

*Jan-Eric Gustafsson*  
Göteborgs universitet  
*Gudrun Erickson*  
Göteborgs universitet

DOI: <http://dx.doi.org/10.5617/adno.6434>

## Nationella prov i Sverige – tradition, utmaning, förändring

### **Sammanfattning**

*Nationella prov har funnits i den svenska skolan sedan mitten av 1900-talet. Syftet med artikeln är att ge en överblick över de förändringar proven genomgått i relation till de olika styrdokument och betygssystem de varit kopplade till. Grundfrågan i texten är på vilket sätt, i vilken utsträckning och hur väl proven förmått möta de utmaningar som dessa förändringar krävt. Detta gäller i synnerhet övergången från ett norm- till ett målrelaterat system, med frågan om likvärdighet i bedömning och betygssättning i fokus. Även frågan om syfte/syften med de nationella proven har en central roll, liksom svenska lärares unika ansvar för bedömning och betygssättning av de egna elevernas kunskaper. Artikeln beskriver och diskuterar slutligen den aktuella situationen, då förändringar är på väg vad gäller regelverk och principer för de nationella provens utveckling, användning och funktion. I detta kan dels en statlig utredning, dels införandet av ett systemramverk för proven komma att få betydelse.*

*Nyckelord: det svenska skolsystemet; nationella prov; styrsystem; förändring; bedömning; kvalitet*

## National tests in Sweden – tradition, challenge, change

### **Abstract**

*In the Swedish school system, national tests have been in existence since the mid 1900s. The aim of this article is to give an overview of the gradual changes of these tests related to the different national curricula and grading systems that they have been connected to. The core question is in what way, to what extent, and how well the tests have managed to meet the challenges required by these developments. In particular, this deals with the transition from a norm- to a goal-referenced system, with the question of equity in assessment and grading in focus. In addition, the issue of aim/aims of national tests is an important aspect,*

*as is the unique responsibility of Swedish teachers in assessing and grading their own students' competences. Finally, attention is paid to the current situation, when several changes are underway regarding explicit regulations and principles for the development, use and function of the national test. In this, a governmental inquiry and the introduction of a common framework for national tests may become important.*

*Keywords: the Swedish school system; national tests; national regulations; change; assessment; quality*

## Inledning

Den svenska skolan har haft nationellt tillhandahållna prov sedan mitten av förra seklet, alltså ca 70 år. Före läroplansreformen 1994 rubricerades dessa som standardprov (i grundskolan) och centrala prov (i gymnasiet), medan samlingsbeteckningen nationella prov använts därefter. Systemet har under denna tid genomgått både större och mindre förändringar, och det har även de betygssystem som de nationella proven haft en nära koppling till. Denna artikel har som syfte att ge en överblick över förändringarna i systemen med nationellt tillhandahållna prov och hur dessa förmått att möta de utmaningar som förändrade styrdokument och betygssystem skapat. Detta gäller särskilt övergången från ett norm- till ett målrelaterat system, med likvärdighet i bedömning och betygssättning som den största utmaningen. Frågor kring de nationellt tillhandhållna provens syften har också en central roll, liksom svenska lärares ansvar för bedömning och betygssättning av de egna elevernas kunskaper, vilket är ett sätt att organisera bedömning som knappast finns i andra länder.

## Det relativa betygssystemet

En av de mest grundläggande och insiktsfulla texterna kring prov och betyg som publicerats i Sverige är slutbetänkandet från utredningen SOU 1942:11 ("Betänkande om betygssättningen i folkskolan"). Utredningsuppdraget var att föreslå ett system för betygssättning i folkskolan som gav så jämförbara betyg att urvalsproven till realskolan skulle kunna avskaffas. Utredningens ordförande, Frits Wigforss, som var seminarielektor i matematik vid lärarutbildningen i Kalmar, hade under lång tid utvecklat och undersökt prov i matematik. Han hade särskilt fokuserat på utveckling av diagnostiska prov att användas av lärarna för det som idag kallas formativ bedömning (Kilpatrick & Johansson, 1994). Men Wigforss hade även undersökt i vilken grad olika lärares betygssättning var jämförbar och dragit slutsatsen att det fanns stora skillnader i betyg mellan

klasser som hade samma provresultat. Han konstaterade att lärarna har god förmåga att särskilja olika prestationsnivåer bland elever i den egna klassen, men att de inte har tillräcklig kännedom om hur andra klasser i landet presterar för att kunna sätta betyg som är nationellt likvärdiga. Som en lösning på detta problem föreslog utredningen att så kallade standardprov skulle införas med hjälp av vilka lärarna kunde få information om de egna elevernas prestationsnivå jämfört med övriga elever i landet. Tanken var dock inte att de enskilda elevernas betyg skulle grundas på deras provresultat, utan att provresultaten skulle vara vägledande för klassens betygsfördelning och baseras på en sammanvägning av all tillgänglig information, som exempelvis resultat på andra prov, muntliga prestationer i klassrummet, liksom resultatet på standardprovet.

Systemet med standardprov infördes successivt under 1940- och 1950-talen och gav då stöd åt den sjugradiga bokstavs-betygsskalan, som för folkskolans del introducerats år 1897, genom rekommenderade procentandelar för de olika betygsstegen. Vid införandet av grundskolan år 1962 ändrades betygs-skalan till en femgradig sifferskala med en föreskriven procentuell fördelning över betygsstegen baserad på normalfördelningen. Samtliga betygssteg angav godkända prestationer, vilket innebar att distinktionen mellan underkända och godkända prestationer enligt tidigare betygs-skalor avskaffades. Ett relativt betygssystem, som i all huvudsak var identiskt med det som infördes för grundskolan 1962, infördes i gymnasiet i och med 1965 års läroplan.

Det relativa betygssystemet utsattes för stark kritik från såväl lärare och skoladministration som från politiker och allmänhet. En kritikpunkt var att lärarna inte förstod att använda systemet som det var tänkt, utan istället applicerade procentfördelningen inom klasser utan hänsyn till resultaten på standardprov. En annan invändning var att systemet inbjöd till konkurrens snarare än till samarbete. Ytterligare punkter för kritik var att betygen inte sade något om nivån på kunskaper och färdigheter, i det att betygen behöll samma medelvärde år från år, oberoende av om kunskapsnivån ökade eller minskade. Inte heller tydliggjordes kunskapsinnehållet i de olika nivåerna. En hel del kritik avsåg dock snarare betyg generellt än det relativa systemet specifikt, och det framfördes att det inte skulle komma att finnas behov av betyg för urval från grundskolan till gymnasieskolan eftersom varje elev skulle kunna få sitt förstahandsval tillgodosett (SOU 1977:9).

Kritiken mot det relativa betygssystemet medförde att Skolöverstyrelsen (SÖ) fick i uppdrag av regeringen att utreda betygsfrågan. SÖ (1972) föreslog ett system med s.k. kursrelaterade betyg, som skulle vara målrelaterade. Inom ramen för SÖ-projektet "Målbestämning och utvärdering i skolan" (MUT) gjordes försök att utveckla målbeskrivningar i så precisa termer att de skulle kunna användas som underlag för betygssättning. Kritiker menade dock att så detaljerade målbeskrivningar skulle få en alltför styrande inverkan på undervisningens uppläggning och SÖ lade år 1975 ner MUT-projektet. I en lång rad utredningar under de kommande decennierna diskuterades det relativa

betygssystemets brister (se Andersson, 1991), dock utan att något konkret alternativ utvecklades. I en departementsrapport (Ds 1990:60) konstaterade man dock att det relativa betygssystemet spelat ut sin roll och att det skulle ersättas med ett annat system.

