

SNS UTBILDNINGSKOMMISSION



Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter

JAN-ERIC GUSTAFSSON, CHRISTINA CLIFFORDSON, GUDRUN ERICKSON

SNS FÖRLAG

Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter

JAN-ERIC GUSTAFSSON
CHRISTINA CLIFFORDSON
GUDRUN ERICKSON

SNS FÖRLAG

SNS Förlag
Box 5629
114 86 Stockholm
Telefon: 08-507 025 00
info@sns.se
www.sns.se

SNS är en politiskt oberoende ideell förening som genom forskning, möten och bokutgivning bidrar till att beslutsfattare i politik, offentlig förvaltning och näringsliv kan fatta välgrundade beslut baserade på vetenskap och saklig analys.

SNS UTBILDNINGSKOMMISSION är ett flerårigt forskningsprogram med fokus på den övergripande frågan hur Sverige kan stärkas som kunskapsnation. Ambitionen är att arbetet ska mynna ut i tydliga och konkreta policyförslag för den svenska utbildningspolitiken. Fem forskningsrapporter varav en policyrapport kommer att produceras inom ramen för projektet.

Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter

Jan-Eric Gustafsson, Christina Cliffordson, Gudrun Erickson
Första upplagan
Första tryckningen

© 2014 Författarna och SNS Förlag
Omslag och grafisk form: Allan Seppa
Tryck: Books on Demand, Tyskland

ISBN 978-91-86949-55-6

Innehåll

Utgivarens förord	5
Sammanfattning	7

1. INTRODUKTION	11
2. KUNSKAPSBEDÖMNING PÅ INDIVIDNIVÅ	17
3. KUNSKAPSBEDÖMNING PÅ SYSTEMNIVÅ	45
4. ANTAGNINGSSYSTEMET TILL HÖGRE UTBILDNING	70
5. SAMMANFATTANDE DISKUSSION OCH FÖRSLAG	99

Ordlista	109
Referenser	113

Utgivarens förord

SNS Utbildningskommission är ett flerårigt forskningsprogram med fokus på den övergripande frågan hur Sverige kan stärkas som kunskapsnation. Detta är den tredje rapporten från forskningsprogrammet och handlar dels om det svenska utvärderingssystemet för skolan, dels om antagningssystemet för högre utbildning. Författare till rapporten är Jan-Eric Gustafsson, professor i pedagogik, Göteborgs universitet, Christina Cliffordson, professor i pedagogik, Högskolan Väst och Gudrun Erickson, docent i pedagogik, Göteborgs universitet.

SNS hoppas att rapporten kan bidra till en fördjupad diskussion om hur det svenska utvärderingssystemet för skolan bör vara utformat samt hur antagningssystemet för högre utbildning kan förbättras. För analys, slutsatser och rekommendationer svarar helt och hållet rapportens författare. SNS som organisation tar inte ställning till dessa. SNS har som uppdrag att initiera och presentera forskningsbaserade analyser av viktiga samhällsfrågor. Projektet har kunnat genomföras tack vare bidrag från den referensgrupp som följt SNS Utbildningskommission. I denna grupp ingår Ernst & Young, KTH, LO, Lärarförbundet, SKL, Sveriges Skolledarförbund, Skolinspektionen, Skolverket, Stockholms stad, Theeducation AB, Tieto och Viktor Rydbergs Stiftelse. Referensgruppen leds av ordförande Peter Gudmundsson, rektor KTH.

Projektet har även följts av en expertgrupp bestående av Anders Björklund, professor i nationalekonomi, Institutet för social forskning, SOFI, Stockholms universitet; Arvid Carlsson, professor, Nobelpristagare i medicin 2000; Anna Ekström, generaldirektör, Skolverket; Inger Enkvist, professor i språk och litteratur, Lunds universitet; Jan-Eric Gustafsson, professor i pedagogik, Göteborgs universitet; Lars Heikensten, vd, Nobelstiftelsen; Torkel Klingberg, professor i neurovetenskap, KI; Lars Leijonborg, ordförande, Friskoleutredningen; Pär Nuder, f.d. finansminister; Bengt Samuelsson, professor, Nobelpristagare i medicin 1982, och Sverker Sörlin, professor i miljöhistoria, KTH.

SNS tackar för det engagemang som referensgruppens och expertgruppens ledamöter visat under arbetets gång. Lämnade synpunkter har varit till stort stöd för författarna. Varken referensgruppen eller expertgruppen ansvarar för innehållet i rapporten.

6

SNS vill även tacka Peter Fredriksson, professor i nationalekonomi, Stockholms universitet, Sverker Härd, enhetschef, Skolverket, Kristian Ramstedt, f.d. enhetschef, Skolverket, och Jonas Vlachos, docent i nationalekonomi, Stockholms universitet, som deltagit i en särskild referensgrupp som också följt projektet och bidragit med synpunkter. Tack också till Jörgen Tholin, universitetsdirektör vid Göteborgs universitet samt docent i pedagogik, som granskat rapporten och framfört sina synpunkter på ett akademiskt seminarium.

Stockholm i mars 2014

Camilo von Greiff, forskningsledare, SNS

Sammanfattning

Att bedöma elevers kunskaper med hjälp av betyg och nationella prov har stor betydelse både för de elever som deltar i utbildningen och för att utbildningen ska kunna utvecklas och förbättras. Betygen är dessutom viktiga för elevernas möjligheter till fortsatta studier på gymnasie- och högskolenivå, för deras möjligheter att välja attraktiva skolor eller lärosäten och utbildningsprogram och för deras studiemotivation och självkänsla. Betygens stora betydelse gör det angeläget att betygssystemet uppfyller högt ställda kvalitetskrav: att de är rättvisa och att de på ett pålitligt sätt uttrycker elevens uppnådda kunskaper och färdigheter. Detsamma gäller för de nationella proven.

Även kunskapsbedömningar på systemnivå, till exempel nationellt, kan få stor betydelse för hur utbildningspolitiken utformas, varför även dessa kunskapsbedömningar måste uppfylla högt ställda kvalitetskrav. De internationella jämförande studierna, som PISA och TIMSS, är exempel på sådana kunskapsbedömningar.

Syftet med denna rapport

Denna rapport analyserar den svenska skolans kunskapsbedömningssystem på elev- och systemnivå. Vi kommer att visa på brister i dagens system men också ge förslag till hur det kan förbättras. Analysen motiveras av tre faktorer. För det första har den svenska skolan sedan slutet av 1980-talet genomgått en mängd förändringar som inte minst rört systemen för kunskapsbedömning både på elev- och systemnivå. Vi anser att det är angeläget att belysa och begrunda vad dessa förändringar inneburit och vilka konsekvenser de fått för skolan och kunskapsbedömningen. För det andra ser vi och andra bedömare (t.ex. Nusche m.fl., 2011) flera uttryck för att den svenska skolans kunskapsbedömningssystem inte uppfyller de kvalitetskrav som är rimliga att ställa. För det tredje ser vi att en del av kvalitetsproblemen

skulle kunna lösas om man använde sig av hittills outnyttjade möjligheter som finns i den internationella kunskaps- och metodutvecklingen inom den pedagogiska mätläran.

Rapporten behandlar utformningen av system för betygssättning, nationella prov, nationell utvärdering/kunskapsbedömning och antagningssystemet till högre utbildning. Vi fokuserar i första hand på grundskolan och gymnasieskolan, med särskild tyngdpunkt på grundskolan. Men vi berör också den högre utbildningen då gymnasiebetygens huvuduppgift är att utgöra underlag för urval och antagning till högre utbildning.

Rapportens slutsatser

(O)LIKVÄRDIGA BETYG OCH NATIONELLA PROV

Likvärdigheten i betygen uppvisar stora brister. De målrelaterade betygen lider av bristande likvärdighet mellan lärare och skolor, och i synnerhet gymnasiebetygen och vissa av grundskolebetygen är påverkade av inflation. Det är därför inte möjligt att använda betygen för att studera kunskapsutveckling över tid. Och att använda betygen för urval till gymnasium och högre utbildning kan vara problematiskt ur både rättvise- och effektivitetssynpunkt.

Vi identifierar även likvärdighetsproblem i de nationella proven. För prov där eleverna skriver längre texter vars kvalitet ska bedömas, är graden av överensstämmelse mellan lärarnas bedömningar och externa bedömningar inte tillfredsställande. I de flesta ämnen är lärarnas bedömningar generösare än de externa bedömningarna. Ett annat problem med de nationella proven är att det i de flesta ämnen finns en stor variation i provbetyg från ett år till ett annat. Vidare är andelen elever som får ett icke godkänt provbetyg betydligt högre i exempelvis matematik än andelen som får ett icke godkänt slutbetyg eller kursbetyg. Dessa omständigheter indikerar att de nationella proven i åtminstone vissa ämnen ger ett begränsat stöd till lärarnas betygssättning samt att de heller inte kan uppfylla sitt andra syfte: att ge underlag för en analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, på huvudmannanivå och på nationell nivå.

BEHOV AV ETT NATIONELLT KUNSKAPSBEDÖMNINGSSYSTEM

Bristerna i likvärdighet i betyg och nationella prov gör att inget av dessa instrument är helt tillförlitliga för att bedöma elevers kunskaper, vare sig

på elev- eller systemnivå. Inte heller de internationella jämförande undersökningarna, som kommit att bli den viktigaste källan till information om kunskapsutvecklingen i den svenska skolan, kan på egen hand visa kunskapsutvecklingen på nationsnivå på ett tillfredsställande sätt. Även om dessa studier ger värdefull information visar de enbart delar av vissa ämnesområden, och tyngdpunkterna i proven motsvarar inte alltid fullt ut tyngdpunkterna i den svenska läroplanen.

Med denna utgångspunkt drar vi slutsatsen att Sverige behöver utveckla ett koherent ramverk för uppföljning och utvärdering som integrerar relativt fristående komponenter och nivåer till ett sammanhängande system.

ANTAGNINGSSYSTEMET TILL HÖGRE UTBILDNING BEHÖVER REFORMERAS FRÅN GRUNDEN

Vår huvudsakliga slutsats när det gäller antagningssystemet till högre utbildning är att det allvarligt brister både i rättssäkerhet och effektivitet, och därför bör bli föremål för en total genomlysning.

Rapportens rekommendationer

KUNSKAPSBEDÖMNING PÅ ELEV- OCH SYSTEMNIVÅ

- * Inför en försöksverksamhet med ett rullande, stickprovsbaserat, nationellt kunskapsbedömningssystem för grund- och gymnasieskolan inom ett par ämnen. Låt en utvärdering av försöksverksamheten ligga till grund för att utveckla ett kunskapsbedömningssystem för andra ämnen.
- * Det nationella kunskapsbedömningssystemet liksom det nationella provsystemet ska omfatta såväl provdelar där resultaten uttrycks i poäng på skalor, som provdelar som ger information av kvalitativ art.
- * Koppla samman det nationella provsystemet med det nationella kunskapsbedömningssystemet i syfte att de nationella proven bättre ska stödja en likvärdig och över tid stabil betygssättning.
- * Samla in resultat från de nationella proven storskaligt och på uppgiftsnivå för att främja uppföljnings- och forskningsändamål.
- * Formulera om direktiven för betygssättning i styrdokumentet så att ett betyg på en kurs speglar en sammanvägd bedömning av elevens kunskaper på kursens olika områden och inte som nu enbart elevens sämsta område.

ANTAGNINGSSYSTEMET TILL HÖGRE UTBILDNING

- * Låt en offentlig utredning, helst med parlamentarisk sammansättning, göra en total översyn av antagningssystemet till högre utbildning med syfte att komma till rätta med de allvarliga rättssäkerhets- och effektivitetsproblem som föreligger.
- * En sådan utredning bör bland annat ge förslag på hur kraven på allmän och särskild behörighet till olika högre utbildningar kan kopplas till gymnasieutbildningarnas innehåll och kravnivåer för olika betyg.
- * Renodla Högskoleprovets användning som urvalsinstrument till att utgöra en ”andra chans”, enligt de ursprungliga intentionerna. Detta bör göras genom att dels minska andelen sökande som antas till högre utbildning via Högskoleprovet till 10–15 procent av det totala antalet platser, dels genom att införa en nedre åldersgräns för att få genomföra provet.

1. Introduktion

I utbildningssammanhang är de system som används för att bedöma de studerandes kunskaper och färdigheter av stor betydelse både för de elever som deltar i utbildningen och för dem som ska utveckla och förbättra utbildningen. Betyg är en bedömning av uppnådda kunskaper efter en mer omfattande studieperiod, som en kurs, en termin, ett läsår, ett stadium, eller en 9-årig grundskola. Sådana sammanfattande (eller "summativa", se ordlista) kunskapsbedömningar har stor betydelse för elevernas möjligheter till fortsatta studier på gymnasie- och högskolenivå, för deras möjligheter att välja attraktiva skolor/lärosäten och utbildningsprogram, och för deras studiemotivation och självkänsla. På engelska används uttrycket *high-stakes* för sådana system som har betydelsefulla konsekvenser för enskilda personer, och hit hör tveklöst betygen i den svenska skolan. Betygens stora betydelse gör det angeläget att betygssystemet är rättvist och förmår att uttrycka relevant information om uppnådda kunskaper och färdigheter.

Även under lärandets gång bedöms elevernas kunskaper, både av eleverna själva och av lärarna. Dessa bedömningar görs bland annat i syfte att identifiera missuppfattningar och luckor i elevens lärande för att det ska vara möjligt att ge en detaljerad återkoppling som kan influera det fortsatta lärandet. Denna typ av bedömning brukar benämnas "formativ bedömning" (se ordlista) eller "bedömning för lärande". Forskning visar att det finns en stor potential för att formativ kunskapsbedömning har positiva effekter på utvecklingen av kunskaper och färdigheter (se t.ex. Hattie, 2009).

En annan form av kunskapsbedömning är de internationella jämförande studier av olika länders kunskapsresultat, som PISA ("Programme for International Student Assessment") i OECD:s regi och TIMSS ("Trends in International Mathematics and Science Study") som genomförs av IEA ("International Association for the Evaluation of Educational Achievement"). Syftet med dessa undersökningar är dels att informera om vilka resultat som uppnås i olika länder, dels att lämna underlag för att närmare kunna analysera vilken betydelse olika faktorer har för resultaten. Med den

terminologi som vi introducerade ovan kan man säga att de internationella studierna är kunskapsbedömningar på utbildningssystemnivå, som görs med både summativa och formativa syften. Även bedömningar på systemnivå har mycket stort genomslag och kan få stor betydelse för hur utbildningspolitiken utformas. Därför måste även dessa kunskapsbedömningar uppfylla högt ställda kvalitetskrav.

Denna rapport handlar om kunskapsbedömning i och av den svenska skolan. Vi fokuserar i första hand på grundskolan och gymnasieskolan, med särskild tyngdpunkt på grundskolan. Men vi berör också den högre utbildningen eftersom gymnasiebetygens allt annat överskuggande uppgift är att utgöra underlag för urval och antagning till högre utbildning.

De frågor vi tar upp rör i första hand utformning av system för betygssättning, nationella prov, nationell utvärdering/kunskapsbedömning samt antagningssystemet till högre utbildning. I utbildningspolitiska diskussioner kan sådana frågor betraktas som eviga, så det krävs en motivering för varför vi återigen tar upp dem. Det finns tre huvudsakliga skäl till detta. Det första är att den svenska skolan sedan slutet av 1980-talet genomgått en lång serie förändringar som inte minst har rört systemen för kunskapsbedömning både på elev- och på systemnivå. Detta gör det angeläget att belysa och begrunda förändringarnas innebörd och konsekvenser. Det andra skälet är att vi och andra bedömare ser flera uttryck för att den svenska skolans kunskapsbedömningssystem inte helt uppfyller de kvalitetskrav som är rimliga att ställa. Det tredje skälet är att vi ser vissa hittills outnyttjade möjligheter att dra nytta av den internationella kunskaps- och metodutvecklingen inom den pedagogiska mätläran när det gäller att lösa en del av kvalitetsproblemen.

Förändringar av den svenska skolan under de senaste decennierna

Från att ha varit ett av världens mest centraliserade skolsystem förvandlades det svenska skolsystemet under början av 1990-talet till ett av världens mest decentraliserade och avreglerade skolsystem. Skolans organisation och styrning förändrades radikalt när ansvaret för att genomföra utbildning decentraliserades till kommuner och fristående huvudmän och en ny statlig skoladministration skapades. Vidare infördes en avreglerad resursfördelning, valfrihet mellan kommunala och fristående skolor, fri etableringsrätt på skolmarknaden med skolpengen som ekonomiskt incitament, nya läroplaner, en kursutformad gymnasieskola, ett nytt betygssystem och nya

system för uppföljning och utvärdering av skolans resultat.

