svar

finns inte. Klart är dock att olika syften ställer olika krav på provens utformning. Vad gäller de svenska nationella proven har de länge haft tre övergripande syften: betygsstödjande, utvärderande och formativa. Mycket kort kan sägas att det betygsstödjande syftet har med rättvisa och likvärdighet att göra och ställer mycket höga krav på giltighet och trovärdighet, det vill säga validitet och reliabilitet. Ett utvärderande syfte förutsätter en mycket god så kallad domäntäckning, det vill säga att proven innefattar så mycket som möjligt av respektive ämnes kursplan och att det finns möjlighet att följa resultat över tid. Ett formativt syfte har bland annat att göra med att proven ska bidra till ökad måluppfyllelse för eleverna, vilket handlar om innehåll, tidsaspekter, återkoppling med mera. Att kombinera allt detta med lika hög kvalitet i enskilda prov innebär uppenbarligen stora utmaningar och låter sig svårligen göras, vare sig principiellt eller inom de ramar som finns. Detta har i olika sammanhang påpekats både från Skolverket och olika forskare och finns också diskuterat på ett generellt plan i internationell litteratur kring prov och bedömning (se till exempel Gustafsson, Cliffordson och Erickson, 2014; Hattie och Timperley, 2007; Newton, 2007; Skolverket, 2003; SOU 2016:25).

Om de första ca tio åren med de nya nationella proven präglades av implementerande ambitioner och oro för alltför mycket styrning, kan tiden därefter sägas illustrera något av det motsatta: kritiken mot de mångfacetterade syftena tilltog (se t.ex. SOU 2007:28) och rättvis- och likvärdighetsaspekter kom att lyftas fram alltmer. Sammanfattningsvis kan utvecklingen sägas vara att färre syften successivt blivit överordnade och att de som tidigare haft denna status har placerats under en rubrik som säger att de nationella proven också kan bidra till vissa av de tidigare huvudsyftena, dvs. uttryckt som en möjlighet, inte som ett direktiv. Detta gäller även idag, hösten 2018, då endast en funktion, eller ett syfte, anges på Skolverkets hemsida, nämligen att de nationella proven ”är ett stöd för en likvärdig och rättvis betygssättning av en elevs kunskaper”. I anslutning till detta konstateras vidare att ”genom analyser av provresultaten i relation till uppnådda kunskapskrav på skolnivå, huvudmannanivå och på nationell nivå kan de nationella proven bidra till att stärk skolornas kvalitetsarbete”.

Vad gäller de nationella provens roll för lärare i deras arbete har formuleringen ’stöd för betygssättning’ funnits med från 1990-talet. Detta är en tydlig markering av att proven inte är examina i den traditionella bemärkelsen; provresultaten är alltså ingalunda tänkta att vara det enda underlaget för individuella betyg utan ska kombineras med lärarens kontinuerliga iakttagelser. Någon precisering eller kvantifiering av detta stöd har dock inte gjorts förrän 2018, då en skrivning om att provresultaten ska ”särskilt beaktas” vid betygssättningen införts. Formulering är hämtad från SOU 2016:25, i vilken en genomlysning gjordes av frågan om betygsstöd, liksom ett flertal andra kring nationella prov. Vi återkommer till detta senare i texten, under rubriken Utredningar och åtgärder.

Material i de nationella proven

Inför skiftet till ett målrelaterat betygssystem på 1990-talet fördes diskussioner om hur och i vilken omfattning de nya nationella proven skulle komma att påverkas av förändringen.

Åsikterna gick isär, delvis beroende av vilket ämne som avsågs och i vilken utsträckning enskilda kursplaner förändrats gentemot tidigare. Medan det stundtals hävdades att helt nya sätt att pröva krävdes, framfördes också synpunkten att det inte var uppgifterna i sig som behövde genomgå någon radikal förändring utan i stället tolkningen och användningen av provresultaten.