Trots all kritik mot det relativa betygssystemet framstår det i vissa avseenden som framgångsrikt. Framförallt uppfyllde de relativa betygen högt ställda krav på likvärdighet mellan lärare, skolor och kommuner (Skolverket, 2009a), och det var också kravet att betygen skall vara jämförbara som Wigforss i SOU (1942:11) identifierade som den största utmaningen. Grundskolebetygen hade också god prognosförmåga av studieframgång i gymnasieskolan, och gymnasiebetygen predicerade studieframgång i högre utbildning bättre än antagningsprov som Högskoleprovet (se Gustafsson, Cliffordson & Erickson, 2014, för en forskningsöversikt). Avnämares tilltro till de relativa betygen var också hög, vilket återspeglas i att några krav på införande av avgångsexamina från grund- eller gymnasieskola inte framförts. I den allmänna kritiken av det relativa betygssystemet har det hävdats att det ledde till en deprofessionalisering av lärarna eftersom lärarnas bedömningar ersattes av standardprov (Carlgren & Marton, 2001, s. 103 ff; Lundahl, 2011). Kilpatrick och Johansson (1993) hävdar dock tvärtom att det relativa betygssystemet gav stort utrymme för lärarnas professionalitet, eftersom det gav ett kraftfullt verktyg för lärarna att lösa likvärdighetsproblemet samtidigt som det gav frihet för lärarna att använda sin professionella kompetens för att bedöma elevernas kunskaper på grundval av all tillgänglig information.

Varför fick då det relativa betygssystemet utstå så mycket kritik? En förklaring till detta är att det var ett unikt system utan internationell motsvarighet. I beskrivningar och kritik av systemet (t.ex. Andersson, 1991; Carlgren & Marton, 2001; Lundahl, 2011) har detta lett till tveksamma jämförelser med länder där externa prov har direkt påverkan på enskilda elevers betyg, och där proven reducerar snarare än stärker lärarnas professionalitet. En annan förklaring till kritiken är att det var svårt att förstå vilket syfte standardproven hade, och hur dessa var avsedda att användas. Förmodligen var det dock framförallt antagandet att kunskaper och färdigheter skulle vara normalfördelade i populationen som väckte motvilja mot det relativa betygssystemet. Detta antagande kom att förknippas med föreställningar om intelligens och determinism, när det i själva verket var ett antagande baserat på matematiska och statistiska överväganden.

## Det mål- och kunskapsrelaterade betygssystemet (Lpo94/Lpf94)

År 1994 infördes mål- och kunskapsrelaterade ämnesbetyg i grundskolans åk 8 och 9 (Lpo94) och kursbetyg i gymnasieskolan (Lpf94). För grundskolan fastställdes en tregradig betygsskala med betygsstegen Godkänd, Väl godkänd

och Mycket väl godkänd. I gymnasieskolan användes också betygssteget Icke godkänd. Även i grundskolan fanns ett osynligt betygsssteg (Ej uppnått målen) för icke godkända resultat. För gymnasieskolans del angav läroplanen att studierna skulle vara kursutformade, med ett betyg för varje kurs.

Ämnesbetygen i grundskolan och kursbetygen i gymnasiet var avsedda att ge information om den enskilda elevens studieresultat, och gymnasiebetygen skulle användas för urval till högre utbildning. Men betygen skulle också ge information om skolors och kommuners resultatutveckling liksom om resultatutvecklingen på nationell nivå. Vidare skulle den publicerade informationen om skolors betygsmedeltal ge föräldrar och elever ett adekvat underlag för att bedöma olika skolors kvalitet inför elevernas val av skola. Det målrelaterade betygssystemet hade sålunda ett betydligt bredare uppdrag än det relativa betygssystemet och skulle både certifiera enskilda elever och fungera som ett lokalt och nationellt utvärderingssystem.

Det nya betygssystemet var förankrat i mål och kriterier, vilket skulle ge förutsättningar för en likvärdig betygssättning, och ett underlag för uppföljning av utvecklingen av kunskaper och färdigheter över tid på såväl lokal som nationell nivå. Skolverket fick ett nytt uppdrag om nationella prov, där provens funktion ändrades till att vara stödjande och inte styrande som de tidigare standardproven och centrala proven. Enligt uppdraget skulle syftena med det nationella provsystemet vara att kontrollera om eleverna hade nått den kunskapsnivå som kursplanerna föreskrev, att ge underlag för betygssättningen samt att ge lärarna stöd i att diagnostisera elevernas svårigheter. För gymnasieskolan skulle proven fortsätta att ge en så god grund som möjligt för en rättvis betygssättning, eftersom betygen var viktiga som urvalsinstrument till högre utbildning. Även andra uppgifter tillkom, som att kommunicera läroplanen och bedömningsprinciperna till lärarna. I flera ämnen ändrade också de nationella proven karaktär och fick större inslag av komplexa provuppgifter som ställde krav på omfattande skriftlig och muntlig elevproduktion, medan andelen uppgifter med fasta svarsalternativ minskade.

Kort efter att det nya betygssystemet införts uppdagades allvarliga problem med betygens likvärdighet, bland annat i form av stora skillnader mellan lärare, skolor och huvudmän (Skolverket, 2000, 2002). Det fanns också tecken på betygsinflation i form av ökande betygsmedelvärden från ett år till ett annat, vilket i synnerhet gällde gymnasieskolan (Cliffordson, 2004). Flera tänkbara förklaringar till likvärdighetsproblemen anfördes, som att lärare och skolledare inte varit tillräckligt väl förberedda för att kunna hantera det nya betygssystemet, att det saknades kriterier för betygssteget MVG och att kriterierna gav utrymme för olika tolkningar. Frågor restes också kring de nationella provens roll för att stödja en likvärdig betygssättning.

Regeringen tillsatte 2006 en utredning som skulle se över grundskolans mål- och uppföljningssystem. Denna utredning gjorde bedömningen att syftena med de nationella proven borde begränsas till att dels stödja en likvärdig bedömning

av elevernas kunskapsutveckling och en rättvis betygssättning, dels ge underlag för en analys av i vilken utsträckning kunskapskraven nås på skolnivå, på huvudmannanivå och på nationell nivå (SOU 2007:28).

## Den nu gällande läroplanen (Lgr 11/Lgy 11)

SOU 2007:28 kom att utgöra ett av underlagen för de nya läroplaner som gäller från 2011. De nuvarande kurs- och ämnesplanerna (de förra för grundskolan, de senare för den gymnasiala utbildningen)<sup>1</sup> inleds med en kort karakteristik och situering av ämnet och dess betydelse utanför skolan, i samhället/världen, och har därefter tre avdelningar med rubrikerna Ämnets syfte, Centralt innehåll och Kunskapskrav. Syftesdelen avslutas med ett antal punkter som definierar vilka förmågor eleverna genom undervisningen ska ges förutsättningar att utveckla. Dessa motsvarar mål men har inte detta som officiell beteckning. Avsnittet Centralt innehåll introducerades i 2011 års kurs- och ämnesplaner, detta efter en lång diskussion om den lokala nivåns frihet och förmåga att besluta om innehåll och metod för undervisningen, i synnerhet sett ur ett bedömnings- och likvärdighetsperspektiv (SOU 2007:28; Tholin, 2005). Innehållet karakteriseras i Skolverkets information som en sammanfattning av vad eleverna har rätt att lära sig och knyter därmed an till de tidigare definierade förmågorna. Däremot sägs också att lärare och elever tillsammans kan besluta dels om hur det centrala innehållet ska behandlas, dels om att lägga till fler aspekter.