Ideologiskt vägledande för dessa förändringar var en stark tilltro till ett decentraliserat beslutsfattande och till mål- och resultatstyrning. Läroplanerna formulerade målen för grund- och gymnasieskolan, medan det var upp till huvudmännen att utforma organisation och undervisning så att målen nåddes. Mot bakgrund av den starka betoningen av mål- och resultatstyrning är det måhända överraskande att mycket av förändringarna när det gäller kunskapsbedömningar på både individ- och systemnivå innebar att man tog avstånd från systematiska observationer och mätningar. Detta kan förmodligen delvis hänföras till det paradigmskifte inom pedagogisk forskning som inneburit att kvantitativt inriktad forskning inte anses ge svar på viktiga frågor, utan i stället att pedagogiska forskningsfrågor företrädesvis bör studeras med kvalitativa ansatser. Detta paradigmskifte inleddes på 1970-talet och fick ett omfattande genomslag i forskningen om skolan, liksom i utformningen av systemen för kunskapsbedömning inom skolan på 1990-talet (se Gustafsson 2013b för en mer omfattande diskussion).

Under senare delen av 2000-talet har en ökad reglering med nya läro- och kursplaner ägt rum. Man har också stärkt uppföljningen av elevernas kunskapsutveckling, bland annat genom en ny betygsskala, betyg från årskurs 6, skriftliga omdömen från årskurs 1 och nationella prov i flera ämnen. Dessutom har kontrollen och uppföljningen av skolornas verksamhet ökat i omfattning med en kraftigt utökad skolinspektion med förstärkta sanktionsmöjligheter, och en ny, mer detaljerat reglerande, skollag (SFS 2010:800). Dessa förändringar motiverades delvis av att resultaten för de svenska grundskoleeleverna sedan mitten av 1990-talet visat på en successiv försämring inom matematik, naturvetenskap och läsning i internationella kunskapsmätningar. För flera av dessa åtgärder saknas dock forskningsunderlag som ger grund för att förvänta sig positiva effekter på elevernas resultat. Det är också osäkert i vilken utsträckning det finns information som gör det möjligt att utvärdera effekterna av de vidtagna åtgärderna (SOU 2013:30).

Kvalitetsproblem i den svenska skolans kunskapsbedömningssystem

Nusche m.fl. (2011) identifierade många positiva egenskaper när de för OECD:s räkning utvärderade det svenska systemet för uppföljning och utvärdering. De konstaterade att det decentraliserade svenska skolsystemet

arbetar enligt målstyrningsprinciper som leder till att alla nivåer inom skol-systemet är engagerade i uppföljning och utvärdering. På den nationella nivån finns en hög grad av öppenhet och transparens när det gäller rapporteringen och att göra data tillgängliga. Rapportförfattarna menade också att det nationella provsystemet, som bygger på en hög grad av förtroende för lärarna i att genomföra och bedöma proven, stödjer utveckling av lärarproffessionalitet. Men de pekade också på att kvaliteten av lärarnas bedömningar av proven varierar och att betygssättningen brister i likvärdighet. För att lösa dessa problem behövs prov och andra hjälpmedel som kan stödja en pålitlig och rättvis betygssättning.

De påpekade också att det finns ett system för nationell kunskapsbedömning av den svenska skolans resultat som bygger på data från bland annat internationella studier, rapporter från Skolinspektionen och sammanställningar av resultat på nationella prov och betyg. Men de ställer sig tvivksam till att man i så stor utsträckning förlitar sig på betygsdata och att annan information om skolans verksamhet och resultat samlas in i så begränsad omfattning.

Resultat från internationella jämförande studier har kommit att bli den viktigaste källan till information om kunskapsutvecklingen i den svenska skolan. Efter 1995 har antal och frekvens av sådana studier ökat. Det faktum att studierna ger information om resultatförändringar inom länder har gjort dem mer användbara för att beskriva och analysera kunskapsutvecklingen i Sverige och att jämföra den med utvecklingen i andra länder. Även om de internationella studierna ger värdefull information är de dock begränsade till att omfatta endast vissa delar av vissa ämnesområden, och tyngdpunkterna i proven motsvarar inte alltid fullt ut tyngdpunkterna i den svenska läroplanen. Det finns därför anledning att överväga om Sverige ska komplettera de internationella studierna med ytterligare stickprovsbaserade undersökningar av kunskapsutvecklingen inom olika områden.

I utvärderingen behandlar Nusche m.fl. (2011) en lång rad andra aspekter, som lokal utvärdering på skol- och kommunnivå och lärarbedömning, och de lägger fram flera förslag till förbättringar. Ett huvudförslag är dock att Sverige behöver utveckla ett koherent ramverk för uppföljning och utvärdering som integrerar relativt fristående komponenter och nivåer till ett sammanhängande system.

Betyg från gymnasieskolan har stor betydelse för möjligheten att få tillträde till högre utbildning. De avgör allmän och särskild behörighet för tillträde till utbildningar och de är ett viktigt instrument för att rangordna dem som söker till utbildningar när antalet sökande är fler än utbildnings-

platserna. Bristerna i likvärdighet i betygssättningen gör dock att betygens användbarhet för antagning kan ifrågasättas. Det finns även andra problem med användning av betyg i detta sammanhang, som att betyg från olika betygssystem eller enligt olika betygsskalor inte fullt ut är jämförbara. Sedan några år tillbaka sker vidare meritvärderingen av betygen enligt en modell där sökande får tillgodoräkna sig extrapoäng (så kallade meritpoäng, se ordlista) om de läst vissa kurser inom gymnasieskolan. Som vi diskuterar mer utförligt i kapitel 4 finns det anledning att befara att detta försämrar både rättvisa och effektivitet när betyg används för urval till högre utbildning. Det huvudsakliga alternativet i detta sammanhang är Högskoleprovet, som är konstruerat för att mäta kunskaper och färdigheter som är betydelsefulla för framgång i högre studier. Men forskningen visar att även Högskoleprovet är problematiskt när det används för urval.

Det finns alltså skäl att misstänka att det svenska antagningssystemet till högre utbildning inte är så effektivt som det skulle kunna och borde vara, vare sig på individ- eller systemnivå, liksom att vissa grupper av sökande drabbas av orättvisor. Med tanke på den stora individuella, praktiska, ekonomiska och politiska betydelse som antagningssystemet har, framstår det därför som nödvändigt att identifiera brister i såväl enskilda komponenter som i hela systemets funktions sätt, liksom att skissera på möjliga sätt att i ett helhetsperspektiv komma tillrätta med bristerna.

Metodutvecklingen inom den pedagogiska mätläran

Under de senaste decennierna har det skett en betydande metodutveckling inom området pedagogiska mätningar. Denna metodutveckling, vars mer konkreta innebörd vi kommer att återkomma till, gör det bland annat möjligt att konstruera prov som består av olika uppgifter men som ändå ger mätresultat på samma skala. Det innebär i sin tur att elever kan besvara en mindre del av en större mängd uppgifter och att resultaten sedan kan sättas samman till en helhet. En annan fördel är att det är förhållandevis enkelt att konstruera prov som ger jämförbara resultat över tid, vilket förbättrar möjligheterna att undersöka förändringar i kunskapsnivåer på systemnivå.

Internationellt används dessa metoder i ökande utsträckning men de har ännu inte börjat användas i Sverige. Här finns dock möjligheter som det finns anledning att ta tillvara vid utformningen av ett svenskt system för kunskapsbedömningar.

Disposition

Vi börjar med att i kapitel 2 diskutera kunskapsbedömning på individnivå, och då framför allt likvärdighet i summativ bedömning i form av betyg och nationella prov. Vi fokuserar på likvärdighetsfrågor både när det gäller skillnader i kunskapsbedömningar mellan lärare och skolor, och stabilitet i betyg och nationella prov över tid.

16

I kapitel 3 går vi in på kunskapsbedömning på systemnivå, och särskilt på nationell kunskapsbedömning. Efter några exempel på nationella kunskapsbedömningssystem diskuterar vi hur ett svenskt system med rullande kunskapsbedömningar inom olika ämnen skulle kunna utformas. Vi diskuterar också vilken roll de nationella proven kan spela för att ge information om nationell utveckling av kunskapsnivåer. Vår slutsats är att den är begränsad. Däremot öppnar sig intressanta möjligheter att knyta samman utveckling av nationella prov med det nationella kunskapsbedömningssystemet, vilka vi kortfattat diskuterar.

I kapitel 4 tar vi upp antagningssystemet till högre utbildning och pekar på gymnasiebetygens viktiga roll, både för allmän och särskild behörighet och för urval. Likvärdighetsproblemen i betygen utgör emellertid ett hot mot betygens funktion i antagningssystemet. Ett annat hot som vi diskuterar är det nyligen introducerade meritvärderingssystemet som ger vissa sökande möjlighet att tillgodoräkna sig extrapoäng på grundval av kurser som de valt i gymnasiet. Vi diskuterar också de negativa konsekvenser som förändringar av betygssystem och betygsskalor medför. Ett alternativt urvalsinstrument är Högskoleprovet, vars prognosförmåga dock är så svag att det borde användas så lite som möjligt. Huvudslutsatsen är att betygens roll i antagningssystemet bör stärkas, under förutsättning att likvärdighetsproblemen kan bemästras.

I kapitel 5 sammanfattar vi vår diskussion och fokuserar på hur likvärdigheten i kunskapsbedömningarna kan stärkas.

I slutet av rapporten finns en ordlista som förklarar och kommenterar vissa av de begrepp vi använder.

2. Kunskapsbedömning på individnivå

Vi inleder med att redogöra för betygens funktioner och diskuterar två olika system för betygssättning. Vi tar sedan upp de nationella proven och deras funktioner. Avslutningsvis diskuterar vi vilka problem som är förknippade med betyg och nationella prov och vilka förändringar som är nödvändiga för att få ett mer rättssäkert och effektivt system för kunskapsbedömning på individnivå.

Betyg

Betyg används i utbildningssammanhang för att ge en sammanfattande bedömning av en elevs kunskaper och färdigheter. Dessa låter sig dock inte enkelt mätas eller bedömas på en fastställd skala, vilket innebär att bedömningarna måste göras mot jämförelse- eller referenspunkter. Beroende på hur dessa jämförelsegrunder väljs erhålls olika betygssystem. Vi kan skilja mellan två huvudvarianter av betygssystem:

- * *Normrelaterad* (eller grupprelaterad) betygssättning som bygger på jämförelse mellan individer inom en grupp (till exempel alla de som ett visst år läser ett visst ämne).
- * *Målrelaterad* betygssättning som är baserad på bedömning av individens prestation i förhållande till i förväg preciserade mål.

Betygen fyller flera olika funktioner. De används för att ange behörighet för tillträde till utbildningar och som grund för urval till nästa nivå i utbildningssystemet, från grundskola till gymnasieskola och från gymnasieskola till högre utbildning. En annan funktion är att ge information om utbildningens resultat både på nationell och lokal nivå (klass, skola och kommun). Betygen ska också ge information om elevens studieresultat till eleven själv och till elevens föräldrar. Vissa ser också betygen som ett belöningsystem

för elevernas studieprestationer, medan andra är mer tveksamma till om betygen kan och bör uppfylla denna funktion.

För att betygen ska kunna uppfylla dessa funktioner väl måste de uppfylla högt ställda kvalitetskrav. De måste ge rättvisa uttryck åt elevernas kunskaper och de måste vara jämförbara både mellan lärare och skolor och över tid. Det är också viktigt att betygen förmår att förutsäga studieframgång när de används för urval, liksom att de behörighetsregler som grundas på betygen uttrycker utbildningarnas faktiska krav på förkunskaper. Informationen behöver också vara av sådan art att den kan användas för att inrikta det fortsatta studiet på de områden som behöver förbättras. För att betygen ska kunna ge information om kunskapsutvecklingen på nationell och lokal nivå måste de kunna återspegla förändringar i kunskapsnivåer. Dessa delvis motstridiga krav är inte lätta att uppfylla. Vi diskuterar nedan hur de två betygssystemen förmår att göra det.

DET RELATIVA BETYGSSYSTEMET

Fram till mitten av 1990-talet användes det så kallade relativa, normrelaterade, betygssystemet i den svenska skolan. Det stipulerade hur stor andel av det totala antalet elever som skulle erhålla vart och ett av de olika betygsstegen. Denna fördelning avsåg samtliga elever i landet som läste det aktuella ämnet, medan betygsfördelningen inom den enskilda klassen kunde förväntas vara annorlunda. Som hjälpmedel för att fastställa klassens nivå i relation till den nationella nivån fanns i vissa ämnen standardprov i grundskolan och centrala prov i gymnasiet.

Den femgradiga sifferskalan infördes 1962. Men redan 1949 fattade riksdagen beslut om att införa det relativa betygssystemet i folkskolan, vilket också gjordes under början av 1950-talet. Detta skedde på grundval av förslag som lades fram i en utredning 1942 under ledning av Frits Wigforss (SOU 1942:II). En av anledningarna till utredningen var att man konstaterat att betygen i folkskolan brast i jämförbarhet, och att de därför inte kunde användas för urval till realskolan.

Utredningen diskuterar de norm- och målrelaterade betygssystemens för- och nackdelar, och drar slutsatsen att målrelaterade betyg inte kan göras jämförbara utan att mycket noga precisera kursernas innehåll och kunskapskrav: "Det är sannolikt omöjligt att genom allmänna uttryck så precisera den kvalitativa godhet som fordras för de olika betygsgraderna, att betygen därigenom kunna bli jämförbara" (SOU 1942:II, s. 53). De undersökningar som utredningen genomförde visade på en mycket stor variation i olika lärares betygssättning, men också att lärarna hade god förmåga att be-

döma eleverna i klassen i förhållande till varandra. Utredningen drog därför slutsatsen att lärarna som hjälp för sin betygssättning måste ha tillgång till standardprov som kan ange klassens prestationsnivå.

Det normrelaterade betygssystem som föreslogs av utredningen från 1942 fortlevde med endast smärre förändringar ända till 1994. Men trots sin livslängd har det relativa betygssystemet inte varit oomtvistat (se Gustafsson, 2006b, för en mer utförlig beskrivning av kritiken). Ett problem som uppmärksammades var att de relativa betygen inte säger något om den faktiska nivån på elevernas prestationer, utan endast hur en elev presterar i förhållande till andra elever. Förändringar i kunskaper och färdigheter från ett år till ett annat i landet som helhet kunde inte avläsas i form av skillnader i betygsnivåer. Ett annat problem som lyftes fram var att det relativa betygssystemet i princip förutsätter att alla elever i landet läser ämnet, eftersom det endast är då som det kan finnas någon grund för normalfördelningsantagandet (se ordlista). Ett ytterligare problem som identifierades var att standardprov endast förekom i tre ämnen (svenska, matematik och engelska), vilket gjorde det svårt att uttala sig om klassens nivå i förhållande till andra klasser för övriga ämnen. I debatten framhölls också att normalfördelningen ibland felaktigt applicerades på skol- eller klassnivå ("femmorna tog slut") samt att systemet i sig uppmanade till konkurrens mellan elever.

DET MÅL- OCH KUNSKAPSRELATERADE BETYGSSYSTEMET

Under 1990-talets början utvecklades och infördes på kort tid ett nytt betygssystem, det mål- och kunskapsrelaterade (se Gustafsson, 2006b, för en översiktlig beskrivning av dess tillkomst). År 1994 fattade riksdagen beslut om att införa målrelaterade betyg för både grundskola och gymnasieskola i anslutning till de nya läroplanerna.¹ För grundskolan fastställdes en tregradig betygsskala med betygsstegen Godkänd, Väl godkänd och Mycket väl godkänd. I gymnasieskolan användes även Icke godkänd. Betyg gavs första gången på höstterminen i årskurs 8, och därefter gavs betyg varje termin. För slutbetyget från vårterminen i årskurs 9 fastställdes mål och kriterier centralt, medan kriterierna för terminsbetygen utvecklades lokalt. Betygsättningen i årskurs 9 hade också stöd av nationella ämnesprov i svenska/svenska som andraspråk (fortsättningsvis hänvisar vi till dessa med samlingsbeteckningen svenska), engelska och matematik. För gymnasieskolans del beslöts att studierna skulle vara kursutformade, med ett betyg för varje kurs, och att nationella kursprov skulle stödja betygssättningen i svenska, engelska och matematik.

¹ Lpo 94 för grundskolan och Lpf 94 för gymnasieskolan.