Centrala frågor i diskussionen av provtyper handlade dels om hur frågor skulle ställas, dels om hur eleverna skulle avge sina svar. Den första frågan knöt på ett uppenbart sätt till kursplanernas beskrivning av kunskap, inte minst till relationen mellan mera faktainriktade frågor och sådana som fokuserade på förståelse, färdighet och förtrogenhet, detta i enlighet med de förarbeten som gjorts till de nya styrdokument (SOU 1992:94). Frågan om responssätt avsåg i vad mån elever skulle producera egna svar eller välja bland ett antal alternativa möjligheter, och i så fall på vilket sätt. En aspekt av den senare frågan var i vad mån flervalstuppgifter kunde fånga de olika kunskapskvaliteter som beskrevs i kursplanerna. En annan diskussion gällde om språket som sådant skulle komma att utgöra ett väsentligt bedömningsunderlag i uppgifter med öppna svar, snarare än den kunskap som avsågs att fokuseras. Någon konsensus kring dessa frågor nåddes inte, utan de olika provgrupperna valde sina egna lösningar. Den övergripande effekten av systemskiftet var dock att väsentligt fler så kallade performansuppgifter (även kallade autentiska uppgifter) infördes, dvs. komplexa uppgifter där elever skulle prestera tämligen omfattande skriftliga eller muntliga svar, inte välja ett korrekt svarsalternativ, vilket var vanligt i det relativa systemet. De autentiska uppgifterna kunde inte sällan knytas på ett ganska direkt sätt till kursplanernas formuleringar. Detta blev ännu tydligare efter införandet av nya styrdokument 2011, i vilka ämnenas innehåll konkretiserades på ett mera explicit sätt än tidigare. Det fanns dock effekter av ökningen av autentiska uppgifter, som var mindre positiva från mätsynpunkt, nämligen dels att antalet uppgifter minskade, vilket ökar risken för mätfel, dels att svaren krävde kvalitativ bedömning av lärare, vilken visade sig kunde variera på ett påtagligt sätt. Det senare har påpekats bland annat av Skolinspektionen, som på regeringens uppdrag utrett användningen av de nationella proven från 2009 och framåt (Skolinspektionen, 2017). Det är dock viktigt att påpeka att andra studier visat på väsentligt bättre överensstämmelse mellan bedömare. Denna fråga är alltför komplex för att fokusera på i denna text, men det kan konstateras att inte minst metoden att studera överensstämmelse har en stor betydelse för vad som ses och hävdas (Gustafsson & Erickson, 2013; Skolverket, 2009a).

Framtagande av nationella prov

De olika nationella proven utvecklas vid olika universitet och högskolor i Sverige på uppdrag av Skolverket. Gemensamt för de olika processerna är nära samverkan med olika grupper av experter, inte minst lärare och elever. Vidare företas olika utprövningar av allt tänkbart provmaterial. Dessa utprövningar varierar i omfattning, vilket också gäller de metoder som används för att analysera resultaten. Utprövningarna har dock som gemensamt, övergripande syfte att öka kvaliteten i de prov som ges genom att låta de verkliga brukarna använda dem och kommentera dem från olika synvinklar. Ett annat exempel på samverkan är att de olika betygsgränser som följer med proven tas fram i samarbete med grupper av aktiva lärare tillsammans med erfarna provutvecklare och forskare. Denna så kallade kravgränssättning, eller

standard setting, går till så att lärare 'gör' uppgifterna och analyserar dem dels utifrån kurs- eller ämnesplanens skrivningar om innehåll och krav, dels på basis av sina egna erfarenheter av elevers kunskapsnivåer. De ger därefter individuella förslag på gränser för olika betygssteg, vilka sedan diskuteras i hela gruppen. Det hela kompletteras sedan med olika kvalitativa och kvantitativa analyser som har gjorts på utprovingsdata. Detta förfaringssätt i vid bemärkelse används ofta även internationellt för identifikation och fastställande av olika kvalitativa nivåer.

Slutligen kan nämnas att resultaten på de nationella proven presenteras dels som ett sammanslaget provbetyg, dels i nedbruten form, till exempel i profiler, som synliggör olika aspekter av elevernas resultat i relation till respektive kurs-/ämnesplan. Sammanvägningen till ett provbetyg görs på olika sätt och enligt delvis olika principer i de olika proven, vilket är något som fokuseras i det systemramverk som tagits fram för de nationella proven och som behandlas senare i texten. Viktigt att påpeka är att provbetyget är en sammanvägning av elevens delresultat på det nationella provet enligt en given modell, vilket följaktligen inte är detsamma som den process det innebär när läraren i enlighet med läroplanernas riktlinjer vid betygssättningen 'utnyttjar all tillgänglig information om elevens kunskaper i förhållande till de nationella kunskapskraven och gör en allsidig bedömning av dessa kunskaper'.

Användning av de nationella proven

De svenska nationella proven är, som redan nämnts, inte avsedda att fungera som examensprov som ensamma avgör enskilda elevers slutgiltiga betyg. Däremot är syftet att de ska utgöra ett så gott stöd som möjligt för lärare vid betygssättning. En aspekt av detta är givetvis att acceptansen för proven ligger på en hög nivå, vilket enligt olika undersökningar också tycks vara fallet.