Kunskapskraven slutligen har, enligt Skolverket, ambitionen att anknyta såväl till Ämnets syfte som till Centralt innehåll och är formulerade på ett sätt som undviker detaljer och som uttrycker kvalitetsskillnader med hjälp av generiska s.k. värdeord (fetstilta i texten), som lärare ska tolka för att komma fram till en adekvat och likvärdig bedömning. Kraven är formulerade för tre nivåer, E, C, A, i den sexstegade betygsskala (F-A), som introducerades samtidigt som de nya styrdokumentet 2011. Mellanliggande, godkända steg (D och B) avgörs av lärarna enligt principen att kunskapskravet för den underliggande nivån skall vara helt uppfyllt, liksom till övervägande del det för den högre nivån.

Exempel på betygsskiljande formuleringar i kursplanen i engelska för årskurs 9, vad gäller förståelse av muntligt språk ("i måttligt tempo samt i lättillgängliga texter i olika genrer") är att eleven för betygssteget E kan förstå **"det huvudsakliga innehållet och uppfatta tydliga detaljer"** och för C **"det huvudsakliga innehållet och uppfatta väsentliga detaljer"**; motsvarande krav för betygssteget A är att eleven ska **"förstå såväl helhet som detaljer"**. Ett exempel från ämnet biologi i årskurs 9 är att eleven för betygssteget E **"har grundläggande kunskaper om evolutionsteorin och andra biologiska**

<sup>1</sup> För styrdokument, kommentarmaterial och statistik, se Skolverkets hemsida ([www.skolverket.se](http://www.skolverket.se))

sammenheng og viser det gjennom at **ge eksempel og beskrive** dessa med **viss** anvendning av biologiens begrepp, modeller og teorier”. Motsvarande formulering for betygssteget C er at eleven har ”**goda** kunnskaper om evolutionsteorien og andre biologiske sammenheng og viser det gjennom at **förklara** og **visa på samband inom** dessa med **relativt god** anvendning av biologiens begrepp, modeller og teorier”. Nivå A kräver at eleven har ”**mycket goda** kunnskaper om evolutionsteorien og andre biologiske sammenheng og viser det gjennom at **förklara** og **visa på samband inom** dessa **och något generellt drag** med **god** anvendning av biologiens begrepp, modeller og teorier.”

Styrdokumentet beskriver sålunda i en innehållslig del *vad* elevena ska lära sig og lärarna ska undervisa om, og en nivåbeskrivande del, som definierar *hur väl* detta innehåll ska behärskas för olika betygsnivåer. Detta har sin motsvarighet i de engelska begreppen *content* respektive *performance standards*. Det finns alltid – og ska också finnas – ett visst mått av överlapp mellan kategorierna, inte minst vad gäller de aktuella kunskapsdomänerna. Inte sällan används också i kommenterande material og prov autentiska exempel på elevprestationer (s.k. *benchmarks*) för att illustrera både innehålls- og nivåmåssiga aspekter.

Beskrivningarna av kunnskapskraven i kursplanerna framstår dock som problematiska. Enligt SOU 2007:28 borde kunnskapskraven vara så tydligt uttryckta att lärarna inte ska behöva tolka og omformulera dem på lokal nivå, men det är uppenbart att det finns ett stort tolkningsutrymme i de kunnskapskrav som finns formulerade i kursplanerna. Detta har också visats i en omfattande studie genomförd på regeringens uppdrag av Skolverket i samarbete med externa forskare (Skolverket, 2016). I denna konstaterades att det ökade antalet betygssteg uppskattades av de flesta – trots vissa problem med stegen D og B – liksom kunnskapskravens formuleringar som stöd vid kommunikation med elevena om ämnens innehåll. Det framkom dock med stor tydlighet att andelen lärare som kände sig säkra i sina förklaringar av vad som krävs för olika betyg var väsentligt lägre år 2015 än 2009. Mindre än hälften av lärarna upplevde vidare kunnskapskraven som tydliga, og den huvudsakliga svårigheten sades vara att konkretisera og tolka de värdeord/progressionsuttryck som skiljer betygsstegen åt. Det kanske största problemet var dock den regel som funnits sedan 2011, att samtliga aspekter av kunnskapskraven ska vara uppfyllda för att en elev ska kunna få ett visst betyg; någon kompensasjon tillåts alltså inte. Detta problem har påtalats i flera tidigare sammenheng (Gustafsson, Cliffordson & Erickson, 2014; Vlachos, 2013) og bekräftades alltså i studien. Sammanfattningsvis konstaterades i rapporten att majoriteten lärare ser stora problem med den nationella likvärdigheten vad gäller betygssättning. Skolverket föreslog bland annat översyn såväl av kunnskapskrav som av regler för betygssättning.

## De nationella provens syften

Som redan nämnts hade de nationella proven initialt flera olika syften, men efter hand utkristalliserades två syften: dels att stödja en likvärdig bedömning av elevernas kunskapsutveckling och en rättvis betygssättning, dels att ge underlag för en analys av i vilken utsträckning kunskapskraven nås på skolnivå, på huvudmannanivå och på nationell nivå. I den begreppsapparat som Tveit och Olsen (2018) beskriver motsvaras dessa syften av ”sertifiering” och ”styrning”. Det är dock viktigt att understryka att de svenska nationella proven är lärarnas verktyg för att stödja en likvärdig betygssättning, och att läraren skall utnyttja all tillgänglig information som är relevant för beslutet om vilket betyg eleven skall ha.

Det är av stort intresse att veta i vilken utsträckning proven har lyckats uppnå dessa syften, vilket vi diskuterar nedan.

### **Hur väl stödjer de nationella proven likvärdig betygssättning?**

Riksrevisionen (2004) redovisade resultat från flera studier av likvärdighet i betygssättning, och bland annat visar en jämförelse mellan nationella provbetyg (dvs. det samlade betyget på det nationella provet) och slutbetyg på stor variation mellan skolor och kommuner. Slutsatserna i rapporten var kritiska till såväl regeringens som Skolverkets insatser för likvärdig betygssättning, både rörande stödmaterial och utnyttjande av de nationella proven. I motsvarande granskning av Riksrevisionen (2011) drogs slutsatsen att betygssättningen fortfarande inte sker på ett likvärdigt sätt.

Skolverket (2007) analyserade sambandet mellan nationella prov och slutbetyg i grundskolan, liksom även i gymnasieskolan (Skolverket, 2009b). Båda rapporterna visar på stor variation, både inom och mellan skolor, i hur provresultaten/provbetyget påverkar slutbetyget/kursbetyget. Inga påtagliga skillnader noterades mellan kommunala och fristående skolor. Stora skillnader fanns dock mellan ämnen. I engelska var överensstämmelsen provbetyget–slutbetyget/kursbetyget störst och i matematik minst. Det bör dock påpekas att några generella regler för graden av stöd för lärares betygssättning inte finns.