De centrala begreppen i detta betygssystem är mål och kriterier. Målen anger vad eleven ska kunna och betygskriterierna beskriver de kvaliteter i måluppfyllelsen som svarar mot olika betyg, det vill säga hur väl eleverna ska kunna utföra det som finns formulerat i målen. Kriterierna utgör grunden för en rättvis och likvärdig betygssättning, men Skolverket (1996) underströk – och understryker alltjämt – att kriterieformuleringarna inte kan vara så precisa att likvärdighet i betygssättningen kan uppnås endast genom dessa. En stor del av ansvaret för att uppnå en likvärdig betygssättning lades således på lärarna och skolorna själva, exempelvis genom att organisera professionella samtal om hur kriterierna skulle tolkas.

Det nya betygssystemet kom dock att möta problem inom både grund- och gymnasieskolan redan vid implementeringen, som skedde med kort varsel och med begränsad information till och kompetenshöjning av de lärare som skulle tillämpa systemet. Det har också konstaterats att det fanns brister i konstruktionen av själva målsystemet och systemet med kriterier, liksom brister i likvärdigheten i betyg satta av olika lärare vid olika skolor. I en utredning (SOU 2007:28) inventeras dessa problem och den lägger fram förslag till hur systemet ska kunna förbättras. Bland annat pekar utredaren på att kursplanerna i alltför hög grad saknade hänvisningar till ämnesinnehåll och arbetssätt, liksom att läroplanen erbjöd en mångfald av mål och en svårbemästrad komplexitet där vissa mål uttrycks mycket abstrakt och otydligt. Vidare, menar utredaren, bör kursplanemålen tydligt ange vilka ämneskunskaper som undervisningen ska inriktas mot, och kursplanerna ska ange tydliga krav för godtagbara kunskaper för olika betygssteg. Genom att mål- och kunskapskrav anges i kursplanerna blir också mål- och resultatstyrningen tydligare (SOU 2007:28).

BETYG ENLIGT 2011 ÅRS LÄROPLANER

Från 2011 gäller nya läroplaner för grund- och gymnasieskolan som har sin grund i ovan nämnda utredning. Läroplanen för grundskolan, Lgr11, består av tre delar. Den första är *Skolans värdegrund och uppdrag*. Här anges de grundläggande värden som skolan vilar på, det uppdrag som skolan har, och de principer som ska följas när uppdraget fullgörs. Den andra delen har rubriken *Övergripande mål och riktlinjer för utbildningen*. Här anges de normer och värden samt de kunskaper som alla elever bör ha utvecklat när de lämnar grundskolan, bland annat generella kompetenser som att lära sig att utforska, lära och arbeta både självständigt och tillsammans med andra, samt att befästa en vana att självständigt formulera ståndpunkter grundade på såväl kunskaper som förnuftsmässiga och etiska överväganden. Dessa

övergripande mål (eller kompetensmål) avser sådana kunskaper och färdigheter som inte är direkt knutna till ett visst undervisningsämne.

I den tredje delen av läroplanen presenteras kursplanerna, vilka alla har samma grundläggande struktur. Inledningsvis karaktäriseras ämnet på ett fåtal rader, och därefter beskrivs mer utförligt syftet med undervisningen. Så kallat centralt innehåll anges därefter för de olika stadierna, och sedan följer kunskapskraven för de olika betygsstegen i de olika årskurserna.

Med de nya läroplanerna infördes också en ny sexgradig betygsskala med beteckningarna A, B, C, D, E och F. A–E betecknar godkända resultat och F betecknar icke godkänt resultat. Ett streck ska användas i fall då inget betygsunderlag finns att tillgå. Den nya skalan gäller från läsåret 2012/2013. Från och med detta läsår ges betyg från årskurs 6.

Det är intressant att närmare granska hur kunskapskraven är formulerade. Vi använder engelska i årskurs 6 som exempel. Först beskrivs i ett stycke de kunskapskrav som gäller för receptiva färdigheter (lyssna och läsa), och därefter i ett andra stycke kraven på produktiva och interaktiva färdigheter (tala/samtala och skriva). Ett tredje stycke handlar om aktivt bruk av information om företeelser där engelska används samt jämförelse med egna kunskaper och erfarenheter.

För betyget E anges exempelvis följande kunskapskrav på förmåga att förstå talad engelska: ”Eleven kan förstå **det mest väsentliga av innehållet**² i tydligt talad, enkel engelska i lugnt tempo samt i enkla texter om vardagliga och välbekanta ämnen. Eleven visar sin förståelse genom att **i enkel form** redogöra för och kommentera innehållet ...”. För betyget C anges på samma punkt följande kunskapskrav: ”Eleven kan förstå **det huvudsakliga innehållet och uppfatta tydliga detaljer** i tydligt talad, enkel engelska i lugnt tempo samt i enkla texter om vardagliga och välbekanta ämnen. Eleven visar sin förståelse genom att **i enkel form** redogöra för och kommentera innehåll **och detaljer** ...”. För betyget A anges motsvarande kunskapskrav på följande sätt: ”Eleven kan förstå **helheten och uppfatta väsentliga detaljer** i tydligt talad, enkel engelska i lugnt tempo samt i enkla texter om vardagliga och välbekanta ämnen. Eleven visar sin förståelse genom att **översiktligt** redogöra för och kommentera innehållet **och detaljer** ...”.

Av detta exempel framgår att skillnaden mellan kraven för de olika stegen beskrivs med hjälp av komparationer (till exempel för E ”**det mest väsentliga av innehållet**”, för C ”**det huvudsakliga innehållet och uppfatta**

2 Fetstil i originalet.

tydliga detaljer”, och för A ”helheten och uppfatta väsentliga detaljer”). Dessa jämförelser är dock i sig själva tämligen oprecisa och abstrakta, och det är ytterst tveksamt om de förmår att ge ett tillräckligt stöd för likvärdig betygssättning.

Betyget D innebär enligt kursplanen att kunskapskraven för betyget E och till övervägande del för betyget C ska vara uppfyllda, och betyget B innebär att kunskapskraven för betyget C och till övervägande del för betyget A ska vara uppfyllda. Det bör också påpekas att för att ett visst betyg ska ges, måste samtliga kunskapskrav för detta betygssteg vara uppfyllda. Om en elev inte fullt ut uppfyller alla kunskapskrav för exempelvis de receptiva färdigheterna kan inte eleven kompensera för detta genom att ha mycket goda produktiva färdigheter, utan ska tilldelas ett betyg som reflekterar de svagare receptiva färdigheterna. Vanligtvis är dock en elev bra på vissa saker och mindre bra på andra saker inom ett ämne, varför bedömningen av en elevs prestationer i ett visst ämne i stället borde bygga på en sammanvägning av kvaliteten av elevens prestationer inom ämnets olika delar. Som systemet nu är utformat fokuseras elevens svagaste sidor. Dessutom blir systemet känsligt för att en elev mer eller mindre slumpmässigt inte når upp till samma nivå på alla aspekter av kunskapskraven. Det finns skäl att anta att vissa lärare tvekar inför att använda sig av en sådan princip, vilket i så fall blir en källa till likvärdighetsbrister i betygssättningen.

Alltsedan de målrelaterade betygen introducerades har frågan om deras likvärdighet varit central, och information från många olika källor (se nedan) pekar på betydande brister. Huvudskälet till detta är att den viktigaste grunden för att åstadkomma likvärdighet i betygssättningen är en för alla lärare gemensam uttolkning av de mål och kunskapskriterier/kunskapskrav som anges i läro- och kursplanerna och dessa kan inte ”... så precisera den kvalitativa godhet som fordras för de olika betygsgraderna, att betygen därigenom kunna bli jämförbara”, för att återigen citera SOU 1942:11 (s. 53). Även om de nya läro- och kursplanerna skulle ge ett förbättrat stöd för en likvärdig betygssättning pekar ovanstående redovisning och diskussion på att det finns anledning att tvivla på om de kommer att innebära några förbättringar i detta hänseende.

Vid sidan av de kriterier/kunskapskrav som anges i kursplanerna har lärarna tillgång till flera olika hjälpmedel för att åstadkomma en likvärdig betygssättning, som kommentarmaterial av skilda slag samt bedömningsanvisningar och nationella prov. Inte minst de senare anses spela en mycket viktig roll som stöd för lärarnas betygssättning.

Nationella prov

Sverige har en lång tradition av centralt tillhandahållna prov som stöd för lärares betygssättning (Marklund, 1987). Som tidigare nämnts tillhandhölls standardprov i vissa ämnen från och med 1950-talet som en del av det relativa betygssystemet. Antalet ämnen och årskurser varierade, men standardproven, liksom motsvarande så kallade centrala prov i gymnasiet, fanns med som riktningsgivare för betygen fram till införandet av de nationella proven i anslutning till det målrelaterade systemet i mitten av 1990-talet.

23

DE NATIONELLA PROVEN S SYFTEN

De centralt tillhandahållna proven i det normrelaterade systemet hade sitt bestämda syfte, nämligen att verka som riktningsgivare för betygssättning på gruppnivå. Indirekt ansågs det också viktigt att proven på ett rimligt sätt återspeglade kursplanernas innehåll, men detta var inte uttryckt i anvisningarna för provkonstruktion. I samband med övergången till nationella prov i mitten av 1990-talet inträdde en markant förändring, då provens pedagogiska betydelse som stöd för lärande och undervisning fick en framskjuten position i syftesbeskrivningen. Under ett antal år angavs fem syften för de nationella proven, nämligen att

- * 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
- * ge underlag för en analys av i vilken utsträckning kunskapsmålen nås på skolnivå, på huvudmannanivå och på nationell nivå.

Vidgningen av det traditionella bedömningsbegreppet var inte någon isolerad eller lokalt svensk företeelse, utan starka strömningar i denna riktning var märkbara på flera håll i västvärlden. Caroline Gipps bok *Beyond Testing* (1994) kom till exempel att i hög grad präglade de tidiga diskussionerna kring de nationella proven, liksom något senare arbeten av bland andra Paul Black och Dylan Wiliam, som liksom Gipps var medlemmar i Assessment Reform Group i Storbritannien (se Gardner, 2012). Fokus i denna grupps arbete var att uppmärksamma den formativa bedömningens pedagogiska potential och att därmed verka för det som kom att rubriceras som ”bedömning för lärande”. I linje med detta betonades de svenska nationella provens roll av att vara goda förebilder vid implementeringen av de nya styrdokumentet, medan de likvärdighets- och uppföljningsrelaterade syftena tonades ner.

Under 1990-talet fördes intensiva diskussioner om de nationella provens utformning i relation till de vidgade syften som formulerats. Skolverket ville tona ner mätaspekten och provkaraktären och i stället betona provens implementerande funktion och att de skulle bli en naturlig och integrerad del av undervisningen. I detta betonades starkt validitetsaspekten, som ofta, och delvis oriktigt, ställdes mot reliabilitet.³ Detta tog sig många konkreta uttryck. Ett var att poängbegreppet ansågs reflektera ett ålderdomligt sätt att värdera kunskap och det användes därför inte i vissa prov; poänggränser för olika betygssteg var angivna som rekommendationer snarare än som strikta gränser i somliga prov; och sekretessaspekter tonades ner för att främja en didaktisk användning av de nationella proven.

Ambitionen att tona ner provkaraktären hos de nationella proven yttrade sig också i en preferens för ”performance”-prov, eller ”autentiska” prov, där eleverna får ett begränsat antal uppgifter, som i allmänhet kräver en relativt omfattande produktion av text eller tal, antingen ensam eller tillsammans med andra elever. För att bedöma elevproduktionen krävs sedan en kompetent bedömare. Denna typ av uppgifter ger möjlighet att komma åt kvaliteter i elevernas resultat som inte 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 svenska eller engelska, där det inte torde vara möjligt eller meningsfullt att använda sig av någon annan typ av prov än att låta eleverna producera uppsatser. Det är dock inte lika självklart att det skulle vara en bra metod att låta eleverna skriva längre texter om exempelvis demokratibegreppet i ett prov i samhällskunskap, eftersom resultatet skulle kunna bli mer påverkat av elevens skrivförmåga än av insikterna i demokratibegreppet.

En annan yttring för att tona ner provkaraktären var – och är fortfarande – att inga standardiserade specifikationer för nationell provutveckling togs fram. Detta innebar en stor variation i provens utformning och utvecklingsprocesserna vid de olika universitetsinstitutioner som ansvarade för proven. I vissa fall byggde proven på mindre utprövningar vid ett fåtal skolor och i andra fall baserades de på stora utprövningar i slumpvisa urval. Vid vis-

3 Validitet avser frågan huruvida forskaren mäter det hon eller han har tänkt mäta; mäter exempelvis ett diagnostiskt lästest verkligen skillnader i läsförmåga? Reliabilitet handlar om huruvida resultaten blir desamma om undersökningen genomförs på nytt; begreppet beskriver alltså hur säker mätningen är. För en fördjupad beskrivning, se ordlista.

sa institutioner användes breda metodarsenaler för att analysera provdata, medan andra institutioner endast använde enklare analyser.

I takt med att frågan om de målrelaterade betygens bristande likvärdighet alltmer kommit att uppmärksammas har också frågan om de nationella provens mångfaldiga syften kommit att fokuseras. Även denna fråga diskuteras i SOU (2007:28), där slutsatsen är att de nationella proven i första hand ska utgöra stöd för likvärdig betygssättning och att de ska ge underlag för att analysera i vilken utsträckning kunskapskraven uppfylls. I linje med detta formuleras de nu aktuella syftena med de nationella proven på följande sätt på Skolverkets hemsida:

Syftet med de nationella proven är i huvudsak att

- * stödja en likvärdig och rättvis bedömning och betygssättning
- * ge underlag för en analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, på huvudmannanivå och på nationell nivå.

De nationella proven kan också bidra till

- * att konkretisera kursplanerna och ämnesplanerna
- * en ökad måluppfyllelse för eleverna.

Trots detta är utformningen av de nationella proven delvis oförändrad i förhållande till de principer för provkonstruktion som etablerades när det mål- och kunskapsrelaterade betygssystemet infördes under 1990-talet.

OECD:S KRITIK MOT DE NATIONELLA PROVENS UTFORMNING

Nusche m.fl. (2011) konstaterade att de svenska nationella provens tyngdpunkt på komplex elevproduktion ger anledning att befara att dessa prov har begränsad reliabilitet, generaliserbarhet och validitet. Det finns en omfattande forskning som visar att ”performance”-prov (se ordlista) (eller autentiska prov, som vi hädanefter kommer att benämna dem) ofta inte mäter de förmågor och processer som de har för avsikt att mäta, bland annat av det skälet att kravet att producera text eller tal ställer krav på andra kunskaper och förmågor än de som provet syftar till att mäta. Nusche m.fl. (2011, s. 51) understryker dock att bedömningar av validiteten i de nuvarande nationella proven skulle kräva mer djupgående undersökningar. I sin genomgång pekar de också på att autentiska prov är mindre tillförlitliga och svårare att generalisera utifrån än andra provtyper, beroende på att antalet uppgifter i sådana prov i allmänhet är begränsat.

Trots kritiken mot det svenska nationella provsystemet föreslår dock

inte Nusche m.fl. (2011) att de autentiska proven ska avskaffas, eftersom de kan ge information som inte är åtkomlig på andra sätt. En bättre lösning skulle vara att komplettera de befintliga proven med mindre komplexa uppgifter som ställer lägre krav på bedömning av elevernas svar. Sådana kombinationer av de båda provtyperna skulle kunna ge fördelar i form av både god reliabilitet och god validitet. Vi instämmer med detta och kommer att vidareutveckla resonemanget i kapitel 3.

SKOLINSPEKTIONENS OMRÄTTNINGAR AV NATIONELLA PROV

Ett problem med de autentiska proven är att de ställer höga krav på bedömningen av den omfattande och komplexa elevproduktionen i form av bland annat text och tal. Skolinspektionen har flera år på uppdrag av regeringen rättat om ett slumpmässigt urval av de nationella proven för att kontrollera hur pålitliga de ursprungliga bedömningarna är (Skolinspektionen, 2010, 2011, 2012, 2013). Nusche m.fl. (2011) genomförde sin studie av Sveriges utvärderingssystem strax efter att Skolinspektionen (2010) redovisat resultaten från den första omrättningsomgången, vilka pekar på att det finns stora problem med pålitlighet och objektivitet i lärarnas bedömningar.

Med hjälp av externa bedömare har omrättningar genomförts av relativt stora slumpmässiga urval av prov där elevens lärare redan gjort en ursprunglig bedömning. Omrättningarna har genomförts på delprovsnivå. Endast delprov där elevernas prestationer finns dokumenterade har kunnat inkluderas, vilket innebär att delprov med exempelvis muntlig kommunikation, laborationer och liknande elevproduktion har fått uteslutas. I första hand har alltså skriftliga delprov inkluderats i omrättningarna, med antingen längre elevsvar i form av uppsatser, korta svar i form av några få ord, eller i form av val mellan på förhand givna svarsalternativ.