Lärares synpunkter på proven

Skolverket redovisade 2014 en omfattande sammanställning och analys av de olika lärarenkäter (drygt 22 000) som sändes in i samband med 2013 års prov. Resultaten stämmer väl överens med de årliga analyser som görs av de olika provgrupperna vid de institutioner som utvecklar proven. Vid den tidpunkt då Skolverkets undersökning gjordes var frågorna i respektive lärarenkät inte identiskt formulerade, men det är ändå fullt möjligt att ringa in frågeområden med samma innehåll. I synnerhet tre frågor är här av största intresse, nämligen en allmän om lärares uppfattningar om provet som helhet, en om relationen till den aktuella kurs-/ämnesplanen och slutligen en med fokus på den grad av stöd lärare anser sig ha av proven och resultaten. Vad gäller samtliga dessa frågor visar svaren att i regel väl över 90 % av lärarna var positiva, såg en tydlig relation mellan proven och de underliggande styrdokumenterna samt ansåg sig få gott stöd vid betygssättning. Detta gällde i synnerhet de prov som funnits under en längre tid, medan vissa av de mera nyligen introducerade proven i natur- och samhällsorienterande ämnen i grundskolan uppvisade något lägre siffror, dock fortfarande på en hög nivå (över 70 %). Vidare kunde noteras att de allra flesta lärare var nöjda med de betygsgränser som angavs för de olika proven, dock föga förvånande med vissa variationer mellan delprov och prov.

I Skolverkets undersökning 2014 var det ytterst få lärare som var uttalat kritiska till proven som sådana. Däremot gavs många positiva kommentarer. Kritiska synpunkter förekom givetvis också, dock främst vad gällde frågor om den arbetsbörda proven medförde. Under senare år har emellertid också andra aspekter av de nationella proven kommit i fokus, inte minst i relation till den bestämmelse som införts i samband med 2011 års läroplansreform vad gäller elevers slut-/kursbetyg och som innebär att alla kunskapskravens delar ska vara uppfyllda för att elevens ska få ett visst betyg i det aktuella ämnet. Flerdimensionalitet i kunnande, eller tydligare uttryckt, ojäma/'taggiga' profiler inom ämnens ram har följaktligen negativa effekter i betygssammanhang. Detta kritiserades från flera håll redan på ett tidigt stadium (Gustafsson m.fl., 2014; Vlachos, 2013) och identifierades även som problematiskt av Skolverket i en bred utvärdering om den nya betygsskalan samt kunskapskravens utformning, som publicerades 2016. Vissa lärare upplever även att relationen mellan denna s.k. tröskelregel och de nationella proven är problematisk, eftersom proven tillämpar en kompensatorisk princip, som innebär att en svag prestation kan vägas upp av en starkare inom ramen för ämnets innehåll. Frågan är vid skrivandet av denna text (nov 2018) under utredning och förändringar kan komma att ske.

Analys av provresultat

En förutsättning för att de nationella proven ska fylla sin funktion som stöd för lärares betygssättning, förutom att de mottas väl av brukarna, är att de håller hög och stabil kvalitet, såväl inom prov som över tid. Detta innebär att de ska ha tydlig så kallad *construct coverage* samt uppvisa goda mätmässiga egenskaper. Det förra handlar om att proven så väl som möjligt ska reflektera kurs-/ämnesplanernas beskrivning av den aktuella kunskapsmängden (*the construct*), det senare om att resultaten som genereras är stabila och trovärdiga indikatorer på elevers kunskaper. Sammantaget kan proven därmed bidra till rättvisa och likvärdighet. I arbetet att utveckla material som lever upp till detta måste i synnerhet två risker beaktas, nämligen dels att ett för litet eller för smalt underlag av ämnets olika angivna delar inkluderas, dels att något annat än ämnets angivna innehåll står i fokus. Dessa två företeelser refereras i facklitteratur till som *construct under-representation* respektive *construct irrelevant variance* och anses vara de två huvudsakliga hoten mot provens giltighet, eller validitet (Messick, 1989; Skolverket, 2017).

Provresultaten kan studeras på olika sätt och från olika synvinklar, till exempel på basis av elevers prestationer inom och mellan ämnen samt med fokus på stabilitet över tid. Vad gäller elevers prestationer, dvs. deras uppvisade måluppfyllelse, finns uppenbara skillnader mellan ämnen. Detta torde till stor del kunna relateras till den bakomliggande kurs- eller ämnesplanen, inklusive kunskapskraven som de nationella proven är satta att pröva, men provens utformning och de olika årskullarnas egenskaper har givetvis också betydelse.

I Tabell 1 visas utfallet av de nationella proven i grundskolans åk 9 mellan åren 2013 och 2017 i form av det sammanvägda provbetyget, vars maximala värde är 20.0. Observera att ämnesgrupperna varierar i storlek¹.