Frågan om betygets och provens stabilitet från ett år till ett annat är också av stort intresse, både när det gäller likvärdig betygssättning och utvärdering. Av olika redovisningar från Skolverket<sup>2</sup> framgår att i synnerhet provbetygets medelvärde uppvisar stora variationer i vissa ämnen<sup>3</sup>. Beräkningar för åren 2013-2017 visar att skillnaderna i engelska, svenska och geografi är relativt små, medan de är av betydande omfattning för de övriga ämnena. Givet att

<sup>2</sup> Skolverkets årliga sammanställningar av nationella provresultat återfinns på Skolverkets hemsida, under Publikationer och nyhetsbrev. Skolverkets statistik är tillgänglig under rubriken Statistik.

<sup>3</sup> Totalinsamlingar av resultat görs för ämnena engelska, matematik, svenska och svenska som andraspråk. Resultat för de samhällsorienterande ämnena i grundskolan föreligger för 1/4 av kohorten för vardera ämne, och motsvarande för de tre naturorienterande ämnena för 1/3 av kohorten.



prestationsskillnaderna på nationell nivå i ett ämne från ett år till ett annat tenderar att vara av blygsam omfattning, framstår dessa resultat som förvånande.

Det finns också en tendens att medelvärdet för slutbetyget följer medelvärdet för provbetyget, men med kraftigt dämpad variation. Detta pekar på att lärarnas betygssättning influeras av provresultaten, men att lärarna också baserar sin betygssättning på annan information, vilket är i enlighet med intentionen. Den variation i provbetygets medeltal från ett år till ett annat som är så tydlig för de flesta ämnen kan ha flera förklaringar. En skulle kunna vara att provets svårighetsgrad varierar mellan åren, och att det därför är olika svårt att nå ett visst betyg. En annan möjlig förklaring är att poänggränserna för de olika betygsstegen fastställts med bristande precision. Den poängskala som bokstavsbetygen översätts till är inte heller ekvidistant, med en skillnad på 10 poäng mellan det icke godkända betyget F och det lägsta godkända betyget E, och med en ökning om 2,5 poäng för vardera av betygsstegen D-A. De stora skillnaderna i betygsmedeltal mellan åren för vissa ämnen kan därför vara uttryck för att det finns stora skillnader i andelen F-betyg mellan åren.

Stora skillnader har också konstaterats för olika ämnen vad gäller påverkan av provresultaten på slut- och kursbetyg. Några direktiv vad gäller detta har inte funnits, men det är likväl intressant att notera konsekventa mönster över tid, framförallt vad gäller andelen elever som får högre respektive lägre slutbetyg än provbetyg (Erickson, 2017). För ämnesproven för åk 9 i grundskolan vårterminen 2017 varierade exempelvis procentandelen elever som fick högre slutbetyg än provbetyg mellan 13 % (engelska) och 39 % (matematik). Andelen elever som däremot fick ett lägre slutbetyg än provbetyg varierade från 13 % i engelska till 1 % i matematik. För engelska var det följaktligen lika stora andelar elever som fick lägre respektive högre slutbetyg än provbetyg, medan det för övriga ämnen var en betydligt mindre andel vars slutbetyg var lägre än provbetyget än som fick ett högre slutbetyg än provbetyg.<sup>4</sup>

Från det att jämförelser av detta slag började publiceras av Skolverket är några mönster stabila. Detta gäller främst lärares tydliga uppgradering av slutbetygen i relation till provbetygen i alla ämnen utom engelska. Detta är intressant av flera skäl, inte minst eftersom lärare i samtliga ämnen i enkäter förklarar sig nöjda med det stöd för betygssättning de anser sig få av resultaten på de nationella proven, vilka också anses stämma väl överens med de egna iakttagelserna (Skolverket, 2014). Det är dock väsentligt att upprepa att några fasta regler kring den tyngd som ska ges provresultaten inte existerat. Icke desto mindre är det intressant att konstatera att landets lärare i flera ämnen anser att 30-40 % av eleverna är värda ett högre slutbetyg i åk 9 än vad deras provbetyg på det nationella provet anger.

Vid införandet av det målrelaterade betygssystemet ifrågasattes förslaget om en gräns för godkända betyg i grundskolan (se t.ex. Emanuelsson, 2006). Till

---

<sup>4</sup> Se [www.skolverket.se](http://www.skolverket.se) - Statistik

försvar för förslaget underströks dock att gränsen skulle vara så lågt satt att praktiskt taget samtliga elever skulle få godkända betyg. Resultaten från den första omgången målrelaterade betyg visade dock på en dramatiskt annorlunda bild (Skolverket, 1999). Tjugo procent av de elever som gick ut grundskolan våren 1998 nådde inte målen i ett eller flera obligatoriska ämnen. Endast 91 procent av eleverna uppfyllde behörighetskravet för tillträde till gymnasieskolan (godkända betyg i svenska, engelska och matematik). Resultaten på ämnesproven i svenska, matematik och engelska visade vidare att 93, 88 respektive 96 procent av eleverna uppnått målen, vilket pekar på en betydande variation mellan de tre ämnena.

Den första omgångens svaga resultat möttes med förvåning och bestörtning, och en lång rad åtgärder har vidtagits för att förbättra resultaten. Några framgångar har dock inte nåtts i ansträngningarna att minska andelen elever som drabbas av icke godkända betyg. År 2017 var exempelvis andelen elever som inte nådde målen i ett eller flera obligatoriska ämnen 26 procent, vilket är en betydande ökning jämfört med år 1998 (Skolverket, 2017a). Oss veterligen har dock inte gränserna för godkänt resultat grundats i noggranna analyser av vilken kompetens i olika ämnen som är nödvändig för att klara gymnasieskolan, och inte heller har vi sett några argument för varför andelen elever med godkända resultat bör variera så mycket mellan olika ämnen som är fallet.

Denna korta genomgång pekar på att det finns brister i de nationella provens förmåga att ge ett adekvat stöd för likvärdig betygssättning. Det finns flera olika förklaringar till detta, och vi diskuterar några nedan.

### Provuppgifternas art

Sedan 1990-talets introduktion av ett nytt provsystem har proven haft en stor andel uppgifter av s.k. *performance*-karaktär, även kallade autentiska uppgifter. I dessa får eleverna ett begränsat antal uppgifter som i allmänhet kräver en relativt omfattande produktion av text eller tal, antingen ensam eller tillsammans med andra elever. Typen av uppgifter ger möjlighet att fånga kvaliteter i elevernas resultat som inte på samma sätt är åtkomliga med standardiserade provtyper, där eleven svarar med några få ord, eller väljer bland olika svarsalternativ. Ett uppenbart exempel på detta är uppsatsprov i språk, där det inte torde vara meningsfullt att använda någon annan typ av prov än att låta eleverna producera texter. Det är dock inte lika självklart att det skulle vara en bra metod att låta eleverna skriva längre texter om till exempel demokratibegreppet i ett prov i samhällskunskap eller fotosyntes i biologi eller kemi, eftersom resultatet skulle kunna bli drabbat av s.k. *construct irrelevant variance*, det vill säga vara mer påverkat av elevens skrivförmåga än av insikterna i de efterfrågade begreppen (Messick, 1989; Skolverket, 2017b).

Även om autentiska uppgifter har många attraktiva egenskaper tenderar de att påverka provets reliabilitet negativt. Ett skäl till detta är att tidsåtgången för eleverna att besvara dem minskar möjligheterna att använda tillräckligt många

uppgifter totalt för att säkerställa en adekvat reliabilitet. Ett annat skäl är att autentiska prov kräver komplexa bedömningar vilka är såväl tidskrävande som felbemängda (se vidare nedan). Det är för övrigt intressant att notera att Wigforss (1939) diskuterar val mellan olika provtyper på sätt som fortfarande är relevanta.