Skolinspektionen (2013) hävdar att de resultat som framträder över de fyra omgångarna är tämligen entydiga. För uppsatsprov i årskurs 9 och i gymnasiet föreligger, i synnerhet för ämnet svenska, stora skillnader på så sätt att lärarnas ursprungsbedömningar har ett högre medelvärde än de externa bedömningarna. Vidare är graden av överensstämmelse kring bedömning av de enskilda uppsatserna låg mellan de interna och de externa bedömare. För uppsatsdelen i svenska i gymnasieskolan sammanfaller exempelvis betygen satta av läraren och omdömarens endast i 38 procent av fallen. Vidare visar analyserna att graden av överensstämmelse är lägst för betyget MVG. För prov som endast kräver korta svar är medeltalsskillnaden mellan de två bedömkategorierna mindre, och graden av överens-

stämme mellan de två bedömarna är större. För praktiskt taget samtliga undersökta delprov ligger dock lärarnas bedömningar högre än de externa bedömningarna.

Skolinspektionens (2013) rapportering lämnar dock mycket övrigt att önska. Huvuddelen av de empiriska resultaten presenteras i en svåröverskådlig tabellbilaga, medan tolkningar, slutsatser och rekommendationer återfinns i själva rapporten, där dock resultatredovisningen ger ett ostrukturerat intryck. Två typer av avvikelser mellan ursprungsbedömning och omrättning är särskilt intressanta. Den ena är om det finns systematiska skillnader i grad av stränghet mellan de två bedömningarna, och den andra avser graden av överensstämmelse på elevnivå mellan de två provbetygen. Tyvärr brister Skolinspektionens redovisningar när det gäller att särskilja dessa, och tyngdpunkten i redovisningen ligger på överensstämmelsen på elevnivå. Utifrån sammanställningar av data från 2010 och 2011 års omrättningar som Hinnerich och Vlachos (2013, tabell 1-3) gjort har vi därför beräknat storleken på medelvärdesskillnaden mellan de två provbetygen (se tabell 1).

Dessa resultat skiljer sig från dem som Skolinspektionen (2013) presenterar. Någon generell skillnad i resultat mellan uppsatser och prov med kortvarsformat finns inte, men däremot finns det skillnader mellan ämnen. För engelska är skillnaderna mellan bedömningarna små för båda provtyperna i både årskurs 9 och gymnasiet, medan skillnaderna för svenska genomgående är stora, oberoende av provtyp. För de naturvetenskapliga ämnenas teoriprov i årskurs 9 är skillnaderna i generositet i bedömningarna nästan lika stora som skillnaderna i svenska. Matematik uppvisar en splittrad bedömning med små skillnader för två av delproven i årskurs 9, en större skillnad för det tredje delprovet i årskurs 9, och en betydande skillnad för gymnasieprovet.

För att förstå de stora skillnaderna i bedömningarnas generositet är det nödvändigt att analysera likheter och skillnader mellan olika ämnen. Hinnerich och Vlachos (2013) pekar på att de naturvetenskapliga proven är relativt nya och att ursprungsbedömarna var ovana vid uppgiften, vilket kan vara en tänkbar förklaring till skillnaderna mellan bedömningarna. En annan möjlig förklaring kan vara att dessa prov kräver relativt omfattande skriftliga svar som ska tolkas och bedömas. Den stora skillnaden i utfall för de olika matematikproven pekar också på att olika provformat har betydelse för hur stora skillnader som observeras mellan de två bedömningarna.

För att kunna förklara den stora skillnaden i resultat för svenska och engelska skulle vi behöva genomföra en detaljerad analys och jämförelse mellan provuppgifter och bedömningsanvisningar i de två ämnena, men

Tabell 1. Medelvärdeskillnader mellan ursprungsbedömning och kontrollbedömning.

Årskurs 9		Gymnasiet	
Engelska kortsvarsdel B	0.07	Engelska kortsvarsdel	0.08
Engelska uppsatsdel C	0.12	Engelska uppsatsdel	0.10
Svenska kortsvarsdel B	0.33		
Svenska uppsatsdel C	0.30	Svenska uppsatsdel	0.43
Matematik B1	0.07	Matematik 1 och 2	0.32
Matematik B2	0.09		
Matematik C	0.18		
Fysik Teori	0.36		
Fysik Planering	0.11		
Biologi Teori	0.28		
Biologi Planering	-0.14		
Kemi Teori	0.32		
Kemi Planering	-0.10		

Not. Differensen mellan betygen är beräknad som skillnaden mellan lärarbetyg och kontrollbetyg, och är uttryckt i termer av standardavvikelseenheter* (Cohens *d*). En positiv differens betyder att elevens lärare gjort en generösare bedömning av elevens provresultat än den externa bedömaren.

* Se ordlista för förklaring.

det är inte är möjligt här. Viss information går dock att få från utfallet av den andra omrättning som genomförs då det föreligger en stor diskrepans mellan ursprungs rättning och kontrollrättning. Skolinspektionen (2013, diagram 4, s. 18) redovisar hur stor andel andra-omrättningar som stödjer den ursprungliga rättningen eller omrättningen. För prov av kortsvarskaraktär i årskurs 9 stöds omrättningen oftare än ursprungsrättningen i både engelska och svenska (knappt 60 procent). En tänkbar tolkning är att bedömarna gjort slarvfel, de har läst fel eller missat saker i den ursprungliga bedömningen. Det var stor skillnad mellan ämnena i hur stor andel av proven som gick till en andra omrättning: 12 procent av proven i svenska men endast 2 procent av proven i engelska. Detta tyder på att kortsvarsuppgifterna i

svenskprovet lättare inbjuder till slarvfel vid bedömningen än de engelska kortsvarsuppgifterna.

Skolinspektionen visar också att för uppsatsprovet i engelska i årskurs 9 var det en lika stor andel av andra-omrättningarna som stödde ursprungs-rättningen som kontrollrättningen. För uppsatsprovet i svenska var det en mindre andel som stödde ursprungs-rättaren än som stödde kontrollrättaren. Effekten av denna skillnad förstärks av att det även här var en betydande skillnad i andelen prov som gick till en andra omrättning (23 procent för svenska jämfört med 7 procent för engelska). Dessa resultat pekar på att det är betydligt svårare att bedöma uppsatser i ämnet svenska än i engelska i årskurs 9. En förklaring kan vara att bedömningsanvisningarna är tydligare och enklare att följa i engelska. En annan förklaring kan vara en högre grad av komplexitet i kursplanen för svenska.

MISSGYNNAS VISSA ELEVGRUPPER AV LÄRARNAS BEDÖMNINGAR?

Huruvida egenskaper hos eleverna, lärarna och skolorna påverkar utfallet av lärarnas bedömningar har också undersökts närmare, senast av Hinnerich och Vlachos (2013) som har analyserat samtliga tillgängliga prov. Deras analyser visar att lärarnas ursprungliga provbetyg var något lägre än omrättarnas för pojkar, elever med utländsk bakgrund, elever vars föräldrar saknar högre utbildning och för elever födda sent på året. Skillnaderna var mer markanta för uppsatser och mindre markanta, och ofta inte statistiskt säkerställda, för delprov med kortsvarsformat. Dessa resultat pekar på att lärarna i sina bedömningar i viss utsträckning missgynnar elevgrupper som tenderar att uppvisa relativt svaga studieresultat. I linje med detta visade också analyserna att gymnasieelever med höga grundskolebetyg tenderar att få en generösare provbedömning av sina lärare än av omrättarna.

Hinnerichs och Vlachos (2013) analyser visar också att fristående skolor i genomsnitt sätter högre provbetyg än kommunala skolor, relativt omrättarnas bedömningar. Denna effekt var starkare än de effekter som identifierades för olika elevenskaper, och i stort sett hela den observerade skillnaden i betygsresultat mellan fristående och kommunala skolor, efter kontroll för skolornas elevsammansättning, kan förklaras av att lärare i fristående skolor bedömer elevernas nationella prov mer generöst.

Bortsett från effekten av huvudman och höga grundskolebetyg är de skattade effekterna mycket svaga. En del är vidare statistiskt säkerställda antingen endast på gymnasiet eller i årskurs 9. Att lärarnas bedömningar tycks gynna elever med goda förutsättningar och goda prestationer är dock

oväntat. Tidigare forskning har snarare pekat på att lärares betygssättning tenderar att gynna elever vars resultat riskerar att falla under gränsen för godkänt (Klapp Lekholm & Cliffordson, 2008).

SKOLINSPEKTIONENS FÖRSLAG TILL ÅTGÄRDER

Med stöd i resultaten från omrättningarna lägger Skolinspektionen (2013) fram flera förslag till åtgärder. Den fastslår att överensstämmelsen mellan lärarnas bedömning och omrättningen bör uppgå till minst 70 procent för uppsatsdelarna, medan överensstämmelsen för kortvarsdelarna bör uppgå till minst 80 procent beroende på hur komplexa bedömningarna är. Avvikelser som är större än så anser Skolinspektionen vara ”obefogade”.

Detta är ett mycket olyckligt förslag. För det första är de fastslagna gränserna tämligen godtyckliga och endast svagt förankrade i tidigare forskning. För det andra utgår de angivna procenttalen från IG–MVG-skalan, och de kommer inte att vara överförbara på den mer finfördelade F–A-skalan. För det tredje utgår förslaget från att den bristande överensstämmelsen till stor del beror på lärarnas ofullkomliga bedömningar, och att detta skulle gå att åtgärda genom bättre utbildning och mer omsorgsfullt iakttagande av bedömningsanvisningarna. Men det är väl känt från forskningen om bedömningar av kvalitet i exempelvis uppsatser på gymnasie- eller högskolenivå, och ansökningar om forskningsmedel, att oenighet mellan bedömare i första hand har sin grund i motiverbara skillnader i värderingar och använda kvalitetskriterier, och endast i mindre utsträckning i ofullkomligheter i den enskilda bedömningen (se t.ex. Jayasinghe m.fl., 2001). Den mest framkomliga vägen att öka säkerheten i bedömningar torde således vara att väga samman flera oberoende bedömningar.

Skolinspektionen har valt att fokusera på tillförlitligheten i den enskilda bedömningen. Den hävdar att bristande överensstämmelse mellan lärarbedömningar och externa bedömningar orsakas av ofullkomligheter i lärarnas bedömningar. Med detta synsätt blir det möjligen naturligt att sätta upp gränser för ”obefogade” avvikelser gentemot kontrollrättningarna. Men eftersom huvuddelen av den bristande överensstämmelsen har sin grund i att bedömningsuppgiften är så pass komplex, försätts lärare och skolor i den omöjliga situationen att förbättra samstämmigheten i bedömningarna, när i stället andra åtgärder skulle vara en mera framkomlig väg. Ett exempel på en sådan åtgärd är att öka antalet bedömningar av varje uppsats.

Skolinspektionen föreslår också att lärarna ska få bättre utbildning i bedömning och att bedömningsanvisningarna förbättras, vilket vi håller med om. Skolinspektionen föreslår vidare att skolorna i så stor utsträckning som

möjligt använder sig av så kallad sambedömning, det vill säga att fler än en lärare bedömer samma prov eller ger sin syn på en aktuell bedömning. För att stärka likvärdigheten är det dock nödvändigt att de olika lärarna gör sina bedömningar oberoende av varandra, men så har inte Skolinspektionen formulerat förslaget.

Ytterligare ett förslag från Skolinspektionen är att elevens identitet ska vara okänd för läraren som bedömer provet. Detta förslag har sin grund i Hinnerichs och Vlachos (2013) studie av missgynnande av vissa elevgrupper som vi redogjorde för i förra avsnittet.

Skolinspektionen föreslår vidare att en utredning tillsätts för att utreda för- och nackdelar med att låta uppsatsskrivande ingå i de nationella proven. Detta mot bakgrund av att man konstaterat att de nationella proven endast i begränsad utsträckning bidrar till att stärka likvärdigheten i betygssättningen. Ett problem med förslaget är dock att det enbart lyfter fram de problem som finns i uppsatsdelarna av de nationella proven. Det finns även produktiva provdelar i andra ämnen än språk där eleverna exempelvis ska kommunicera muntligt eller genomföra experimentella uppgifter, men där det inte varit möjligt att låta dessa ingå i omrätningsverksamheten, eftersom det inte finns någon dokumentation av elevsvaren som kan utsättas för omdömning. Eftersom det finns anledning att tro att problemen med bristande överensstämmelse i dessa bedömningar är minst lika stora som för uppsatsbedömningarna framstår det som föga ändamålsenligt att begränsa frågan till att bara gälla uppsatsernas roll i det nationella provsystemet.

REFLEKTIONER KRING SKOLINSPEKTIONENS OMRÄTTNINGAR

Gustafsson och Erickson (2013) analyserade de två första omgångarna av omrättningar, och konstaterade att det finns flera problem med dem. Huvudproblemet är att det inte är möjligt att hävda att Skolinspektionens omrättningar är mer korrekta än lärarnas ursprungliga rättningar. De bedömare Skolinspektionen anlitate utgjorde inget representativt urval av bedömare, och det är rimligt att anta att omrättningsuppdraget i sig självt innebar användning av generellt strängare bedömningsnormer. Graden av överensstämmelse mellan de två bedömarna var också samma som påvisats i andra undersökningar av komplexa bedömningar, exempelvis av examensarbeten inom högskolan (Härnqvist, 1999).

Gustafsson och Erickson (2013) konstaterade också att Skolinspektionens undersökning 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 ”naming, shaming and blaming”. I inspektion utgår man från att metoderna 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.

Det finns alltså anledning att förhålla sig kritisk till Skolinspektionens genomförande av omrätningsuppdraget. Samtidigt är det uppenbart att det finns problem förknippade med de nationella provens förmåga att ge ett adekvat stöd för likvärdig betygssättning. Dessa problem beror dock förmodligen i första hand på de nationella provens syften, utformning och användning. 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.

Betygens likvärdighet

Sedan det målrelaterade betygssystemet infördes har frågan om betygens likvärdighet alltmer kommit i fokus. En betydande mängd forsknings- och utredningsarbete har ägnats åt att beskriva och analysera just betygens likvärdighet. I detta avsnitt ger vi en kort översikt över resultaten av denna forskning.

VAD ÄR LIKVÄRDIG BETYGSSÄTTNING?

Innan vi mer konkret går in på frågan om betygens likvärdighet finns det anledning att diskutera innebörden i begreppet likvärdighet i detta sammanhang. En intuitiv definition av likvärdig betygssättning är att elever eller grupper av elever som i samma grad uppnått kunskapskraven ska ha samma betyg. Innebörden är här densamma som att betygen ska vara rättvisa. I skollagen (2010 kap 1, 9 §) står: ”Utbildningen inom skolväsendet ska vara likvärdig inom varje skolform och inom fritidshemmet oavsett var i landet den anordnas”. Någon allmänt omfattad definition av begreppet ”likvärdig utbildning” finns inte, utan det kan ges flera olika innebörder. Den underliggande betydelsen av begreppet kan sägas vara att skapa jämlikhet i livschanser. Man kan urskilja en lägre ambitionsnivå som avser ”lika tillgång” till

utbildning, och en högre ambitionsnivå som avser ”lika möjligheter”. Det senare innebär att skolan ska kompensera för effekten av ovidkommande faktorer som kön, klass och etnicitet (von Greiff, 2009).

En av betygens huvudfunktioner är att vara ett instrument för urval till nästa nivå i utbildningssystemet. Eftersom betygen påverkar individers tillgång till utbildning innebär detta att även den lägre ambitionsnivån avseende likvärdig utbildning hotas av bristande likvärdighet i betyg. När man använder betygen som ett mått på utbildningsutfall hotas även den högre ambitionsnivån av orättvisa betyg. En likvärdig betygssättning är därför central för en likvärdig skola.

Denna slutsats ifrågasätts dock av Lundahl (2006, 2011) och Waldow (2011). De baserar sitt ifrågasättande på jämförande studier av hur betyg tolkas och används i Sverige och Tyskland. De menar att trots att de tyska betygen brast i likvärdighet samtidigt som de hade en mycket stor betydelse för fördelningen av livschanser över eleverna i den starkt differentierade tyska skolan var det en fördel att det inte gjordes några försök att med hjälp av prov åstadkomma en mer likvärdig betygssättning i Tyskland. Detta bevarade, enligt Lundahl och Waldow, lärarnas professionalitet. I Sverige infördes i stället det relativa betygssystemet, understött av standardprov, för att lösa problemen med den bristande likvärdigheten i betygssättningen. Detta innebar en deprofessionalisering av läraryrket, eftersom lärarperspektivet ersattes av provresultat, menar Lundahl (2006).