¹ Samtliga elevers resultat samlas in för de olika ämnena, vilket innebär > 90 000 för engelska, matematik och svenska. Eleverna gör ett av tre no-prov och ett av fyra so-prov (slumpvis tilldelning till skolor), varför antalet

Tabell 1. Sammanvägt provbetyg (Pb) för nationella prov i årskurs 9, 2013–2017

Ämne	Pb 2013	Pb 2014	Pb 2015	Pb 2016	Pb 2017	Medelvärde per ämne
Engelska	14.9	15.0	15.1	15.2	15.0	15.0
Matematik	12.2	11.4	10.7	12.1	10.8	11.4
Svenska	13.3	13.6	13.5	13.6	13.6	13.5
Svenska som andraspråk	9.2	9.4	9.2	8.9	7.7	8.9
Biologi	10.9	12.5	11.9	13.0	12.4	12.1
Fysik	11.9	11.4	11.8	13.1	13.2	12.3
Kemi	11.2	11.4	13.3	13.3	12.6	12.4
Geografi	13.0	13.1	12.6	13.2	13.1	13.0
Historia	11.2	12.4	10.7	12.9	12.5	11.9
Religionskunskap	11.3	12.8	13.8	13.5	13.0	12.9
Samhällskunskap	13.8	12.6	12.9	15.0	13.3	13.5
Medelvärde samtliga prov/år	<i>12.1</i>	<i>12.3</i>	<i>12.3</i>	<i>13.1</i>	<i>12.5</i>	12.5

Trots vissa svagheter, som olika antal resultat per ämne samt varierande bortfall, visar tabellen på flera intressanta företeelser. En sådan är den tydliga skillnad i resultat mellan ämnena, som varit ungefär densamma sedan det målrelaterade systemets införande på 1990-talet. Om man ser till de tre ämnena som funnits med i provsystemet sedan starten, nämligen engelska, matematik och svenska, är det uppenbart att skillnaden i resultat mellan engelska och matematik är betydande och konsekvent över tid. Detta kan tolkas på olika sätt och bör diskuteras utifrån olika utgångspunkter: kunskapsteoretiska, pedagogiska, metodiska osv. Frågor som kan ställas är till exempel om matematik är ett mera kognitivt krävande ämne än språk, i detta fall engelska i en svensk kontext; om ämnenas kursplaner skiljer sig åt vad gäller struktur, kravnivåer, antal aspekter i kunskapskraven etc; om proven i engelska är för lätta/proven i matematik för svåra i relation till sina respektive kursplaner; om lärarna lyckas olika väl i de två ämnena; om

resultat per prov i dessa fall är väsentligt lägre. Svenska som andraspråk studeras av < 10 000 elever. Trots dessa skillnader kan antalet resultat per ämne sägas vara tillräckligt för att intressanta observationer ska kunna göras och vissa slutsatser dras.

läromedlen är olika vad gäller närhet till styrdokumentet; om exponering för ämnena utanför skolan skiljer sig åt på något väsentligt sätt; om traditionerna i de båda ämnena är olika på viktiga sätt osv (Erickson, 2017). Något enkelt svar på dessa frågor torde inte gå att finna utan snarare kombinationer av möjliga förklaringar, som är kontextkänsliga på olika sätt.

Antalet provtagare i ämnet svenska som andraspråk är väsentligt lägre än i övriga ämnen men dock fullt tillräckligt för att generera viktig information. En fråga är givetvis varför resultaten är så påtagligt mycket lägre än i övriga ämnen och varför de dessutom tycks ha sjunkit markant under den senaste tiden. I denna diskussion är förstås diverse kontextuella faktorer väsentliga, liksom ramfaktorer i vid bemärkelse (allt från till exempel tid och resurser till attityder och organisationsformer).

Slutligen kan, vad gäller resultat som indikation på måluppfyllelse, konstateras att de tre proven i naturorienterande ämnen uppvisar likartade resultat, i synnerhet under de senaste åren, medan skillnaderna är något större mellan de samhällsorienterande ämnena, där historia har de lägsta resultaten och samhällskunskap de högsta.

Frågan om stabilitet respektive variabilitet mellan år är komplex och innehåller aspekter av styrdokumentens utformning i olika ämnen, provutveckling, inte synnerhet vad gäller art och grad av standardisering, analyser och inte minst fastställande av kravnivåer (Erickson, 2017). Skillnaderna mellan ämnen är här tämligen stora, med minst variation vad gäller proven i svenska, engelska och geografi. Den största variationen återfinns i so- och no-ämnena (med undantag av geografi och i någon mån fysik). Detta ska till viss del förstås mot bakgrund av att dessa prov introducerades betydligt senare än de andra och därmed inte haft samma tid att utveckla rutiner t.ex. för kravgränssättning. Dessutom kan situationen påverkas av hur kunskapskraven är utformade.

Provresultat och betyg

Dagens nationella prov är avsedda att stödja betygssättning genom att fungera som ett komplement till lärarens allsidiga bedömning av elevernas kunskaper. Den fråga som från början ställts angående detta syfte är olika varianter av 'Vad är lagom stöd?', en fråga som följaktligen kräver precisering och kanske i någon mån även viss kvantifiering. Några svar har dock inte givits, vilket innebär att det egentligen bara finns en typ av artikulerad funktion som är mot regelverket, nämligen användning av provresultaten som ensamt avgörande vad gäller betyg. Denna typ av användning förbjuds de facto av läroplanernas olika formuleringar om allsidighet, i Lgr 11 uttryckt på följande sätt (avsnitt 2.7, Bedömning och betyg: Riktlinjer):

Läraren ska

- vid betygssättningen utnyttja all tillgänglig information om elevens kunskaper i förhållande till de nationella kunskapskraven och göra en allsidig bedömning av dessa kunskaper.