Den stora andelen autentiska uppgifter som präglar de svenska nationella proven kritiserar av OECD efter granskning av utvärdering och bedömning på nationell nivå i det svenska skolsystemet (Nusche m.fl., 2011). Ett sätt att öka reliabiliteten i det nationella provsystemet som rekommenderas är att kombinera "the use of performance-based tasks and some standardised close-ended formats" (Nusche et al., 2011:11). Med det senare avses till exempel flervalssuppgifter av olika slag, bl.a. *multiple choice* och matchning, som varken kräver produktiva svar från eleverna eller kvalitativa tolkningar av bedömarna. Denna typ av uppgifter ifrågasätts dock inte sällan och framställs stundtals som förknippad med ytliga och atomistiska kunskaper, det vill säga uppgifter långt från den handlingsorienterade syn som präglar styrdokumentet (se t.ex. Jönsson m.fl., 2017). Mot detta talar dock dels behovet av att fånga kvalitativa nivåer av förståelse, som elever av olika skäl kan ha svårt att själva verbalisera, dels möjligheten att genom slutna provformat kunna motverka att bedömningen påverkas på ett otillbörligt sätt, t.ex. av skriftlig förmåga, när detta inte är i fokus. Vidare uppvisar väl konstruerade och utprovade uppgifter av flervalstyp, som används i storskaliga bedömnings sammanhang, betydligt större stringens i bedömningen än uppgifter med öppna svar och bidrar därmed till högre reliabilitet (Downing, 2006; Haladyna, 2004). En kombination av uppgiftstyper, så som föreslås i OECD:s rapport, förefaller därför högst rimlig.

### Lärobedömningar

Lärare i Sverige har sedan länge ett mycket stort – i det närmaste odelat – ansvar för bedömning och betygssättning av de egna elevernas kunskaper. Som stöd har lärarna nationella styrdokument, till vilka det finns ett betydande antal stödmaterial av skiftande slag, samt ett omfattande nationellt prov- och bedömningssystem som ska komplettera den egna, kontinuerliga bedömningen. Proven bedöms av elevernas egna lärare, med rekommendationer om samverkan mellan kolleger, men utan krav på medbedömning. Bland annat till följd av den påvisade variabiliteten i användningen av de nationella provresultaten, föreslogs i utredningen SOU 2007:28 att systematiska studier skulle genomföras av lärares bedömning av de nationella proven, också i relation till betygssättning. Regeringen beslutade att följa förslaget och Statens Skolinspektion, inrättad 2008, gavs i uppdrag att under en treårig försöksverksamhet ombedöma upp till en tiondel av de nationella proven.

Under perioden 2009 till 2017 har åtta omgångar av ombedömningar genomförts i vilka bedömare rekryterade av Skolinspektionen gjort bedömningar som jämförts med de ursprungliga lärobedömningarna (Skolinspektionen,

2017). Bedömningarna har fokuserat på de delar av de nationella proven som innebär bedömning av omfattande elevproduktion av text, där delproven som omfattar uppsatsskrivning i svenska och engelska är de främsta exemplen. Resultaten i de olika bedömningsomgångarna är tämligen konsistenta, och visar att andelen lärare som satt ett högre delprovsbetyg än Skolinspektionens ombedömare är större än andelen lärare som satt ett lägre delprovsbetyg.

Det finns två typer av avvikelser mellan ursprungs rättning och omrättning, vilka inte separerats i Skolinspektionens rapportering. Den ena avser systematiska skillnader i grad av stränghet mellan de två bedömningarna, och den andra avser graden av överensstämmelse på elevnivå mellan de två provbetygen. På grundval av data från 2010 och 2011 års omrättningar beräknade Gustafsson, Cliffordson och Erickson (2014) medelvärdeskillnaden mellan de två bedömningarna. Resultaten visade framförallt på skillnader mellan ämnen, med små skillnader i engelska och stora skillnader i svenska. För matematik varierade skillnaderna över olika provtyper.

På grundval av resultaten i de tidiga bedömningsomgångarna argumenterade Skolinspektionen för att stora skrivuppgifter borde tas bort eller minskas i omfattning i de nationella proven. Några sådana förslag ges inte längre, utan Skolinspektionen (2017) föreslår snarare att bedömningsprocesserna på skolorna bör utformas så att skillnaderna mellan bedömare minimeras genom användning av sambedömning och anonymiserade prov. Denna förändring tycks ha sin grund i att Skolinspektionen inte längre ser ombedömningen som den korrekta bedömningen. Att det inte är möjligt att betrakta Skolinspektionens bedömningar som objektiva har dels sin grund i att deras bedömare, som skall uppfylla kraven både att ha relevant erfarenhet av bedömning och att vara tillgängliga för den omfattande bedömningsuppgiften, inte utgör ett representativt urval ur en väldefinierad population, dels i att det är rimligt att anta att omrättningsuppdraget i sig självt innebar användning av generellt strängare bedömningsnormer. Graden av överensstämmelse mellan de två bedömarna låg också på ungefär samma låga nivå som påvisats i andra undersökningar av komplexa bedömningar, exempelvis av examensarbeten inom högskolan (Härnqvist, 1999).

Gustafsson och Erickson (2013) konstaterade att Skolinspektionens ombedömningar har samma syfte som många tidigare vetenskapliga undersökningar av bedömningar av prov, men att den har genomförts med "skolinspektionslogik" snarare än med "forskningslogik". I forskning är deltagarna anonyma, vilket inte är fallet med inspektion, som delvis bygger på namngivning av enskilda skolor, stundtals refererat till som "naming, shaming and blaming" (Elstad, 2009). I inspektion utgår man från att undersökningsmetoderna ger korrekta resultat och de problematiseras inte. I forskningen däremot ägnar man stor uppmärksamhet åt metodernas reliabilitet och validitet. Vidare gör "skolinspektionslogiken" det naturligt att resultaten av inspektionen publiceras direkt i massmedierna, medan forskningsresultat når

medierna först efter att de granskats och diskuterats i vetenskapssamhället. Genom att medierna fokuserat på brister i lärarnas bedömningar verkar det som om problemet orsakas av att lärare är subjektiva och inte förmår att göra korrekta bedömningar. Detta har i sin tur lett till att tilltron till lärarkårens professionalitet skadats.

### **Slutsatser kring de nationella provens stöd för likvärdig och rättvis betygssättning**

I det relativa betygssystemet var standardprovets huvuduppgift att styra betygssättningen så att den var likvärdig och rättvis. Systemet var också utformat för att lösa denna uppgift, och gjorde det på ett framgångsrikt sätt. Vid introduktionen av det mål- och kunskapsrelaterade betygssystemet var stöd för likvärdig betygssättning en underordnad uppgift för provsystemet, eftersom det antogs att likvärdigheten skulle garanteras av styrdokumentens formuleringar kring mål och kriterier. Det nationella provsystemet var inte heller utformat för att vare sig styra eller stödja betygssättningen, eftersom det inte fanns något regelverk kring hur resultaten på de nationella proven skulle kombineras med annan information som läraren samlat kring elevernas resultat.