Denna slutsats anser vi dock är felaktig. Den stora insikten i Wigforss utredning (SOU 1942:11) var att lärare har utmärkt förmåga att bedöma sina egna elevers kunskaper utifrån bland annat observationer i klassrummet – men de kan inte bedöma andra lärares elever. För att lösa likvärdighetsproblemet måste därför lärarna förses med information om graden av duktighet hos andra lärares elever relativt de egna eleverna, samtidigt som de även fortsättningsvis utnyttjar sin professionella kompetens i bedömning av de egna eleverna. Det var så det relativa betygssystemet var tänkt att fungera, och trots all kritik som riktats mot det finns det anledning att tro att det i huvudsak också gjorde det (se Kilpatrick & Johansson, 1994, för en ingående diskussion om det relativa betygssystemet ur likvärdighets- och professionalitetssynpunkt).

ASPEKTER AV LIKVÄRDIGHET

Som vi redan noterat i beskrivningen av hur de nationella proven utvecklats, har synen på behovet av likvärdighet i betygssättningen skiftat under de senaste decennierna. En förklaring till detta är att det finns flera olika aspekter

av likvärdighet, där vissa fokuserats mer än andra.

Den likvärdighetsaspekt som i första hand varit i fokus är skillnader i betygssättning mellan lärare och skolor. Även om det är rimligt att tro att det är olika mekanismer som skapar skillnader i betygssättning mellan lärare å ena sidan och mellan skolor å andra sidan har det på grund av hur statistiken är sammanställd varit svårt att separera de två variationskällorna. Vi kommer därför att behandla dem tillsammans.

En annan uppmärksammas likvärdighetsfråga handlar om huruvida betygen är jämförbara för olika grupper av elever, som pojkar och flickor, elever med olika utbildnings- och migrationsbakgrund, eller elever i skolor med olika huvudmän. Skillnader i betygsresultat kan ofta konstateras mellan olika grupper av elever, men frågan är om detta är uttryck för att betygen är orättvisa, eller om betygen på ett korrekt sätt återspeglar faktiska skillnader i kompetens mellan grupperna.

En ytterligare likvärdighetsaspekt gäller betygens jämförbarhet över tid. Eftersom gymnasiebetyg erhållna vid olika tidpunkter jämförs med varandra vid antagning och urval till högre utbildning är det givetvis viktigt att dessa är likvärdiga. Frågan om systematiska förändringar i de faktiska kraven på kunskaper för att nå de olika betygsstegen ("betygsinflation") har studerats i viss utsträckning, och vi kommer längre fram att sammanfatta resultaten från denna forskning.

Det finns också en annan aspekt av frågan om betygens jämförbarhet över tid. Det har konstaterats att de nationella provens medelvärde i vissa ämnen varierar från ett år till ett annat, vilket också kan förväntas medföra en motsvarande variation i betygen, i den mån de nationella proven stödjer lärarnas betygssättning. Även detta är en viktig likvärdighetsfråga både när det gäller betygen och de nationella proven.

Ytterligare en likvärdighetsaspekt gäller frågan om betygens jämförbarhet över olika ämnen. Är det svårare att få höga betyg i vissa ämnen än i andra, eller är lättare att få ett underkänt betyg i vissa ämnen än i andra? Frågan har knappast alls diskuterats i Sverige. Den har däremot diskuterats livligt i andra länder, exempelvis England.

Under det relativa betygssystemet var endast några av dessa likvärdighetsfrågor aktuella att ställa, bland annat eftersom det inte fanns någon ambition att fånga upp förändring över tid. Men efter införandet av det målrelaterade betygssystemet kom frågan om betygens likvärdighet snabbt i fokus. Politiskt uppmärksammades frågan alltså under sent 1990-tal och därefter i tilltagande omfattning, både i Skolverkets egna kvalitetsgranskningar och i lokala och kommunala studier.

Riksrevisionen (2004) sammanfattade i rapporten ”Betyg med lika värde? En granskning av statens insatser” resultat från ett antal olika studier. Här görs bland annat en jämförelse mellan nationella provbetyg och slutbetyg, som visade 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. Rekommendationerna innefattade bland annat en utökning av allmänna råd, kommentarmaterial och material i den så kallade provbanken. Dessutom konstaterades att Skolverket ”inte i tillräcklig utsträckning tillvaratagit möjligheten att stödja en likvärdig betygssättning genom det nationella provsystemet” (s. 59). Hur detta stöd skulle utformas behandlades dock inte.

Motsvarande granskning av Riksrevisionen (2011) har titeln ”Lika betyg, lika kunskap? En uppföljning av statens styrning mot en likvärdig betygssättning i grundskolan”. Resultatet av granskningen formuleras på följande sätt: ”Betygssättningen sker fortfarande inte på ett likvärdigt sätt. Problemet med bristande likvärdighet i betygssättningen har varit känt under lång tid och omfattande insatser har vidtagits för att åtgärda problemet, hittills utan uppenbara framgångar” (s. 10). Riksrevisionens rekommendationer innebär att regeringen bör förtydliga för skolmyndigheterna hur långt deras mandat sträcker sig samt även se över målen för de nationella ämnesproven och renodla deras syften som betygssjäddjande. Skolverket rekommenderas att använda sitt mandat att styra samt att återkommande följa upp hur olika stödmaterial för bedömning och betygssättning används på skolnivå. Slutligen uttrycker Riksrevisionen en stark rekommendation till Skolverket och Skolinspektionen att mera effektivt samordna sina uppföljnings- och kontrollverksamheter. I rapporten finns en tydlig kritik av myndigheternas försiktighet i relation till den kommunala och lokala nivån, och Riksrevisionen konstaterar att ”delat ansvar mellan stat och huvudman komplicerar styrningen” (s. 10).

RELATIONEN PROVBETYG–SLUTBETYG/KURSBETYG

De standardiserade provens roll i betygssättningen var mycket tydlig i det relativa betygssystemet – klassens resultat på provet styrde klassens betygsfördelning relativt andra klasser. Även för den individuella betygssättningen kunde resultat på provet ha betydelse, men betyget skulle framför allt bestämmas av lärarens samlade bedömning av elevens prestationer i klassrummet. I det målrelaterade systemet finns inga anvisningar om i vilken utsträckning den enskilda elevens betyg ska styras av provbetyget, bara att

läraren ska ha ett allsidigt underlag, det vill säga inte bara resultatet på det nationella provet. Det finns givetvis inte heller några anvisningar om hur betygsfördelningen i klassen ska styras av klassens provresultat.

De första tio åren av det nuvarande styr- och betygssystemet präglades av oro för att proven tilläts styra betygssättningen i alltför stor utsträckning. Under de senaste tio åren har förhållandet varit det motsatta. Oron för bristande likvärdighet har accelererat och de nationella proven har alltmer kommit att återopas som möjlig riktpunkt. Utvecklingen har varit successiv, men något av en vändpunkt var Skolverkets (2007) analys av sambandet mellan nationella prov och slutbetyg i grundskolan, liksom motsvarande rapport för gymnasieskolan (Skolverket, 2009). Båda rapporterna visar på stor variation, både inom och mellan skolor, i hur provresultaten/provbetyget påverkar slutbetyget/kursbetyget. Inga påtagliga skillnader noteras mellan kommunala och fristående skolor. Stora skillnader finns dock mellan ämnen. I engelska är överensstämmelsen provbetyget–slutbetyget/kursbetyget störst, minst är den i matematik, och under senare år de naturorienterade ämnena i årskurs 9.

Frågan om betygens och provens stabilitet från ett år till ett annat är också av stort intresse. Av olika redovisningar från Skolverket framgår att i synnerhet provbetygets medelvärde uppvisar stora variationer i vissa ämnen, och särskilt i matematik. 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.

Den variation i provbetygets medeltal från ett år till ett annat som är så tydlig för vissa ämnen, inte minst matematik kan ha flera olika förklaringar. En skulle kunna vara att provets svårighetsgrad varierar mellan åren, och att det därför är olika svårt att nå ett visst betyg. Svårighetsgraden för proven tycks dock inte variera i så hög grad att detta framstår som en rimlig förklaring. En bättre förklaring är förmodligen att poänggränserna för de olika betygsstegen fastställts med bristande precision, vilket i synnerhet drabbar prov med ett lågt totalt poängantal.

Även en synbar stabilitet kan inrymma förändringar som inte är helt reella. Detta skulle till exempel kunna vara fallet i engelska, där resultaten över tid är tämligen stabila, dock med en successiv, svag ökning av medelvärdet för både provbetyg och slutbetyg mellan åren 1998 och 2012. En intressant fråga är om denna ökning återspeglar en faktisk förbättring av elevernas engelskkunskaper eller om den helt eller delvis har sin grund i fastställandet av poänggränser. Inom ramen för ett pågående projekt vid Göteborgs

universitet pekar preliminära resultat på det senare, i synnerhet för de högre betygsstegen.

Poänggränserna för de olika betygsstegen fastställs innan provet ges. De fastställs utifrån bedömningar, främst gjorda av lärare, av var poänggränserna ska gå för att svara mot de kunskapskrav för olika betygssteg som anges i styrdokumentet (s.k. *standard setting*). Dessa bedömningar kompletteras med data från utprövningar. Betygsgränserna fastställs inför varje nytt prov enligt dessa principer. Det är dock i dessa bedömningar inte möjligt att vara så precis att inte smärre förskjutningar av gränsvärdena sker från ett år till ett annat. Vi återkommer till dessa problem längre fram i rapporten.

Tyvärr begränsas kraften i analyserna av att det inte har varit möjligt att analysera de prov som gavs under åren 2006–2009. Som en följd av den nedtonade mätkaraktären hos de nationella proven har Skolverket periodvis haft en mycket skeptisk inställning till provsekretess. På 1990-talet togs ordet sekretess tillfälligt togs bort och ersattes med uttryck som ”lärares professionella handhavande”, ”endast användas för sitt syfte” och ”hanteras och förvaras på betryggande sätt”. Detta skapade en irriterande otydlighet kring regelverket för materialet. Men en senare åtgärd kom att ha betydligt allvarligare effekter. Skolverket fattade nämligen beslut att släppa sekretessen för ämnesproven i årskurs 9 för fyra år i följd (2006–2009), vilket omöjliggjorde användning av så kallade länkuppgifter (se ordlista) för länkning mellan proven under dessa år. Detta har bland annat gjort det omöjligt att ge ett fullödigt svar på frågan om hur kunskapsnivåerna i engelska har förändrats mellan 2001 och 2011.

BETYGSINFLATION

Gustafsson och Yang Hansen (2009) har beskrivit hur resultaten i grundskolan utvecklats enligt betygen. De visar att det skett positiva förändringar enligt det totala betygsmättet (meritvärdet) från tidpunkten för införandet av det målrelaterade betygssystemet, men även att utvecklingen varit mycket positiv i de praktisk-estetiska ämnena, och närmast obefintlig i svenska, matematik och engelska.

Den positiva betygsutvecklingen kan tolkas som att elevernas kunskaper och färdigheter har förbättrats över tid. Det finns dock starka skäl att ifrågasätta om betygsförändringarna faktiskt återspeglar kunskapsutvecklingen. Skillnaderna i utveckling för de olika ämnena är orimligt stora och dessutom oförklarliga. Klapp Lekholm (2008) har visat att betygssättning är influerad av olika faktorer utöver elevernas kunskaper och färdigheter. Inte ens i de ämnen där stöd för betygssättning ges av nationella prov finns möjlighet att

säkerställa att betyg och provresultat är jämförbara över tid. En rimligare förklaring till förändringarna över tid är varierande grader av betygsinflation i olika ämnen, möjligen i kombination med faktiska förändringar av kunskaper och färdigheter, vilka kan vara såväl positiva som negativa.

Cliffordson (2010) visade att det genomsnittliga jämförelsetalet i gymnasieskolan successivt har ökat med cirka 0,5 standardavvikelseenheter från 1997 fram till och med 2004, för att därefter ligga på ungefär samma nivå. Enligt Skolverkets statistik är nivån densamma även de nästkommande tre åren, till och med 2012. Studien visar också att ökningen är något större för flickor än för pojkar. Betygen stiger mera för fristående skolor under de första tre till fyra åren för att sedan sjunka och närma sig betygsnivån för kommunala skolor, vilket sannolikt delvis kan förklaras med en förändring av friskolornas elevsammansättning (Vlachos, 2010). Medan differensen i betygsnivå mellan kommunala och fristående skolor minskar över tid, blir nivån något lägre för de yrkesinriktade programmen på fristående skolor i slutet av perioden, medan nivån för de studieförberedande programmen fortfarande är något högre för fristående skolor jämfört med kommunala skolor. Med kontroll för kön samt utbildnings- och migrationsbakgrund minskar skillnaderna i såväl betygsutveckling som betygsnivå mellan kommunala och fristående skolor. När hänsyn också tas till elevernas grundskolebetyg minskar också skillnaderna väsentligt. Nivåskillnaden försvinner då helt i slutet av perioden.

Fördelningen av jämförelsetalet inom respektive avgångskohort har förändrats över tid. Från att ha varit i stort sett normalfördelad 1997 har den blivit alltmer snedfördelad, med en ökning av andelen med höga betyg. För de högsta betygen (jämförelsetal 20,0) har det skett en kontinuerlig ökning över de första 11 åren för att minska något år 2009. Det är dock stora skillnader mellan kommunala och fristående skolor. Medan ökningen för de kommunala skolorna är i stort sett konstant över hela perioden (från 0,02 till 0,72 procent) med undantag för år 2009 då andelen sjönk något, var ökningen betydligt större för fristående skolor fram till och med 2004 (från 0,2 till 1,8 procent), för att sedan avta årligen, bortsett för avgångsåret 2009, då andelen åter ökade något. Skillnaden är dock fortfarande påtaglig år 2009.

Jämförelser av betygens genomsnittliga jämförelsetal eller andelen högsta betyg över tid är ett osäkert mått på om det föreligger någon betygsinflation eller inte. Det har således varit oklart om den kontinuerliga medelvärdesökningen av betygen återspeglar en ökning av elevernas kunskaper och färdigheter, eller om den är uttryck för betygsinflation. Om betygsökningen återspeglar en faktisk ökning av gymnasieelevernas kunskaper och

färdigheter är det rimligt att förvänta att de bättre gymnasiebetygen också återspeglas i bättre studieresultat i den fortsatta utbildningen. En prognosstudie (Cliffordson, 2004) baserad på betyg från åren 1997 till 2001, då det genomsnittliga jämförelsetalet hade ökat från 12,7, till 13,7, visar att ökningen av jämförelsetalet inte svarar mot en höjning av gymnasieelevernas kunskaper och färdigheter. Detta resultat pekar på att ökningen av jämförelsetalet orsakas av betygsinflation.

Diskussion och slutsatser

Nedan diskuterar vi dels likvärdighetsproblemen i betygen och de nationella proven, dels möjliga sätt att åstadkomma ett stöd från de nationella proven för en likvärdig betygssättning.

LIKVÄRDIGHET I BETYG OCH NATIONELLA PROV

Den översikt som presenterats ovan pekar på en lång rad problem med likvärdigheten i betygen. De målrelaterade betygen lider av bristande likvärdighet mellan lärare och skolor, och i synnerhet gymnasiebetygen är påverkade av inflation. Det är därför inte möjligt att använda betygen för att studera kunskapsutveckling över tid. Och att använda betygen för urval till gymnasium och högre utbildning kan vara problematiskt ur både rättvis- och effektivitetssynpunkt.

Svårigheterna att åstadkomma likvärdighet i betygssättningen har sannolikt sin grund i att de verbalt formulerade kunskapskraven inte ger tillräcklig styrning för en jämförbar betygssättning. Även om lärarna inom alla landets skolor utvecklar en samsyn för hur kunskapskraven ska tolkas och användas kvarstår problemet att skillnaderna mellan skolor kommer att vara minst lika stor som tidigare. Någon informationskälla som anger den mer konkreta innebörden i kunskapskraven är sannolikt nödvändig för att förbättra likvärdigheten i betygssättningen.

Praktiskt taget alla studier som redovisats här avser betygssättning enligt 1994 års läroplaner. Frågan om i vilken grad dessa resultat är generaliserbara till 2011 års läroplaner är därför högst relevant. Av de utdrag ur läro- och kursplaner som vi redogjort för i detta kapitel framgår att kunskapskraven är förhållandevis abstrakt formulerade, och skillnader mellan kunskapskraven för de olika betygsstegen anges som ”bra”, ”bättre” eller ”bäst”. En klar innebörd i kunskapskraven för den lägsta nivån är därför svår att utläsa ur texten, liksom skillnaden i kunskapskrav för de olika betygsstegen. Kon-

kretiseringar av kunskapskraven i form av provuppgifter med stigande svårighetsgrad åtföljda av exempel på bedömda och kommenterade elevsvar framstår åtminstone i princip som en möjlig väg att identifiera en mer tydlig innebörd i kunskapskraven. Detta görs till viss del i befintliga kommentarmaterial och nationella prov men behöver förstärkas.