Samma information, något mera precis, ges i Lgy 11 (2.5, Bedömning och betyg: Riktlinjer):

Läraren ska vid betygssättningen

- utnyttja all tillgänglig information om elevens kunskaper i förhållande till de nationella kunskapskrav som finns för respektive kurs,
- beakta även sådana kunskaper som en elev har tillägnat sig på annat sätt än genom den aktuella undervisningen, och
- utifrån de nationella kunskapskrav som finns för respektive kurs allsidigt utvärdera varje elevs kunskaper.

Av stort intresse är givetvis lärares uppfattningar om och användning av provresultaten i den samlade bedömningen. Ett sätt att studera detta är att jämföra de samlade provbetygen med slut- respektive kursbetygen per ämne. Data av denna typ är tillgängliga på Skolverkets hemsida och har under ett tiotal år publicerats i samband med rapporteringen av proven. Eftersom inget regelverk funnits för användningen av provbetyget, har det stundtals ifrågasatts varför statistiken publiceras och även att den ibland använts av Skolinspektionen i granskningar av skolor. Dock anses informationen som en intressant indikation på det nationella provsystemets funktionalitet och som en aspekt av den komplexa frågan om rättvisa och likvärdighet i bedömning av kunskap och kompetens.

I Tabell 2 visas relationen mellan det samlade provbetyget (Pb) och lärarnas slutbetyg (SB) i åk 9 vårterminen 2017. I mittspalten anges den procentuella andelen fall där de två betygen sammanföll, i vänsterspalten där slutbetyget var lägre än provbetyget och i högerspalten när eleverna fick högre slutbetyg än provbetyg.

Tabell 2. Samlat provbetyg (PB) i relation till lärares slutbetyg (SB); åk 9, vt 2017

<i>ÄMNE</i>	<i>SB < PB</i>	<i>SB = PB</i>	<i>SB > PB</i>
Antal prov			
Engelska n = 92 140	13 %	74 %	13 %
Matematik n = 91 856	1 %	60 %	39 %
Svenska n = 67 663	9 %	61 %	30 %
Svenska som andraspråk n = 8 385	5 %	60 %	35 %
Biologi n = 28 247	6 %	60 %	34 %
Fysik n = 29 705	11 %	67 %	22 %
Kemi n = 31 443	7 %	65 %	28 %

Geografi n = 24 257	5 %	66 %	29 %
Historia n = 24 276	6 %	57 %	37 %
Religionskunskap n = 22 572	6 %	64 %	30 %
Samhällskunskap n = 21 696	7 %	64 %	29 %

Som framgår av tabellen är elevernas slutbetyg i åk 9 identiskt med provbetyget i mellan 57 och 74 % av fallen, med ett medelvärde på 64 %. Detta visar i sig att en tydlig majoritet av lärarna anser att provresultaten bekräftar deras egna bedömningar, alternativt att de justerat betygen efter provresultaten. När man analyserar de fall där lärarens betyg inte är detsamma som provbetyget – vilket de givetvis inte behöver vara i ett system där provens funktion är stödjande – framgår dock att eleverna i samtliga ämnen utom engelska får ett högre slutbetyg än provbetyg. Detta är extra tydligt i matematik, där endast 1 % av eleverna får ett lägre betyg, medan 39 % får ett högre, men skillnaderna är stora även i historia, svenska som andraspråk och biologi. I fysik är skillnaden mellan lägre respektive högre slutbetyg väsentligt mindre än i andra ämnen, medan det i engelska är lika många elever som får ett lägre som ett högre slutbetyg i relation till provbetyget.

Utfallet av analyserna av slutbetyg i relation till provbetyg är tämligen likartat över tid, och genom åren har olika förslag till förklaringar av fenomenet framförts. Dessa handlar bland annat om att eleverna markant ökat sina kunskaper efter proven och fram till betygssättningen i slutet av terminen, medan andra innebär att de delar av kursplanerna som inte ingår i proven behärskas bättre av eleverna, vilket återspeglas i slutbetygen. Ingen av dessa förklaringar förefaller dock särskilt trolig. Tiden efter proven tycks alltför kort för att eleverna på något markant sätt skulle kunna öka sina kunskapsnivåer, och mönstret av väsentligt högre slutbetyg än provbetyg är tydligt, relativt oberoende av vad i kursplanen som tillfälligtvis inte ingår i proven. Detta sammantaget pekar på ett behov av ytterligare undersökningar för att bättre förstå bakomliggande orsaker. Inte minst behöver diskrepansen mellan lärares svar på lärarenkäter, där en stor majoritet anser att proven väl speglar styrdokumentet och stämmer väl överens med den egna bedömningen, och deras betygsbeslut studeras.