De studier som Skolverket (2007, 2011) har genomfört av de nationella provens betydelse för likvärdigheten i betygssättningen har framförallt fokuserat på diskrepanserna mellan provbetyg och slutbetyg. En implicit utgångspunkt för dessa jämförelser är att det är proven som är rättesnöret, och att avvikelser mellan slutbetyg och provbetyg är uttryck för brister i betygssättningen. Något stöd för en sådan tankemodell finns dock inte i styrdokumentet, och inte heller uppvisar de nationella proven egenskaper som ger stöd för en sådan tolkning. De påfallande skillnaderna i medelvärden från ett år till ett annat för merparten av de nationella proven visar att det är problematiskt att använda dem som rättesnöre för att bedöma om betygssättningen är likvärdig eller ej. De stora och systematiska skillnaderna mellan lärarnas bedömningar av nivån på elevernas prestationer med slutbetygen å ena sidan och resultaten på de nationella proven å den andra sidan visar vidare att de nationella proven oftast skattar elevernas resultat lägre än vad lärarna gör.

Även om det finns tillfällen då de nationella proven är värdefulla för att undersöka likvärdighet i betygssättning framstår det stöd som ges av proven för att stärka likvärdigheten i betygssättning som begränsat och problematiskt.

### **Slutsatser kring de nationella provens stöd för uppföljning och utvärdering**

Det andra huvudsyftet med det nationella provsystemet har hittills varit att ”ge underlag för en analys av i vilken utsträckning kunskapskraven nås på skolnivå, på huvudmannanivå och på nationell nivå”. Detta kan också betecknas som ett syfte att ge underlag för uppföljning och utvärdering på alla nivåer i utbildningssystemet.

En förutsättning för att de nationella proven skall vara användbara för uppföljning och utvärdering är att de ger resultat som är jämförbara över tid. Denna jämförbarhet är också en förutsättning för att proven skall kunna användas för att kontrollera om det föreligger betygsinflation eller ej. Trots att det redan under början av 2000-talet fanns indikationer på att betygen var utsatta för inflation, och då i synnerhet i gymnasieskolan, betraktade Skolverket länge frågan om betygsinflation som mindre intressant, och hävdade fram till 2012 att det inte existerar någon betygsinflation i Sverige (Skolverket, 2012).

Vi har redan konstaterat att de nuvarande nationella proven, med något enstaka undantag, inte uppfyller kravet på jämförbarhet över tid. Skolverket har vid några tillfällen initierat projekt som haft som syfte att hitta lösningar för att åstadkomma detta. Några större framsteg har dock inte gjorts, vilket möjligen beror på att prov med stora inslag av autentiska uppgifter av det slag som Skolverket prioriterat inte så lätt låter sig ekvivaleras över tid.

Ytterligare ett problem vad gäller det uppföljande/utvärderande syftet är att de nationella proven innehåller ett så begränsat antal uppgifter att täckningen av de olika ämnesdomänen blir svag. Detta problem diskuteras mer utförligt av Gustafsson m.fl. (2014).

Sammanfattningsvis måste vi konstatera att det nationella provsystemet inte bidragit med adekvata underlag för uppföljning och utvärdering på lokal och nationell nivå därför att proven inte uppfyller de grundläggande kraven för detta syfte. I sammanhanget är det intressant att konstatera att Tveit (2018) noterar att det norska nationella provsystemet först efter lång tid lyckades åstadkomma prov som är jämförbara över tid: "Thus, this important aspect of the tests' governing role was achieved 10 years after the first implementation of the national testing programme." (Tveit, 2018, s. 9). Det återstår ännu att se hur lång tid detta tar att åstadkomma i Sverige.

## Aktuellt utrednings- och utvecklingsarbete

Slutsatsen är sålunda att det nationella provsystemet i sin nuvarande utformning inte fullt ut kan sägas uppfylla vare sig det betygsstödjande syftet eller det uppföljande/utvärderande syftet. Dock pågår utrednings- och utvecklingsarbete som åtminstone delvis försöker avhjälpa de kvalitetsproblem provsystemet lider av. Nedan diskuterar vi kort en del av dessa aktiviteter.

### Statlig utredning om nationella prov (SOU 2016:25)

I syfte att åstadkomma ett mera effektivt och tillförlitligt, sammanhållet nationellt system beslutade den svenska regeringen under våren 2015 att tillsätta en särskild utredare för att göra en översyn av de nationella proven för grund- och gymnasieskolan. Ett ytterligare uttalat syfte var att minska lärares administrativa börda vad gäller bedömning, utvärdering och uppföljning av

elevers kunskaper. I kommittédirektivet (Regeringskansliet, 2015, sid 1) fastslogs att utredarens övergripande uppdrag bland annat var att

- se över de nationella provens syften, funktion och omfattning,
- föreslå ett system för löpande nationell utvärdering för trendmätning över tid,
- föreslå hur bedömningen av elevlösningar av nationella prov bör utformas för att säkerställa att den görs på ett likvärdigt sätt,
- utarbeta ett förslag som ökar inslagen av extern bedömning av de nationella proven på ett kostnadseffektivt sätt, samt att
- analysera förutsättningarna för att digitalisera de nationella proven och föreslå hur, i vilken omfattning och i vilken takt digitalisering kan ske.

Utredningsarbetet bedrevs under ledning av den särskilde utredaren Tommy Lagergren, som till sin hjälp hade en brett sammansatt grupp av experter inom olika områden. Utredningens slutbetänkande, med titeln *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning* (SOU 2016:25) presenterades i mars 2016 i två volymer med sammanlagt närmare 800 sidor text. Den första delen innehåller sexton kapitel i vilka direktivets olika punkter diskuteras och där förslag till förändringar läggs fram. Den andra delen omfattar fyra bilagor, som till stor del behandlar frågan om provresultat i relation till lärares betygssättning i ämnet.

Utredningens mest övergripande förslag var ett nytt, tredelat nationellt system för kunskapsbedömning, varje del med ett renodlat syfte. Nationella prov skulle fokusera på stödet för lärares betygssättning och stärka både likvärdighet och individuell rättssäkerhet. För att bidra till lokala och nationella uppföljningar/utvärderingar av kunskapsresultaten och även ge underlag för analyser av utveckling över tid, föreslogs en ny nationell kunskapsutvärdering (NKU) enligt en rullande, stickprovsbaserad, modell där olika grupper av ämnen behandlades olika år. Den tredje delen av systemet skulle utgöras av bedömningsstöd i olika ämnen. Dessa material skulle vara antingen obligatoriska eller frivilliga och av betygsstödjande, diagnostisk eller formativ typ, men syftena måste klargöras på ett bättre sätt än i det rådande systemet. Utredningens huvudförslag var sålunda att det nya nationella provsystemets tre delar skall stödja ett av tre huvudsyften: (1) likvärdig och rättvis betygssättning; (2) användbara underlag för lokal och nationell uppföljning och utvärdering av kunskapsresultat; och (3) stöd för individuella anpassningar av undervisning.

För att nå syfte 1 behöver både validitet och reliabilitet i enskilda prov förbättras, och provresultaten göras jämförbara från år till år. En rekommendation handlade här om utveckling av ett systemramverk med gemensamma kvalitetskriterier för de betygsstödjande proven. Ett ytterligare förslag gällde tillsättande av en kvalitetsgrupp med svenska och utländska

experter med god vetenskaplig kompetens och beprövad erfarenhet inom provutvecklingsområdet.