Vi har också identifierat likvärdighetsproblem i de nationella proven. För prov där eleverna producerar längre texter som ska bedömas är graden av överensstämmelse mellan lärarnas bedömningar och externa bedömningar ofta inte tillräckligt hög, och i de flesta ämnen är lärarnas bedömningar generösare än de externa personernas bedömningar. Ett annat problem med de nationella proven är att det i de flesta ämnen finns en stor variation i medelbetyg från ett år till ett annat. Denna variation kommer delvis till uttryck i variation i elevernas slutbetyg/kursbetyg från ett år till ett annat, men betygen är mer stabila än provbetygen. Vidare är andelen elever som får ett icke godkänt provbetyg betydligt högre i exempelvis matematik än andelen som får ett icke godkänt slutbetyg eller kursbetyg. Detta indikerar att de nationella proven i åtminstone vissa ämnen ger ett begränsat stöd till lärarnas betygssättning. Trots detta förklarar sig lärare i årliga undersökningar oftast mycket positiva till proven, inte minst på grund av deras betygsstödjande funktion. Med tanke på den höga grad av instabilitet som de nationella proven i vissa ämnen uppvisar, både i elevernas resultat och i medelresultaten från ett år till ett annat, är det kanske lämpligt att de inte har mer än en begränsad påverkan på betygen.

Det är dock naturligt att fråga sig varför de nationella proven uppvisar en så hög grad av instabilitet, och om det går att göra någonting åt det. Skolverket har gjort en del för att undersöka dessa frågor. Det huvudsakliga svaret på varför provresultaten inte är stabila från ett år till ett annat är att poänggränserna för de olika betygsstegen inte fastställts med tillräcklig precision (Skolverket, 2013).

Som vi visat ovan är något sjunkande gränser för betygsstegen VG och MVG mellan 2001 och 2006 en trolig förklaring till varför provbetygets medelvärde ökat svagt i engelska i årskurs 9 mellan dessa år. För andra ämnen, i synnerhet matematik, är dock instabiliteten över år ett betydligt större problem än det är för engelska. Detta har huvudsakligen sin grund i att antalet provpoäng är betydligt mindre i dessa ämnen, vilket medför att någon enstaka poängs förskjutning av betygsgränserna har ett stort genomslag i andelen elever som hamnar på de olika betygsstegen. Det begränsade antalet provpoäng har sin grund i att ett begränsat antal stora autentiska uppgifter används, vilka inte ger grund för mer än ett begränsat antal po-

äng. Vidare kräver dessa uppgifter komplexa bedömningar, vilket bidrar till försämrad mätsäkerhet. I de prov som utvecklats på senare tid har dock antalet provpoäng ökats, vilket förbättrar möjligheterna att öka precisionen i fastställandet av betygsgränserna.

De nationella proven har två huvudsakliga syften: (1) att stödja en likvärdig och rättvis bedömning och betygssättning; (2) att ge underlag för en analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, på huvudmannanivå och på nationell nivå. Våra resultat pekar på att de nationella proven i sin nuvarande utformning inte fullt ut förmår att uppfylla något av dessa syften. Det har sin grund i att proven har reliabilitetsbrister och att de inte är jämförbara över tid. Vi diskuterar nedan hur dessa två problem kan lösas.

41

HUR KAN DE NATIONELLA PROVEN GÖRAS MER RELIABLA?

Problemet med reliabilitetsbrister i bedömning av längre uppsatser har uppmärksammats av Skolinspektionen vid alla genomförda omrätningsomgångar. Skolinspektionens huvudförslag har varit att uppsatsskrivning ska avlägsnas från de nationella proven. Problemet med denna lösning är att förmåga att skriva på svenska och engelska är centrala mål i läroplanerna för både grundskola och gymnasium. Det finns inte heller någon annan känd metod för att pröva förmågan att skriva uppsats än just genom uppsatsskrivning. Om uppsatsdelarna tas bort från de nationella proven försämrans därför provens validitet, vilket inte är acceptabelt.

Det är inte heller omöjligt att höja reliabiliteten i uppsatsbedömningarna, vilket bland annat visats när man studerat graden av samstämmighet mellan oberoende bedömare⁴ vid bedömning av uppsatser i engelska (Skolverket, 2009). Skolinspektionen föreslår flera olika metoder att åstadkomma förbättringar i överensstämmelsen mellan bedömarna, som förbättrade bedömningsanvisningar, förbättrad utbildning i bedömning och en ökad grad av samverkan mellan lärare kring bedömningsfrågor. Skolinspektionens förslag att sätta upp en tämligen godtycklig gräns för vad som ska betraktas som ”obefogade” avvikelser mellan interna och externa bedömningar framstår däremot som mindre lämpligt. Huvudskälet till detta är att det finns ett tak för hur hög överensstämmelsen kan och bör vara i komplexa bedömningar, eftersom skillnader mellan bedömare oftast återspeglar olika men legitima synsätt på vad som är kvalitet. Införande av en mer eller mindre godtycklig gräns mellan vad Skolinspektionen tycker är befogade

4 Så kallad interbedömarreliabilitet.

och obefogade avvikelser innebär därför att skolorna får ett uppdrag som i praktiken är omöjligt att genomföra med framgång.

Den främsta, och i vissa fall den enda tillgängliga metoden, att öka tillförlitligheten i komplexa bedömningar är att öka antalet oberoende bedömare. Även i de fall den enskilda bedömningen har låg reliabilitet ger ytterligare en eller två bedömningar grund för en sammanfattande bedömning med god reliabilitet. Eftersom det är viktigt att de olika bedömningarna är oberoende och att de olika bedömarna är representativa för alla landets skolor bör bedömarna inte komma från samma skola.

Ett system med mer än en bedömning av varje uppsats är givetvis mer resurskrävande än ett system med endast en bedömning. Det skulle dock ge fördelar utöver att reliabiliteten i bedömningen av de enskilda elevernas uppsatser skulle öka, framför allt genom att landets lärare skulle få direkt erfarenhet av hur variationen i elevernas förmåga gestaltar sig på nationell nivå. Erfarenheter från flera olika bedömares arbete skulle också sannolikt bidra till en successiv uppbyggnad av lärarnas bedömningskompetens.

Enligt Skolinspektionens resultat framstår reliabilitetsbrister i bedömningar av uppsatser som huvudproblemet. Det måste dock betonas att det finns andra provdelar som erbjuder minst lika stora utmaningar för bedömare. Hit hör exempelvis bedömningar av muntlig förmåga, gruppuppgifter och experimentella uppgifter. Genomförandet av sådana prov dokumenteras inte och har därför inte kunnat ingå i Skolinspektionens omrättningar, men om så hade skett hade säkerligen reliabilitetsbrister kunnat konstateras även här. Den lämpliga lösningen hade dock inte varit att utesluta dessa delprov från de nationella proven, eftersom de återspeglar väsentliga delar av läroplanen. Om de nationella proven i alltför liten grad söker fånga upp de väsentliga målen i läroplanen förlorar de det som är ett provs mest väsentliga egenskap, nämligen validiteten. Det måste dock understrykas att det också finns andra sätt på vilka prov kan tappa i validitet, och ett av dessa är att ha för låg reliabilitet, eftersom reliabilitet är en nödvändig förutsättning för validitet.

Ett annat sätt på vilket prov kan förlora i validitet är att provresultaten systematiskt påverkas av faktorer som egentligen inte hör till det vi vill mäta. Ett exempel kan vara att reflektioner och resonemang kring abstrakta och komplexa fenomen, till exempel i samhälls- eller naturorienterande ämnen, ska uttryckas i skrift, ett annat att ett prov med fokus på hörförståelse skulle kräva att eleverna avgav svar i form av längre texter. Eftersom de nationella proven för alla åldersgrupper i många fall ställer höga krav på eleverna att producera text utgör detta ett potentiellt hot mot validiteten i de prov som

inte har som syfte att mäta elevernas skrivförmåga. Det är angeläget att undersöka hur stort detta problem är, och om det i så fall bör utvecklas prov med alternativa svarssätt. Prov där eleverna avkrävs längre svar kräver också bedömning av svaren, vilket ger upphov till reliabilitetsbrister.

Vi vill sålunda inte förespråka en lösning på reliabilitetsproblemen med de nationella proven som innebär att de komplexa delproven avlägsnas, eftersom detta skulle innebära hot mot provens validitet. Problemen med samstämmighet i bedömningarna av elevproducerade svar bör dock ges ökad uppmärksamhet, och system som bygger på flera oberoende bedömningar bör prövas. En noggrann genomgång bör också göras av de nationella proven för att identifiera sådana hot mot reliabilitet som också utgör hot mot provens validitet, och då framför allt krav på omfattande skriftliga svar när denna förmåga inte är i fokus.

Även om dessa åtgärder kommer att förbättra provens reliabilitet och validitet finns det inte anledning att förvänta sig att detta kommer att leda till en tillräckligt hög reliabilitet för att proven ska kunna ge ett helt adekvat stöd för likvärdig betygssättning, eftersom de grundläggande problemen med prov byggda på autentiska uppgifter kvarstår. Nusche m.fl. (2011, s. 55) föreslår att en lösning kan vara att kombinera autentiska prov med standardiserade prov som inte ställer stora krav på bedömning av elevernas svar. En sådan kombination menar de skulle kunna förena den högre validiteten i autentiska prov med den högre graden av reliabilitet och generaliserbarhet i standardiserade prov.

Enligt vår mening framstår en sådan ansats som väl skickad att bevara de positiva aspekterna av det svenska nationella provsystemet, samtidigt som den erbjuder möjligheter att komma till rätta med de reliabilitetsproblem som finns i systemet. Vi återkommer i kapitel 5 till hur detta kan åstadkommas.

HUR KAN DE NATIONELLA PROVEN GÖRAS STABILA ÖVER TID?

I det svenska provsystemet har det aldrig funnits en ambition att proven ska ge resultat som är jämförbara över tid, även om Skolverket vid flera tillfällen initierat projekt som haft som syfte att hitta lösningar för att åstadkomma detta. Några egentliga framsteg har dock inte gjorts, vilket möjligen beror på att detta inte har varit en högprioriterad fråga. En anledning till detta kan vara att frågan om jämförbarhet över tid framför allt har kopplats till frågan om det är möjligt att använda proven för att studera utveckling av kunskapsnivåer på nationell nivå över tid, mot vilket det finns starka argument (se kapitel 3). Jämförbarhet över tid är dock en mycket önskvärd egen-

skap hos proven, om en av deras uppgifter är att motverka betygsinflation. Skolverket betraktade dock länge frågan om betygsinflation som mindre intressant, och hävdade fram till 2012 att det inte existerar någon betygsinflation i Sverige (Skolverket, 2012).

Även om omfattningen av betygsinflationen är svår att bestämma mer exakt råder det inget tvivel om att det är angeläget att de nationella proven ger resultat som är jämförbara över tid för att proven ska kunna stödja en likvärdig betygssättning. Vi återkommer i nästa kapitel till frågan om hur stabiliteten i de nationella proven kan säkras.

3. Kunskapsbedömning på systemnivå

Med systemnivå avser vi alla nivåer som finns ovanför den individuella nivån, som klass, skola, kommun/huvudman och nation. Tyngdpunkten här ligger dock på den nationella nivån. OECD (2013) identifierar sex olika syften med bedömning av kunskaper på nationell nivå: (1) att vid en viss given tidpunkt beskriva resultatnivåer, och då inte minst för olika grupper av elever med uppdelning på exempelvis region, kön eller föräldrarnas utbildningsnivå; (2) att beskriva förändringar i resultatnivåer över tid; (3) att följa upp och utvärdera effekten av policyförändringar eller interventioner; (4) att samla in demografiska, administrativa och kontextuella data som kan vara användbara för att förklara de observerade utfallen; (5) att skapa och återföra information som är relevant för olika aktörer inom utbildningssystemet; och (6) att använda den insamlade informationen för analys, utveckling och implementering av utbildningspolitik. Dessa olika syften kan dock nås på olika sätt och nedan redogör vi för några internationella exempel.

Kunskapsbedömningssystem – en internationell utblick

Särskilda nationella system för kunskapsbedömning har utvecklats i flera länder, exempelvis Belgien, England, Finland, Holland, Nya Zeeland och USA (Gustafsson, 2006a). Enligt OECD:s (2013) översikt har praktiskt taget samtliga OECD-länder implementerat nationella kunskapsbedömningssystem, vilka kan vara antingen stickprovsbaserade eller populationsbaserade.

USA

I slutet av 1960-talet infördes i USA under ledning av Ralph Tyler en modell för nationell kunskapsbedömning som var innovativ på många sätt, och inte minst metodologiskt. Denna modell har sedan genomgått flera föränd-

ringar, och den har varit en förebild för andra länder som har utvecklat egna system. Därför är det lämpligt att ta Tylers modell som en av utgångspunkterna för diskussioner om kunskapsbedömning på nationell nivå. Nedan beskriver vi kortfattat modellen. För en mer utförlig beskrivning hänvisas till Gustafsson (2006a).

Medan de standardiserade testsystem som var i bruk under 1960-talet fokuserade på att ge säkra resultat för enskilda elever var huvudsyftet med ”National Assessment of Educational Progress” (NAEP) att kunna bedöma nivån på kunskaper och färdigheter på nationell nivå. Eftersom de enskilda elevresultaten i sig inte var intressanta kunde urval göras både av elever och av uppgifter genom så kallad matrissampling (se ordlista), vilket innebar att olika elever besvarade delvis olika uppgifter. Härigenom kunde många uppgifter användas utan att tids- och arbetsbelastningen för den enskilda eleven blev alltför stor. Det faktum att elevurvalen var förhållandevis små gjorde det också möjligt att inte endast använda flervalstuppgifter, utan även uppgifter med elevproducerade svar som krävde bedömning kunde nyttjas.

Erfarenheterna av NAEP under 1970-talet visade att programmet kunde bidra med nationellt och regionalt representativa data om det amerikanska utbildningssystemet. Men samtidigt hade NAEP en undanskymd plats i den utbildningspolitiska diskussionen, och endast i liten grad hade resultaten policymässiga implikationer. En av anledningarna till detta var att NAEPs sätt att rapportera resultaten, där huvudfokus låg på enskilda uppgifter, var mer lämpat för ämnesexperter än för andra intressenter. Rapporteringen gav inte en sammanfattande överblick över resultaten, och det var svårt att göra jämförelser över tid. Det var också svårt att få överblick över skillnader i resultat för olika undergrupper av elever. Uppläggningsen av NAEP gjorde det även svårt att knyta utbildningsresultat till utbildningspolitik och skolpraktik.

Under 1980-talets början kom därför NAEP att få en delvis ny design som baserades på den utveckling inom den pedagogiska mätläran som brukar betecknas som den moderna mätteorin. Den beskrivs utförligare i rutan *Modern mätteori* (s. 48).

Den moderna mätteorin vidareutvecklades av forskare vid Educational Testing Service i USA, och speciella modellvarianter utvecklades för användning inom NAEP. Med hjälp av modellerna kunde man dra nytta av matrissamplings fördelar, samtidigt som nackdelarna förknippade med redovisning på uppgiftsnivå eliminerades. De nyutvecklade metoderna visade sig också vara mycket användbara för att skapa överblickbara resultat, och de förbättrade möjligheterna att undersöka förändringar i nivån

av kunskaper och färdigheter över tid. Dessa fördelar hade dock ett pris i form av en ökad komplexitetsgrad i genomförande av utvärderingarna och i analysen av data.

En stor mängd utvärderingar har genomförts av NAEP. Dessa är överlag mycket positiva, även om en återkommande kritik är att NAEP vuxit till ett alltför komplext system, och att det ålagts alltför många uppgifter. Andra kritiker har pekat på att utvecklingen av NAEP inneburit att många av Tylers ursprungliga idéer om en nationell utvärdering av skolan har förfusksats genom en insnävning av uppgiftstyper, stark fokusering på vissa ämnesområden, abstrakt rapportering och en ökad betoning av ansvarsutkrävande funktioner (Gustafsson, 2006a).