Slutligen kan nämnas att mönstret i gymnasieskolan mycket liknar utfallet i grundskolans åk 9: även här gavs betydligt fler högre kursbetyg än provbetyg, i detta fall i samtliga ämnen med nationella prov (två prov i engelska, nio i matematik, två i svenska och två i svenska som andraspråk). Överensstämmelse mellan de två betygen fanns i mellan 53 och 78 % av fallen (medel 65 %). Lägre kursbetyg än provbetyg gavs i genomsnitt till 4 %, högre kursbetyg än provbetyg till 31 % av eleverna. Den minsta skillnaden mellan andelen lägre och högre provbetyg fanns i kurserna Engelska 5 och 6, den högsta i kursen Matematik 2B.

Utredningar och åtgärder

Som tidigare nämnts, har de nationella proven under senare år successivt kommit att kritiserats från flera håll, till största delen på ett systemrelaterat snarare än ett innehållsligt plan. Skolinspektionen har fokuserat på bristande överensstämmelse mellan bedömare; lärare och skolledare har påtalat att proven innebär en mycket stor arbetsbörda, och forskare och andra har riktat stark kritik mot den icke-kompensatoriska princip för betygssättning, som indirekt även kom att skapa osäkerhet kring de nationella provens funktion. OECD (Nusche m.fl., 2011) identifierade i en genomlysning av det svenska bedömningssystemet både positiva och negativa drag och förordade framförallt en ökad grad av standardisering av de nationella proven samt åtgärder för att öka reliabiliteten. Det senare skulle bland annat åstadkommas genom utveckling av ett gemensamt ramverk för det nationella provsystemet. Regeringen beslutade i samband med detta att tillsätta en utredning för att belysa ett antal frågor kring provens funktion samt även föreslå åtgärder för att förtydliga och förbättra systemet. Resultaten av utredningen redovisades i SOU 2016:25 med titeln 'Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning'. Rapporten bestod av ca 800 sidor och var uppdelad i två volymer, den andra med fokus på relationen mellan de nationella proven och lärares betygssättning. Utredningen har efter en omfattande remissomgång resulterat i ett antal konkreta åtgärder och beslut. Ett av dessa innebär att de nationella provresultaten från och med höstterminen 2018 ska 'särskilt beaktas' vid betygssättning. Detta förklaras på följande sätt i Skolverkets Allmänna råd (2018, s. 37):

”Provresultatet ska inte helt styra betyget utan vara ett stöd vid betygssättningen. Resultatet på ett nationellt prov kan alltså inte vara det enda underlaget, men det går inte att bortse från resultatet om det inte finns särskilda skäl för det. Anledningen är att nationella prov är välkonstruerade, noga utprovade och har utvecklats just för att vara ett stöd vid betygssättningen. Resultat på ett nationellt prov har därför större betydelse än andra enskilda underlag vid lärarens allsidiga utvärdering av elevens kunskaper vid betygssättningen.”

En ytterligare åtgärd som föreslogs i SOU 2016:25 är en försöksverksamhet med extern bedömning. En sådan kommer att genomföras i anslutning till den digitalisering av provsystemet som ska vara avslutad 2022. I samband med detta ska proven avidentifieras och någon annan än elevens lärare ska bedöma proven. Båda dessa åtgärder syftar till att motverka det som stundtals refereras till som positiv *bias*, dvs. att läraren påverkas – inte sällan i positiv riktning – av att veta vilken elev som har gjort provet. Frågan om detta är en realitet och i så fall i vilken omfattning det påverkar resultaten har inget självklart svar, varför det är viktigt att i försöksverksamheten pröva modeller av skilda slag. Kopplat till detta finns också tankar om medbedömning, dvs. att två lärare bedömer samma prov². I slutänden är det dock givetvis viktigt att den betygssättande läraren genom de olika åtgärder som vidtas får ytterligare information och stöd för sin slutgiltiga betygssättning, där de

² För information om olika förändringar och förslag, se Skolverkets hemsida.

nationella proven fyller en viktig funktion, inte bara resultatmässigt utan också vad gäller olika aspekter av den mångfacetterade kunskap som beskrivs i ämnenas styrdokument.

Tanken på ett systemramverk för de nationella proven, alltså ett regelverk som konformerar och standardiserar delar av systemet har funnits sedan början av 2000-talet och arbete inleddes under 2014 med att skapa ett sådant, som skulle fungera som en övergripande modell för de olika nationella proven. Ett förslag, utvecklat av en grupp forskare med lång erfarenhet av bedömning och prov, antogs av Skolverket i december 2017 och används för närvarande för utveckling av ramverk för individuella nationella prov. Avsikten med dessa provramverk är att såväl utvecklingsprocesser och analyser som produkter i form av uppgifter och prov i olika ämnen ska få en mera standardiserad utformning för att tjäna sitt syfte på bästa sätt. Detta innebär bl.a. att reliabiliteten ska ytterligare öka och variabiliteten minska, så att proven ännu bättre än hittills kan bidra till rättvisa och likvärdighet.