Utredningen menade också att provens roll för lärares betygssättning måste explicitgöras, eftersom den stora variationen mellan år, ämnen, skolenheter och klasser inte är försvarbar ur rättvise- och likvärdighetsperspektiv. Grundprinciper för hur relationen mellan provbetyg och lärarbetyg på gruppnivå kan utformas presenteras och diskuteras i utredningen medan arbetet med konkretisering och implementering föreslås tas om hand av Skolverket.

Utredningsbetänkandet sändes ut för granskning till ett stort antal remissinstanser, och förslagen mottogs i huvudsak positivt. Dock ställde sig Skolverket negativt till förslaget om en nationell kunskapsutvärdering, med hänvisning bland annat till att detta skulle medföra behov av en kostsam kompetensuppbyggnad inom Skolverket. Man invände också mot förslaget om en kvalitetsgrupp som stöd i arbetet med de nationella proven.

### **Systemramverk för nationella prov**

En överenskommelse ingicks i februari 2015 mellan Skolverket och en grupp forskare (Gudrun Erickson, Jan-Eric Gustafsson och Peter Nyström) vid Göteborgs universitet kring utveckling av ett systemramverk. Den primära målgruppen för ramverket var Skolverket samt de lärosäten som på Skolverkets uppdrag utvecklar proven. Uppdraget skulle redovisas i två delar, en allmänt utredande del och en del som innehöll konkreta förslag till riktlinjer. Den senare skulle tjäna som stöd för Skolverket i arbetet med överenskommelser och analyser av material samt för institutionerna i utvecklingen av konkreta provramverk och provmaterial.

Uppdraget redovisades dels i form av en delrapport som beskriver och diskuterar utgångspunkterna för utvecklingen av ett ramverk, dels som ett operativt förslag till systemramverk. Det senare avlämnades till Skolverket den 30 juni 2017 (Dnr 2015:00077). Förslaget hade arbetats fram i kontinuerlig kommunikation med de i provsystemet medverkande universitetsinstitutionerna och i samråd med en extern expertgrupp med bred erfarenhet och kompetens i bedömningsfrågor.

Förslaget till systemramverk innehöll åtta avsnitt, de första med fokus på grundläggande begrepp och överväganden. Därefter följer avsnitt om provutveckling, bedömning och rapportering. Rapporten avslutas med en översiktlig del kring riktlinjer för provens användning och genomförande samt en referenslista, som bland annat hänvisar till de internationella *Standards* som haft stor betydelse för systemramverkets utformning (främst *Standards for educational and psychological testing* (AERA, AEA & NCME, 2014) och de *Standards* och *Guidelines* som publicerats av Educational Testing Service (ETS).

Systemramverket utvecklades utan kännedom om exakt vilka syften som skulle komma att gälla för det nationella provsystemet. Det var dock rimligt att



utgå från antagandet att det primära syftet skulle vara att proven ska ge stöd för likvärdig och rättvis betygssättning, och systemramverkets formuleringar kring krav på validitet och reliabilitet har detta som utgångspunkt. Det understryks dock också att även de betygsstödjande proven måste kunna ge provresultat som är jämförbara över tid, och att detta gör det nödvändigt att införa robusta statistiska metoder för ekvivalering av prov.

Institutionerna fick ta del av texten för användning i sitt arbete med utveckling av preliminära förslag till enskilda provramverk. Detta sammanföll i tid med att Skolverket antog förslaget till systemramverk (Skolverket, 2017b). Under våren 2018 har en serie *peer reviews* mellan de olika institutionerna genomförts. Därefter har institutionerna vidareutvecklat sina provramverk, bland annat på basis av kollegiala iakttagelser och kommentarer, och i juni 2018 levererat en andra version av provramverk. Planen är att dessa ska bearbetas ytterligare, granskas och slutligen fastställas i juni 2019.

Det är sålunda ännu oklart i vilken utsträckning system- och provramverken kommer att medföra att de kvalitetsbrister som för närvarande kännetecknar delar av det svenska nationella provsystemet kommer att avhjälpas, och hur lång tid detta kommer att ta.

## Slutord

Under den period om ca 40 år som det relativa betygssystemet var i bruk utvecklades en trygg förvisning om att de betyg som lärarna fattade beslut om i huvudsak var likvärdiga och rättvisa, samtidigt som det relativa betygssystemet som sådant var utsatt för en massiv kritik. En hypotes är därför att det kom som en överraskning att betygen i det mål- och kunskapsrelaterade betygssystemet inte hade samma goda egenskaper, och en lång rad åtgärder vidtogs av statsmakterna för att stödja lärarna i ansträngningarna att sätta likvärdiga och rättvisa beslut. En sådan åtgärd var att låta de nationella proven få till uppgift att stödja lärarnas betygssättning även i det nya betygssystemet, vilket dock inte fullt ut löst problemet med olikvärdig betygssättning.

Vi har pekat på att dessa problem framförallt är förknippade med de nationella provens brist på reliabilitet och stabilitet över tid, vilket är två centrala grunder när det gäller stöd för rättvis och likvärdig bedömning. Någon tydlig princip för på vilket sätt och i vilken utsträckning de nationella proven skall stödja betygssättningen har heller inte uttalats, även om en regel från och med höstterminen 2018 införts att provresultaten ska "särskilt beaktas" vid betygssättning.

Statsmakternas uppdrag till de universitetsinstitutioner som utvecklar proven har fram till 2018 inneburit krav att förena alltför många och alltför olika syften. Proven har i regel inte förmått att fullt ut leva upp till dessa krav, vilket ska ses i relation till den ofta påtalade omöjligheten och/eller olämpligheten i att förena

flera syften inom ramen för ett och samma prov (t.ex. Gilmore, 2002; Newton, 2007; SOU 2016:25; Tveit, 2018). Det finns dock anledning att understryka att de nationella proven och bedömningsstöden generellt erhållit mycket goda omdömen från lärare, som ofta påtalar provens roll som gott stöd både i kommunikation kring ämnet/styrdokumentet och kartläggning av olika kvaliteter i elevers kunskande.

De förändringar som beslutats på senaste tid innebär också en viss åtstramning av syftena, där det individuella betygsstödet anges som huvudsyftet, men också att de nationella proven som ett andra syfte kan bidra till att stärka skolornas kvalitetsarbete genom analyser av provresultaten i relation till uppnådda kunskapskrav på skolnivå, huvudmannanivå och på nationell nivå. Det finns dock mycket som pekar på att denna andra uppgift svårligen låter sig förenas med uppgiften att stödja likvärdig betygssättning, varför det vore lämpligt att utveckla ett separat uppföljnings- och utvärderingssystem för lokal och nationell utvärdering.

## Om författarna

Jan-Eric Gustafsson er professor emeritus i pedagogik ved Göteborgs universitet. Hans forskningsinteresser omfatter blant annet individuelle differenser i kognitive færdigheter, og studier av faktorer som har betydelse for utdanningsresultat på såvel individ- som systemnivå.

Institutionsanknytning: Institutionen for pedagogik og specialpedagogik, Göteborgs universitet, Box 100, 405 30 Göteborg, Sverige

E-post: [jan-eric.gustafsson@ped.gu.se](mailto:jan-eric.gustafsson@ped.gu.se)

Gurrun Erickson er professor emerita i pedagogik med inriktning mot språk og bedømming ved Göteborgs universitet.

Institutionsanknytning: Institutionen for pedagogik og specialpedagogik, Göteborgs universitet, Box 100, 405 30 Göteborg, Sverige

E-post: [gurrun.erickson@ped.gu.se](mailto:gurrun.erickson@ped.gu.se)

## Referenser

AERA, AEA & NCME (2014). *Standards for educational and psychological testing*.