NYA ZEELAND

Det är naturligt att som nästa land ta upp Nya Zeeland. År 1995 startade nämligen där ett nationellt kunskapsbedömningssystem med beteckningen "National Education Monitoring Project" (NEMP), som på ett tydligt sätt var inspirerat av Tylers ursprungliga idéer om utformningen av NAEP. Ett syfte var att NEMP skulle beskriva förändringar i kunskaper och färdigheter över tid. Ett annat syfte var att det skulle ge information som kunde användas både i praktisk verksamhet och för policyformulering. Analys och redovisning av resultaten gjordes på uppgiftsnivå. Efter varje tillfälle publicerades ungefär två tredjedelar av uppgifterna, medan återstoden återanvändes i nästa utvärderingscykel för att studera förändring över tid. Rapportering gjordes också för olika undergrupper av elever. Ett utmärkande drag hos NEMP var att projektet i stor utsträckning involverade yrkesverksamma lärare, både för att man ville utnyttja lärarnas kompetens, och för att lärarnas delaktighet sågs som en förutsättning för att NEMP skulle kunna förbättra undervisningen.

NEMP avslutades 2000, men efterträddes samma år av projektet "National Monitoring Study of Student Achievement" (NMSSA). NMSSA liknar i mycket NEMP, men skiljer sig från det i framför allt tre avseenden. En första skillnad är att NMSSA innehåller fler traditionella uppgifter av papper-och-penna-typ som administreras i grupp, och som genomförs av alla elever i urvalet, och där syftet framför allt är att ge en stabil grund för att undersöka förändringar över tid inom ett område. Denna typ av uppgifter används för områden som matematik, läsförståelse och naturvetenskap, medan man inom andra områden använder autentiska uppgifter, och intervju- och observationsmetoder för datainsamling. Den andra skillnaden är att NMSSA i större utsträckning fokuserar på att förklara de observerade

Modern mätteori

En av dem som lade grunden till den moderna mätteorin (eller *Item Response Theory*, IRT) var den danske statistikern Georg Rasch (1960). Rasch visade att det är möjligt att bestämma provuppgifters svårighetsgrad oberoende av nivån på kunskaper hos urvalet av elever. Från detta följer sedan flera betydelsefulla implikationer, som att det är möjligt att föra över resultat från prov som består av olika uppgifter på en och samma mätskala. Detta gör det i sin tur möjligt att jämföra resultat för elever som besvarat olika uppgifter. En annan stor fördel är att det är möjligt att genomföra undersökningarna med en hög ambitionsnivå vad gäller representation av ämnesområdena och samtidigt utöva en begränsad belastning på de elever som ingår i urvalen. Med denna teknik är det också möjligt att undersöka förändringar i kunskapsnivåer inom länder. Detta bygger på att en del av de uppgifter som används vid ett visst tillfälle bevaras hemliga, och återanvänds vid ett senare tillfälle. Dessa uppgifter används då som så kallade länkuppgifter för att föra över resultaten på en och samma skala.

En annan mycket viktig fördel med tekniken i den moderna mätteorin är att metoder också utvecklats för att göra en tydligare bestämning av innebörden i själva skalan. Grundidén är att man för ett antal olika nivåer på skalan bestämmer vilka uppgifter eleverna har en viss lägsta sannolikhet (exempelvis 67 eller 80 procent) att besvara korrekt. Analys av egenskaperna hos uppgifterna på de olika nivåerna ger sedan relativt konkret information om vad elever på de olika nivåerna vet och kan göra. Sådana så kallade kompetensskalor (*proficiency scales*) gör att de kvantitativa resultaten blir lättare att tolka och använda (se t.ex. Wilson, 2005; Wilson & Draney, 2013). De kan också göra det möjligt att mer konkret knyta resultat på skalorna till exempelvis de mål som är formulerade för en utbildning.

Resultat från nationella och internationella kunskapsbedömningar redovisas i allmänhet både som poäng och i kompetensnivåer. I exempelvis PISA 2012 användes i matematikdelen totalt 46 uppgifter med sammanlagt 84 deluppgifter. Med hjälp av modern mätteori-teknik har elevernas svar på dessa uppgifter transformerats till poäng dels på en total matematikskala, dels på sju olika delskalor. Dessa skalor har internationellt ett medelvärde om cirka 500 och standardavvikelsen cirka 100. Varje skala har också delats in i sex olika kompetensnivåer, som beskriver elevernas förmågor. Elever på nivå 1 (minst 358 poäng men mindre än 420 poäng) på den totala matematikskalan beskrivs exempelvis på följande sätt: "På nivå 1 kan eleverna besvara frågor som innehåller kända sammanhang där

all relevant information är given, och frågorna är klara och tydliga. De klarar att identifiera information och att utföra rutinprocedurer enligt direkta instruktioner. I tydligt angivna situationer kan de utföra handlingar som är självklara och omedelbart följer ur uppgiften.” Elever på nivå 4 (minst 545 poäng men mindre än 607 poäng), för att ta ett annat exempel, kännetecknas av följande förmågor: ”På nivå 4 kan eleverna effektivt arbeta med tydliga modeller för komplexa konkreta situationer, som kan innehålla begränsningar eller krav att antaganden måste göras. De kan välja och sammanlänka olika representationsformer och koppla dessa till verkliga sammanhang. Elever på denna nivå kan använda en begränsad uppsättning färdigheter och resonerar med viss insikt, i okomplicerade sammanhang. De kan konstruera och kommunicera förklaringar och argument baserade på sina tolkningar, argument och handlingar” (se Skolverket, 2013, för en detaljerad redovisning).

förändringarna. Information samlas därför in från rektor och lärare om olika kontextuella faktorer på skol- och klassrumsnivå, om samverkan mellan hem och skola, liksom om elevernas bakgrund, familj och intressen. En tredje skillnad är att rapporteringen av resultat i NMSSA i huvudsak sker i form av sådana kompetensskalor byggda med modern mätteori-teknik. I viss utsträckning rapporteras dock också resultat på enskilda uppgifter.

Förändringen av det Tyler-trogna NEMP-systemet till NMSSA år 2010 liknar således i flera avseenden den förändring som Tylers ursprungliga NAEP-system genomgick i USA år 1980, med rapportering i form av skalor, ökat fokus på förklaring och en ökad andel papper-och-penna-uppgifter. Dock ska understrykas att NMSSA fortfarande i stor utsträckning använder sig av olika typer av autentiska uppgifter, observationer och intervjuer vid sidan av papper-och-penna-uppgifter.

Urvalet omfattar varje år ungefär 5 000 elever i 200 slumpmässigt utvalda skolor. På varje skola deltar upp till 28 elever, och 8 av dessa deltar även i en fördjupningsstudie. Under 2012 undersöktes naturvetenskap och skrivning, 2013 låg fokus på matematik och på hälsa och idrott. Utvärderingarna genomförs av erfarna lärare. Ungefär 28 lärare utbildas för uppgiften, och tillbringar under en period om åtta veckor två och en halv dag i var och en av de utvalda skolorna. Eleverna arbetar med tre olika typer av uppgifter. Den första typen, som genomförs av samtliga elever i urvalet, omfattar papper-och-penna-uppgifter liksom frågeformulär om skolan och skolarbetet.

Den andra typen utgörs av mer omfattande autentiska uppgifter, som både genomförs individuellt och i grupper om fyra elever. Den tredje typen är intervjuer, som spelas in på video för att möjliggöra en mer detaljerad analys i efterhand. I de två senare typerna ingår endast de 8 elever per skola som valts ut för de fördjupande studierna.

ENGLAND

50

De ovan beskrivna systemen för nationell kunskapsbedömning är stickprovsbaserade och de är specialutformade för huvudgiften att ge information om utbildningssystemet på nationell nivå. England har ett nationellt kunskapsbedömningssystem ("National Curriculum Assessment") som i många avseenden skiljer sig från dessa system. Basen för det engelska systemet utgörs av en omfattande nationell läroplan med preciserade målbeskrivningar för olika ämnen i olika åldrar eller *stages* (7, 11 och 14 år). Graden av måluppfyllelse mäts dels med hjälp av prov, dels med hjälp av lärarbedömningar. De senare används inom ämnen och ämnesdelar där inte prov ger adekvata underlag för att bedöma i vilken utsträckning de olika målen har nåtts, exempelvis praktisk-estetiska ämnen och muntlig produktion. Enligt de ursprungliga planerna skulle enbart lärarbedömningar användas, och prov skulle endast användas för att kalibrera lärarnas bedömningar, på ett likartat sätt som standardproven användes i det relativa betygssystemet i Sverige för att kalibrera lärarnas betygssättning (DES, 1988). Men redan vid implementeringen av systemet under början av 1990-talet introducerades prov i engelska, matematik och naturvetenskap, och proven kom successivt att få allt större betydelse (Black, 1998). Systemet är fortfarande baserat på lärarbedömningar i andra ämnen än engelska, matematik och naturvetenskap. I historia, geografi, musik och teknik rapporteras lärarbedömningar till föräldrarna för 14-åringar. Prov i engelska, matematik och naturvetenskap ges till 11-åringar och 14-åringar.

Enligt de ursprungliga planerna skulle National Curriculum Assessment uppfylla flera olika syften. De skulle vara diagnostiska och visa på elevers starka och svaga sidor, formativa för att stödja undervisningen, summativa för att ge pålitlig information om enskilda elevers resultat, och utvärderande för att ge information om kunskapsutvecklingen på nationell nivå, liksom för att ge information om kvalitetsskillnader mellan skolor. Stobart (2001) pekar dock på att vissa av dessa funktioner inte uppfylls särskilt väl. Rapporteringskategorierna är alltför grova för att den diagnostiska informationen ska vara av intresse, och den eventuella formativa funktion som skulle kunna nås genom återföring av provresultat till elever och lärare utnyttjas endast i

blygsam utsträckning. Stobart (2001) påpekar att i stället ansvarsutkrävande funktioner, som redovisning av resultat på skolnivå och skoldistriktnivå kommit att dominera. Han pekar på att National Curriculum Assessment praktiskt taget helt kommit att förknippas med redovisning av provresultat, medan lärarbedömningarna kommit i bakgrunden. Tymms (2004) redovisar också forskningsresultat som visar att den tydliga trend till resultatförbättringar på nationell nivå som påvisats kan ha sin grund i *teaching to the test*, eftersom någon motsvarande trend inte går att se i stickprovsundersökningar av kunskapsutvecklingen.

Samtliga elever i de aktuella åldersgrupperna ska delta i National Curriculum Assessment. De uppnådda resultaten på prov och lärarbedömningar har dock inga konsekvenser för den enskilda eleven, vilket innebär att National Curriculum Assessment inte är *high-stakes* på elevnivå. Däremot är systemet *high-stakes* på skolnivå, för skoldistrikt och på nationell nivå (Stobart, 2001, s. 31).

Det engelska nationella kunskapsbedömningssystemet framstår som föga framgångsrikt. När det gäller huvuduppgiften – att beskriva resultatförändringar över tid – hotas korrektheten i resultaten av framför allt två omständigheter. Den ena är vad som ovan beskrevs som *teaching to the test*, att lärarna på grund av trycket att uppnå goda resultat för skolan snävar in undervisningen till att fokusera på sådant som kan förväntas komma på provet. Den andra är svårigheten att upprätthålla en konstant svårighetsgrad i proven från ett år till ett annat. Detta har sin grund i att samtliga elever deltar, vilket kraftigt försvårar möjligheterna att genom förprovningar utveckla prov som är jämförbara från ett år till ett annat. Enligt Stobart (2001, s. 33) är ett "*low-stakes*" stickprovsbaserat system som gör det möjligt att återanvända uppgifter att föredra när det gäller att studera kunskapsutveckling på nationell nivå.

National Curriculum Assessment har genomgått flera successiva förändringar över tid, och bland annat ges numera proven i naturvetenskap endast till ett stickprov av elever. I anslutning till en läroplansreform 2014 kommer också stora förändringar att göras. Systemet med detaljerade beskrivningar av nivåer kommer att överges i syfte att ge skolorna större frihet att utforma undervisningen. National Curriculum Tests kommer att ges i slutet av varje stadium med syfte att studera nivå och förändring av prestationer på både elev- och skolnivå. Systemet kommer därmed även fortsättningsvis att ha starkt fokus på ansvarsutkrävande på skolnivå.

AUSTRALIEN

National Assessment Program (NAP) har nyligen introducerats i Australien med syfte att samla in, analysera och rapportera nationellt jämförbara data över elevers prestationsnivåer i läsning, matematik, naturvetenskap, samhällskunskap och IKT. Programmet består av två olika komponenter. Den ena benämns NAPLAN ("National Assessment Program – Literacy and Numeracy") och i denna ges prov i läsning/skrivning och matematik till samtliga elever i årskurs 3, 5, 7 och 9. Resultaten samlas in och görs tillgängliga på skolnivå i en databas, som också innehåller information om skolorna. Ett syfte är att denna information ska ge underlag för beslut om val av skola, men den ska också redovisa de enskilda elevernas prestationer. Den andra komponenten benämns "NAP Sample Assessments" (eller bara NAP). Som namnet antyder är detta ett urvalsbaserat system där elever i årskurs 6 och årskurs 10 medverkar i undersökningar av kunskaper och färdigheter inom de tre områdena naturvetenskap, samhällskunskap och informations- och kommunikationsteknologi (IKT). Denna del av programmet startade 2003, och varje område undersöks i en rullande treårscykel. Till denna komponent räknas också de tre internationella undersökningarna PISA, TIMSS och PIRLS (se nedan).

NAP-proven är länkade över undersökningsomgångarna så att det är möjligt att studera förändring över tid. Detta gäller även NAPLAN-proven, men dessa är också med hjälp av den moderna mätteori länkade över årskurserna, så att det är möjligt att på en och samma skala undersöka resultatförändringar mellan årskurser. I rapporteringen redovisas också hur stora andelar av eleverna som uppnår olika nationella mål genom att kompetensskalor (se ruta *Modern mätteori*, s. 48) skapats som kopplar elevernas provresultat till de nationella målen.

SVERIGE

Frågan om en nationell utvärdering med fokus på kunskaper och färdigheter har debatterats sedan 1970-talet i Sverige. I mitten av 1980-talet presenterade Skolöverstyrelsen ett förslag till en nationell utvärdering. I förslaget betonades behovet av generella och för hela landet giltiga beskrivningar och analyser såväl av utveckling av kunskapsnivåer över tid, som av likvärdigheten i utbildningen. Den första undersökningen genomfördes 1989 i årskurs 2 och 5. Efter Skolöverstyrelsens nedläggning 1991 fortsatte Skolverket arbetet och under vårterminen 1992 gjordes en datainsamling inom ramen för den andra nationella utvärderingen (NU-92). Detta var en bred undersökning med tyngdpunkt på årskurs 9 som skulle ge en samlad bild av skolans

verksamhet och av elevernas kunskaper (Skolverket, 1992).

Efter några mindre studier genomfördes vårterminen 2003 återigen en mer omfattande utvärderingsomgång (NU-03) (Skolverket, 2004, s. 10). Syftena var att

- * ge en helhetsbild av måluppfyllelsen i grundskolan, ämnesvis och i ett övergripande perspektiv
- * visa på förändringar sedan den nationella utvärderingen 1992
- * peka på behov av insatser.

53

Skolverket gick därmed tillbaka till ursprungstanken med den nationella utvärderingen och lade upp undersökningen för att i så stor utsträckning som möjligt göra en upprepning av NU-92 för att kunna göra jämförelser över tid.

Möjligheterna att använda jämförelsen mellan NU-92 och NU-03 för att studera kunskapsutvecklingen över tid visade sig dock vara begränsade (Gustafsson & Yang Hansen, 2009). Detta berodde dels på att de uppgifts- och provtyper som huvudsakligen användes inte tillät denna typ av jämförelser på grund av att de inte lämpade sig för kvantifiering, dels på att designen även i övrigt uppvisade betydande brister. De resultat som framkom pekade dock på försämrade resultat inom flera områden 2003 jämfört med 1992.

I en återblick på framväxten av systemet för kunskapsbedömningar på nationell nivå tar Öquist (2011), som var projektledare för NU, bestämt avstånd från beskrivningar i kvantitativa termer och det ”medeltalets tyranni” som han menar i alltför hög grad dominerat nationella utvärderingar. Han förespråkar i stället att individer får träda fram och berätta utan någon styrning. En möjlig förklaring till denna lättsinniga hållning till beskrivningar av kunskapsutvecklingen i skolan kan vara att Skolverket såg en annan möjlighet än stickprovsundersökningar för att göra sådana uppföljningar. När det målrelaterade betygssystemet ersatte det relativa betygssystemet öppnade sig nämligen möjligheten att använda betygen och de nationella proven för uppföljning av kunskapsutvecklingen.