Att se med andras ögon

Lärare i Sverige har ett mycket stort och tämligen unikt ansvar för bedömning och betygssättning av elevers kunskaper, vilket är en ytterst ansvarsfull uppgift i synnerhet som betyg i stor utsträckning används för urval till högre utbildning. I andra länder är lärares betyg oftast underordnade externa examina av skilda slag, vars resultat är de som gäller inför beslut vad gäller framtida arbete och utbildning³. Det är följaktligen av yttersta vikt att lärare och skolor i Sverige erbjuds olika typer av stöd i sitt viktiga arbete. Det nationella provsystemet ska ses i ljuset av detta.

I diskussionen om aktörer, brukar man tala om fyra typer av bedömning i skolan, nämligen elevernas individuella självbedömning; den bedömning som görs av och mellan elever (så kallad kamratbedömning eller *peer assessment*); den som utförs av elevernas lärare och slutligen den som kommer utifrån, den externa bedömningen. Alla dessa former bidrar med väsentlig information, både för pedagogiska och rättvise- och likvärdighetsskapande syften. De står inte heller i något motsatsförhållande till varandra, även om syften och metoder varierar och de två senare är de som har en officiell funktion vad gäller betygssättning.

Det finns flera sätt att tala om bedömning och betygssättning, från de strikt vetenskapliga till de myndighets- och förordningsrelaterade och till de metaforiska. Ett exempel på det senare kan vara att beskriva vikten av olika aktörer i bedömningsprocessen som ett försök att, med ett lånat uttryck från ett litet barn, ”bäras åt” (Erickson, 2009). Om man som lärare skulle vara den enda aktören skulle bördan bli blytung, men med hjälp av andra och varandra blir situationen betydligt lättare. Likaså kan man diskutera extern bedömning från olika perspektiv. Många studier visar att bedömningars reliabilitet, eller trovärdighet, befrämjas av att mer än en bedömare är involverad – att fler ögon ser och öron hör. Detta är den stora skillnaden, inte att någon annan ensam gör en bedömning och att resultatet per automatik därmed blir mera ’sant’ eller tillförlitligt. Det stämmer också med många lärares erfarenheter,

³ För information om olika länders skolsystem, se [Eurydice, https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en](https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en)

att man med hjälp av sam-/medbedömning får hjälp att se med andras ögon och därmed blir varse aspekter man kanske inte annars skulle ha uppmärksammat.

Sett över tid, har det svenska nationella bedömnings- och provsystemet genomgått olika stadier, där syftet med den externa bedömningen varierat. Genomgående har dock lärares bedömning av de egna elevernas kunskaper och kunskapsutveckling haft en central roll. Detta är detsamma idag, medan vissa förändringar har gjorts vad gäller det utifrån kommande stödet. Den nya regeln om att resultaten på nationella prov ska tas i särskilt beaktande vid betygssättning samt den planerade försöksverksamheten med extern bedömning och medbedömning är aktuella exempel på detta, liksom införandet av ett ramverk för de nationella provens utformning. Allt detta sammantaget syftar till att ytterligare förstärka kvaliteten i bedömning och betygssättning och därmed bidra till ökad rättvisa och likvärdighet i den svenska skolan.

Referenser

- Andersson, H. (1991). *Relativa betyg. Några empiriska studier och en teoretisk genomgång i ett historiskt perspektiv*. Doktorsavhandling, Pedagogiska institutionen, Umeå universitet.
- Assessment Reform Group (2002). *Assessment for Learning: 10 principles. Research-based principles to guide classroom practice*. Hämtad 18 sept 2018 från <https://www.aiaa.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Pbi Delta Kappan*, 80(2), 139–148. Hämtad 18 sept 2018 från https://www.michigan.gov/documents/mde/Inside_The_Black_Box_-_Black_and_William_601104_7.pdf
- Erickson, G. (2009). ”Att bäras åt” – Om den goda bedömningens flerfaldighet och ömsesidighet. I U. Tornberg m. fl., *Språkdiradaktiska perspektiv. Om lärande och undervisning i främmande språk* (s. 159-173). Stockholm: Liber.
- Erickson, G. (2017). Experiences with Standards and Criteria in Sweden. I Blömeke, S. & Gustafsson, J-E., *Standard Setting in Education. The Nordic Countries in an International Perspective* (s. 123-142). Cham, Switzerland: Springer International Publishing AG.
- Erickson, G. & Gustafsson, J-E. (2017). Bedömningens dubbla funktion – för lärande och likvärdighet. I Lundgren, U. P; Säljö, R. & Liberg, C. (red.), *Lärande, skola, bildning* (4:e utg., s. 551–581). Stockholm: Natur & Kultur.
- Gardner, J. (Ed.) (2012). *Assessment and Learning* (Second edition). London: Sage.
- Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: The Falmer Press.
- Gustafsson, J-E. & Yang Hansen, K. (2009). Resultatförändringar i svensk grundskola. I Skolverket, *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer*, (Kap. 3, s. 39–83). Stockholm: Skolverket. Hämtad 18 sept 2018 från <https://www.skolverket.se/publikationer?id=2260>
- Gustafsson, J-E. & Erickson, G. (2013). To trust or not to trust? – teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability*, 25(1), 69-87.
- Gustafsson, J-E., Cliffordson, C. och Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter*. Stockholm: SNS förlag. Hämtad 18 sept från <https://www.sns.se/aktuellt/likvardig-kunskapsbedomning-i-och-av-den-svenska-skolan-problem-och-mojligheter/>
- Harlen, W. (2012). On the Relationship Between Assessment for Formative and Summative Purposes. I Gardner, J. (red.), *Assessment and Learning* (second edition), s. 87–102. London: Sage Publications Ltd.