Washington, DC: American Educational Research Association. ISBN: 978-0-935302-35-6.

Andersson (1991). *Relativa betyg. Några empiriska studier och en teoretisk genomgång i ett historiskt perspektiv*. Doktorsavhandling, Pedagogiska institutionen, Umeå universitet.

Carlgren, I., & Marton, F. (2001). *Lärare av imorgon*. Stockholm: Lärarförbundet.

Cliffordson, C. (2004). Betygsinflation i de målrelaterade gymnasiebetygen. *Pedagogisk forskning i Sverige*, 9(1), 1-14.

- Downing, S. M. (2006). Selected-Response Item Formats in Test Development. I S. M. Downing & T. M. Haladyna (red.), *Handbook of Test Development* (s. 287-301). Mahwah: Lawrence Erlbaum Associates, Inc.
- Ds 1990:60. *Betygens effekter på undervisningen*. Stockholm: Utbildningsdepartementet.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173–189.
- Emanuelsson, I. (2006). Betyget godkänd i en obligatorisk skola för alla. I E. Forsberg & E. Wallin (red.), *Skolans kontrollregim – ett kontraproduktivt system för styrning* (s. 46-59) Stockholm: HLS.
- Erickson, G. (2017). Experiences with Standards and Criteria in Sweden. I S. Blömeke & J.-E. Gustafsson (red.), *Standard Setting in Education – The Nordic Countries in an International Perspective* (s. 123-142). Cham, Switzerland: Springer International Publishing.
- Erickson, G., Gustafsson, J-E. & Nyström, P. (2017). *Förslag till systemramverk för nationella prov*. Rapport till Skolverket (Dnr 2015:00077).
- Gilmore, A. (2002). Large-scale assessment and teachers' assessment capacity: learning opportunities for teachers in the National Education Monitoring Project in New Zealand. *Assessment in Education*, 9(3), 343-361.
- Gustafsson, J-E., Cliffordson, C. och Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter*. Stockholm: SNS förlag.
- 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.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. New York: Routledge Taylor & Francis Group.
- Härnqvist, K. (1999). *En akademisk fråga: en ESO-rapport om ranking av C-uppsatser: rapport till Expertgruppen för studier i offentlig ekonomi* (Ds 1999:65).
- Jönsson A., Rosenlund D. & Alvé F. (2017). Complement or Contamination: A Study of the Validity of Multiple-Choice Items when Assessing Reasoning Skills in Physics. *Front. Educ.* 2 (48). doi: 10.3389/educ.2017.00048
- Kilpatrick, J., & Johansson, B. (1994). Standardized mathematics testing in Sweden: The legacy of Frits Wigforss. *Nordic Studies in Mathematics Education*, 1(1), s. 6–30.
- Lpf94 (1994). Läroplan för de frivilliga skolformerna Lpf 94. Stockholm: Skolverket.
- Lpo94 (1994), Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet Lpo 94, Stockholm: Skolverket.
- Lundahl, C. (2011). Skolbedömningens praktiska och administrativa dimensioner. I C. Lundahl & M. Folke-Fichtelius (red.). *Bedömning i och av skolan – praktik, principer och politik* (s. 299–314). Lund: Studentlitteratur AB.
- Messick, S. A. (1989). Validity. I Robert L. Linn (Ed.), *Educational Measurement* (Third edition, pp. 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.
- Regeringskansliet (2015). Kommittédirektiv: *Översyn av de nationella proven för grund- och gymnasieskolan, 2015:36*.
- Riksrevisionen (2004). *Betyg med lika värde? – en granskning av statens insatser*. RiR 2004:11. Stockholm: Riksrevisionen.

- Riksrevisionen (2011). *Lika betyg, lika kunskap? – en uppföljning av statens styrning mot en likvärdig betygssättning i grundskolan*. RiR 2011:23. Stockholm: Riksrevisionen.
- Skolinspektionen (2017). *Bedömningsprocessernas betydelse för likvärdigheten. Ombedömning av nationella prov 2016. Redovisning av regeringsuppdrag Dnr U2014/7535/GV*. Stockholm: Skolinspektionen.
- Skolverket (1999). *Beskrivande data om barnomsorg och skola 1999*. Skolverkets rapport nr 173. Stockholm: Skolverket.
- Skolverket (2000). *Nationella kvalitetsgranskningar*. Stockholm: Skolverket.
- Skolverket (2002). *Betygssättning vid 18 fristående skolor*. Stockholm: Skolverket.
- 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 (2009a). *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer* (s. 40-85). Stockholm, Skolverket.
- Skolverket (2009b). *Likvärdig betygssättning i gymnasieskolan? En analys av sambandet mellan nationella prov och kursbetyg*. Stockholm: Skolverket.
- Skolverket (2011). *Redovisning av uppdrag om avvikelser mellan provresultat och betyg i grundskolans årskurs 9*. Redovisning av regeringsuppdrag (Dnr 75-2011:411). Stockholm: Skolverket
- Skolverket (2012). *Betygsinflation - betygen och den faktiska kunskapsutvecklingen*. Stockholm: Skolverket.
- Skolverket (2014). *Så tycker lärarna om de nationella proven 2013*. Rapport 404. Stockholm: Skolverket.
- Skolverket (2016). *Utvärdering av den nya betygsskalan samt kunskapskravens utformning*. Stockholm: Skolverket.
- Skolverket (2017a). *Slutbetyg i grundskolan, våren 2017*. PM Dnr: 2017:01232. Stockholm: Skolverket.
- Skolverket (2017b). *Skolverkets systemramverk för nationella prov*. Stockholm: Skolverket.
- Skolöverstyrelsen (1972). *Betygsättningen i grundskola och gymnasieskola*. Skolöverstyrelsen, Stockholm.
- SOU (1942:11). *Betänkande med utredning och förslag angående betygssättningen i folkskolan*. Stockholm: Ecklesiastikdepartementet.
- SOU (1977:9) *Betygen i skolan. Betänkande av 1973 års betygsutredning*. Stockholm: Statens Offentliga Utredningar
- SOU (2007:28). *Tydliga mål och kunskapskrav i grundskolan - Förslag till nytt mål- och uppföljningssystem*. Stockholm: Utbildningsdepartementet.
- SOU (2016:25). *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning*. Stockholm: Utbildningsdepartementet.
- Tholin, J. (2005). Den innehållslösa grundskolan. I S. Granath, J. Miliander. & E. Wennö (red.), *The Power of Words. Studies in Honour of Moira Linnarud*, (s. 217-223). Karlstad: Karlstad University Press.
- Tveit, S. (2018). Ambitious and Ambiguous: shifting purposes of national testing in the legitimation of assessment policies in Norway and Sweden (2000-2017). *Assessment in Education: Principles, Policy & Practice* 25(3).
- Tveit, S. och Olsen, R. V. (2018). Hvilke formål og roller har eksamen i norsk grunnopplæring? *Acta didactica Norge* 12 (4).
- Vlachos, J. (2013). Betygssystem som bäddar för misslyckanden. *Ekonomistas*, juni 2013.
- Wigforss, F. (1939). *Kunskapsprövningar*. Pedagogiska skrifter utgivna av Sveriges Allmänna Folkskolläraryrkesförenings Litteratursällskap, Häfte 165.