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

Marklund, S. (1987). *Skolsverige. Del 5: Läroplaner*. Stockholm: Liber.

Messick, S. A. (1989). Validity. I Robert L. Linn (Ed.), *Educational Measurement* (Third edition, s. 13-103). New York: American Council on Education/Macmillan.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2), 149-170.

Nusche, D., Halász, G., Looney, J., Santiago, P. & Shewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education – Sweden*. OECD: Paris. Hämtad 18 sept 2018 från <https://www.oecd.org/sweden/47169533.pdf>

Scriven, M. (1967). *The Methodology of Evaluation*. Washington: American Educational Research Association.

Skolinspektionen (2017). *Bedömningsprocessernas betydelse för likvärdigheten. Ombedömning av nationella prov 2016*. (Redovisning av regeringsuppdrag Dnr U2014/7535/GV). Hämtad 18 sept 2018 från <https://www.skolinspektionen.se/sv/Beslut-och-rapporter/Publikationer/Granskningsrapport/Omrattning-av-nationella-prov/bedomningsprocessernas-betydelse-for-likvardigheten/>

Skolverket (1994). *Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet Lpo 94*. Stockholm: Skolverket.

Skolverket (1994). *Läroplan för de frivilliga skolformerna (Lpf 94)*. Stockholm: Skolverket.

Skolverket (2003). *Det nationella provsystemet – vad, varför och varhän?* Dnr 2003:2038.

Skolverket (2009a). *Bedömaröverensstämmelse vid bedömning av nationella prov*. Stockholm, Skolverket (Dnr 2008:286).

Skolverket (2009b). *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer* (s. 40-85). Stockholm, Skolverket. Hämtad 18 sept 2018 från <https://www.skolverket.se/publikationer?id=2260>

Skolverket (2011; rev. 2018). *Läroplan för grundskolan samt för förskoleklassen och fritidshemmet*. Stockholm: Skolverket

Skolverket (2011). *Läroplan för gymnasieskolan*. Stockholm: Skolverket.

Skolverket (2014). *Så tycker lärarna om de nationella proven 2013*. Stockholm: Skolverket (Rapport 404). Hämtad 18 sept 2018 från <https://www.skolverket.se/publikationer?id=3218>

Skolverket (2016). *Utvärdering av den nya betygsskalan samt kunskapskravens utformning*. Stockholm: Skolverket (Dnr 2014:892). Hämtad 18 sept 2018 från <https://www.skolverket.se/publikationer?id=3652>

Skolverket (2017). *Skolverkets systemramverk för nationell prov*. Stockholm: Skolverket. Hämtad 18 sept 2018 från <https://www.skolverket.se/publikationer?id=3890>

Skolverket (2018). *Betyg och betygssättning*. Skolverkets Allmänna Råd med kommentarer Stockholm: Skolverket. Hämtad 5 nov 2018 från <https://www.skolverket.se/publikationer?id=4000>

Skolöverstyrelsen (1972). *Betygsättningen i grundskola och gymnasieskola*. Stockholm: Skolöverstyrelsen.

SOU (1942:11). *Betänkande med utredning och förslag angående betygssättningen i folkskolan*. Stockholm: Ecklesiastikdepartementet. Hämtad 18 sept 2018 från https://weburn.kb.se/metadata/980/SOU_8219980.htm

SOU (1992:94). *Skola för bildning. Betänkande av läroplanskommittén*. Stockholm: Utbildningsdepartementet.

SOU (2007:28). *Tydliga mål och kunskapskrav i grundskolan - Förslag till nytt mål- och uppföljningssystem*. Stockholm: Utbildningsdepartementet. Hämtad 18 sept 2018 från <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2007/05/sou-200728/>

SOU (2016:25). *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning*. Stockholm: Utbildningsdepartementet. Hämtad 18 sept 2018 från <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2016/03/sou-201625/>

Vlachos, J. (2013). Betygssystem som bäddar för misslyckanden. *Ekonomistas*, juni 2013.