Samtidigt som Skolverket tog över den av Skolöverstyrelsen initierade nationella utvärderingen, arbetade verket med att bygga upp ett alternativt system för kunskapsbedömning på nationell nivå, som presenterades av Pettersson & Wallin (1995). I kort sammanfattning innebar förslaget att den nationella utvärderingen skulle fokusera på att förklara och förstå uppnådda resultat. Hur detta skulle genomföras framgår dock inte, även om det är tydligt att det inte handlar om förklaringar i generaliserande, kausala,

termer. Förslaget innebar vidare att beskrivningen av resultatutvecklingen ska göras inom ett nationellt uppföljningssystem, där betygen och de nationella proven ska utgöra källan till information om utvecklingen av elevernas kunskaper och färdigheter. Ämnesbetygen i grundskolan och kursbetygen i gymnasiet var avsedda att ge information inte endast om den enskilda elevens studieresultat, utan även om skolors och kommuners resultat och resultatutveckling, samt om resultatutvecklingen på nationell nivå. Det faktum att det nya betygssystemet var förankrat i mål och kriterier bedömdes vara tillräckligt för att det skulle ge en god grund för uppföljning av utvecklingen av kunskaper och färdigheter över tid. Vidare menade man att den publicerade informationen om skolors betygsmedeltal skulle ge föräldrar och elever ett gott underlag för att bedöma olika skolors kvalitet inför skolval. Det system som Pettersson och Wallin (1995) skisserade kom inte att implementeras. Men betygen har ändå kommit att få en central betydelse i uppföljningen av skolans resultat, och de utgör sedan ett decennium den primära grunden för kunskapsbedömning på nationell nivå.

NATIONELLA SYSTEM FÖR KUNSKAPSBEDÖMNING: DISKUSSION

Denna korta redovisning av några system för kunskapsbedömning på nationell nivå visar att det finns många olika sätt att lägga upp sådana. En huvuddistinktion kan göras mellan urvalsundersökningar å den ena sidan, och totalundersökningar å den andra sidan. Det engelska "National Curriculum Assessment" är ett exempel på den senare typen, och hit kan även det svenska försöket att basera uppföljningen av utvecklingen av kunskaper och färdigheter på de målrelaterade betygen räknas. Det främsta exemplet på stickprovsundersökningar är det amerikanska NAEP, och de liknande nationella kunskapsbedömningssystem som det inspirerat till i andra länder.

Det svenska försöket att utnyttja det målrelaterade betygssystemet som grund för uppföljning av kunskapsutvecklingen på nationell nivå framstår som mindre lyckat, i det att det inte är möjligt att ha någon riktig tilltro till korrektheten i resultaten, på grund av bristande likvärdighet och betygsinflation. Det finns flera anledningar till detta. En anledning är att lärarnas utnyttjande av den nya betygsskalan genomgick stora förändringar under de första åren, från en stor återhållsamhet med att använda det högsta betyget till en högre grad av generositet. En annan anledning är att betygen, och då i synnerhet gymnasieskolans kursbetyg, varit utsatta för ett tryck uppåt från eleverna och deras föräldrar, på grund av den stora betydelse jämförelsetalet har vid ansökan till konkurrensutsatta högskoleprogram. I

samma riktning verkar det faktum att en hög betygsnivå för skolan är ett marknadsföringsargument i skolvalsfrågan. En ytterligare anledning är att de nationella proven under en period inte haft som huvuduppgift att stödja likvärdig betygssättning, och framför allt att de nationella proven inte utformats för att ge resultat som är jämförbara från ett år till ett annat. Skolverket har i sina PM om utfallet på proven också betonat att ”Proven prövar inte alla mål i kursplanen, de prövar inte heller alla mål lika mycket och det är olika uppgifter i proven från år till år. Proven är därför svåra att jämföra över tid och en större variation mellan åren är därför förväntad för proven jämfört med slutbetygen” (Skolverket, 2011).

Inte heller det engelska National Curriculum Assessment framstår som särdeles framgångsrikt. Även här finns misstro mot korrektheten i de resultat som presenteras. Systemet har dessutom deformerande effekter på undervisningen. Det tycks också finnas sidoeffekter i form av negativ påverkan på elevernas motivation (Harlen & Deakin Crick, 2002), och en stark press på lärare och skolledare.

Det som framstår som det grundläggande problemet i både det engelska och det svenska systemet är att de har flera olika syften som är oförenliga. Det svenska betygssystemet är ett *high-stakes*-system på både elevnivå och på skolnivå, vilket leder till ett inflationstryck på betygen. Det engelska systemet är inte *high-stakes* på elevnivå, men väl på lärar- och skolnivå, vilket leder till anpassning av undervisningen så att provresultaten maximeras. Detta innebär i sin tur både missvisande provresultat och suboptimal undervisning i förhållande till de mål som inte mäts av proven. Det faktum att båda systemen omfattar hela populationen av elever gör det också till en stor utmaning att utveckla prov som är jämförbara över tid. I England har man försökt att komma till rätta med detta problem, men med osäker framgång (Stobart, 2004). I Sverige har ännu inte någon seriös ansträngning gjorts att skapa nationella, betygssstödjande, prov som är jämförbara från ett år till ett annat.

De kunskapsbedömningssystem som baseras på stickprovsundersökningar har som huvudsyfte att beskriva förändringar på systemnivå, och är därför inte utsatta för störningar från syften som medför påverkan antingen på proven/betygssättningen eller på undervisningen. Problemet med stickprovsundersökningar kan snarare vara att resultaten på dessa saknar betydelse i så hög grad att eleverna inte anstränger sig tillräckligt att visa vad de kan. I allmänhet är dock syftet att studera förändring över tid, och under antagande att inverkan av denna faktor är konstant påverkar den inte skattning av förändring. I stickprovsundersökningar är det också betydligt lättare att utveckla prov som är jämförbara över tid, exempelvis genom att

använda sådana tekniker som används i de internationella undersökningarna, vilka diskuteras nedan.

Internationella undersökningar av elevers kunskaper och färdigheter

56

Internationella jämförande studier av elevers kunskaper och färdigheter har sedan mitten av 1990-talet kommit att få en ökande betydelse för utvärdering av skolsystem i många länder. Inte minst gäller detta Sverige, där de internationella studierna har infogats som delar i det nationella systemet för kunskapsbedömning (Skolverket, 2006). Kostnads- och kvalitetsmässigt har också de internationella undersökningarna stora fördelar före lokalt utvecklade system genom de synergieffekter och samordnings- och stordriftsfördelar som det internationella samarbetet medför.

Även om de internationella studiernas stora betydelse är ett relativt sentida fenomen påbörjades utvecklingen för mer än 50 år sedan i och med bildandet av IEA ("The International Association for the Evaluation of Educational Achievement") år 1958, se ruta *De internationella studiernas historia* för en historisk beskrivning.

Tyngdpunkten i IEA:s nuvarande verksamhet ligger på genomförande av regelbundet återkommande studier, dels av läsfärdighet i årskurs 4 vart femte år (PIRLS, "Progress in International Reading Literacy Study"), dels av matematik och naturvetenskap i årskurs 4 och 8 vart fjärde år (TIMSS, "Trends in Mathematics and Science Study"). I OECD:s PISA-undersökning studeras vart tredje år läsförståelse, matematik och naturvetenskap för 15-åringar. Syftet med dessa undersökningar är att jämföra länders resultat och att studera förändring av resultat inom länder över tid. Alla undersökningar bygger på de metoder som utvecklades inom NAEP.

Medan Skolöverstyrelsen var en ivrig tillskyndare av svenskt deltagande i de internationella studierna, var Skolverket fram till mitten av 2000-talets första decennium svalare i sin hållning. I TIMSS 1995 deltog därför inte Sverige med något urval av yngre elever, och det dröjde ända fram till 2007 innan urval av svenska elever från årskurs 4 deltog. Sverige var inte heller med i TIMSS 1999. I PIRLS har dock Sverige deltagit i samtliga undersökningsomgångar sedan starten 2001. Detta är också fallet med samtliga genomförda omgångar av PISA, som hade 2000 som startår. Som är väl känt visar resultaten från samtliga genomförda undersökningar på en nedgång i de svenska elevernas resultat från mitten av 1990-talet.

De internationella studiernas historia

Idén med IEA var att studier av variation mellan länder i utformning av utbildningssystem och effekter av sådan variation skulle göra det möjligt att använda "världen som ett pedagogiskt laboratorium". Den svenske pedagogikprofessorn Torsten Husén var ordförande för IEA under många år och under denna tid låg IEA:s huvudkvarter i Stockholm.

Den första IEA-undersökningen genomfördes 1964 med fokus på matematik. Åren 1970–1971 genomfördes den mycket omfattande sexämnesundersökningen, som bland annat undersökte läsförståelse, naturvetenskap, skrivning och franska. Under 1980-talet, ungefär 15 år efter de första undersökningarna, upprepades undersökningarna av matematik och naturvetenskap. I denna inledande fas av de internationella studierna utvecklade forskarna metoder samt samlade in och analyserade data för att kunna ge beskrivningar och förklaringar av olika utfall. Detta var ett pionjärarbete som krävde stora ansträngningar och som också var framgångsrikt i många avseenden. Tanken att använda de jämförande studierna för att dra slutsatser om vad som förklarar variation i kunskaper och färdigheter mellan olika länder blev dock inte någon framgång, eftersom designen med tvärsnittundersökningar (se ordlista) inte tillät slutsatser om orsakssammanhang (Gustafsson, 2008).

Från mitten av 1990-talet skedde en förskjutning av syftet från förklaring till beskrivning, med fokus på utvärdering och jämförelse av resultat mellan olika länder. Tekniken för att göra pålitliga beskrivningar hade också utvecklats genom mer avancerade urvalsmodeller, mätmetoder och analystekniker, och inte minst då genom utvecklingsarbetet i anslutning till det amerikanska systemet (NAEP). IEA sänkte också ambitionsnivån för den internationella rapporteringen till en förenklad deskriptiv rapportering. Efter att resultaten presenterats gjordes data fritt tillgängliga för forskare att göra sekundäranalyser. Från 1995 ökade också antalet deltagande länder, och studierna genomfördes med cyklisk upprepning med intervall på 3 till 5 år. Vidare påbörjade OECD år 2000 PISA-undersökningarna ("Programme for International Student Assessment"), vilka genom sin koppling till den ekonomiska och politiska makten kommit att ytterligare stärka de internationella studiernas betydelse.

När beskrivningar av utbildningsresultat visar på skillnader mellan länder eller förändringar över tid infinner sig omedelbart ett behov att förklara dessa skillnader och förändringar. Inom ramen för de internationella undersökningarna insamlas inte endast information om kunskaper och färdigheter, utan även information från skolledare, lärare, föräldrar och elever om bland annat hemmen, skolan, klassrummen och undervisningen, och om de olika aktörernas bakgrund, attityder och upplevelser. Att i tvärsnittsundersökningar relatera sådana förklaringsvariabler till prestationer inom olika områden säger dock inte så mycket, eftersom sådana samband i allmänhet inte kan ges en tolkning i termer av orsak och verkan. Under senare år har dock en lång rad mer sofistikerade analystekniker utvecklats för att bestämma kausala relationer från icke-experimentella observationsundersökningar och den rika tillgången på data av god kvalitet i de internationella undersökningarna gör det möjligt att med hjälp av dessa tekniker komma längre i analyser av orsaksförhållanden (Gustafsson, 2008, 2013a).

BEGRÄNSNINGAR I DE INTERNATIONELLA STUDIERNAS

Ett ofta uppmärksammat problem med de internationella studierna är att innehåll och provformer i större eller mindre grad står i samklang med läroplaner och praktik i de deltagande länderna. Skolverket har i en serie undersökningar analyserat hur väl de svenska läroplanerna stämmer överens med de internationella studiernas ramverk och prov (se Gustafsson, 2006a, för en mer detaljerad beskrivning). För läsning är det god överensstämmelse med PIRLS-undersökningen för de yngre eleverna i grundskolan, liksom med PISA för de äldre eleverna. När det gäller matematik och naturvetenskap i slutet av grundskolan är överensstämmelsen inte perfekt mellan svensk läroplan och PISA/TIMSS. Men de två internationella studierna kompletterar varandra, så sammantaget representerar de väl den svenska läroplanen.

Rullande kunskapsbedömningar

Sedan NU-03 genomfördes har informationen om kunskapsutvecklingen i den svenska skolan i all huvudsak baserats på betygen och de internationella studierna. Skolverket fick dock av regeringen i uppdrag att föreslå ett system för kunskapsbedömning på nationell nivå, vilket presenterades i en rapport 2006.

Skolverket (2006) diskuterar om det skulle vara lämpligt att fortsätta verksamheten med nationella utvärderingar av det slag som NU-92 och

NU-03 är exempel på. Slutsatsen är att alltför stora läroplansförändringar har skett för att detta ska vara meningsfullt. Man pekar också på att de nationella utvärderingarnas karaktär av resurskrävande punktinsatser inte gör att det i längden är hållbart att driva verksamheten i en sådan form. I stället menar man att ett system med kontinuerliga, rullande, kunskapsbedömningar skulle vara att föredra.

ETT FÖRSLAG TILL RULLANDE KUNSKAPSBEDÖMNINGAR

Inom ramen för regeringsuppdraget till Skolverket fick Jan-Eric Gustafsson i uppdrag att göra en genomgång av internationella erfarenheter av nationella kunskapsbedömningssystem och att på grundval av denna diskutera hur ett svenskt system skulle kunna utformas. Enligt uppdragsbeskrivningen skulle det primära syftet med kunskapsbedömningssystemet vara att beskriva utvecklingen av nivå och spridning i kunskaper över tid på nationell nivå, dels totalt, dels för olika undergrupper (t.ex. kön, social bakgrund, utländsk bakgrund). Ett andra syfte skulle vara att ge underlag för analyser av orsaker till förändringar i kunskapsnivåer. Enligt direktiven skulle samtliga grundskolans ämnen omfattas av den nationella kunskapsbedömningen, och den skulle genomföras årligen, med fokus på olika ämnen år från år.

Den modell som föreslogs av Gustafsson (2006a) var inspirerad både av det amerikanska NAEP-systemet och NEMP-systemet från Nya Zeeland, och utnyttjade också information från de internationella studierna. I modellen görs en distinktion mellan två olika ansatser till nationell kunskapsbedömning som skiljer sig åt metodmässigt i hur de underliggande kunskapsbedömningarna genomförs: en som betecknas som *uppgiftsorienterad*; och en som betecknas som *skalorienterad*. I den uppgiftsorienterade ansatsen ligger fokus på sådana kunskaper och färdigheter som inte låter sig mätas med kontinuerliga skalor. Uppgifterna är i allmänhet komplexa och omfattande, och kräver mångfacetterad bedömning och kvalitativ analys. Rapportering och den praktiska användningen av resultaten sker på en innehållsnära nivå, och den naturliga enheten är den enskilda uppgiften eller en grupp av likartade uppgifter. Några exempel på sådana uppgifter tagna från nationella provmaterial ges nedan.

- * Muntliga uppgifter i matematik, som genomförs i grupper om tre till fyra elever med en provtid per grupp på 20–30 min. Några av uppgifterna besvaras enskilt och andra löses gemensamt utifrån presenterat material. Förmågor som bedöms är problemlösning, metod, begrepp, resonemang och kommunikation.
- * Individuella uppgifter i historia, där eleverna med hjälp av textkällor i skrift

ska dra slutsatser, förklara sina ståndpunkter samt redovisa källkritiska argument. Det som bedöms är elevers förmåga att använda en historisk referensram, hantera källor, reflektera över egen och andras användning av historia samt att använda begrepp.

- * Muntligt prov i svenska i grupper om fyra till fem elever. Maximal provtid 110 minuter; individuellt text- eller bildunderlag; högläsning, kommentarer till text, ledning av och deltagande i diskussion utifrån givna frågor. Det som bedöms är elevers förmåga att presentera samt leda och delta i diskussionen.

I den skalorienterade ansatsen ligger fokus på skalor som representerar nivåer av prestationer inom olika innehållsliga domäner. Dessa skalor bygger i allmänhet på en stor mängd av uppgifter, som var och en är av mindre omfattning och som inte kräver komplex bedömning. Vanliga uppgiftstyper inom ramen för denna ansats är sådana som antingen kräver mycket korta elevproducerade svar eller val mellan olika alternativ. De senare kan vara av traditionell flervalskaraktär, i regel med tre eller fyra svarsalternativ, eller så kallade matchningsuppgifter, där exempelvis olika påståenden ska paras ihop med ett av flera text- eller bildalternativ. Antalet svarsalternativ är i detta fall betydligt större än i flervalsuppgifter. Matchning är dock också en typ av flervalsuppgift, där följaktligen ingen språklig produktion krävs i svaren.

Uppgifter med val mellan olika svarsalternativ framställs stundtals som en negativ provform, förknippad med ytliga och atomistiska kunskaper. Detta är dock en missuppfattning. Formatet kan tvärtom med fördel användas för att fånga kvalitativa nivåer av förståelse, som elever kan ha svårt att själva verbalisera. Läsande ”mellan och bortom raderna” brukar ibland användas som exempel på detta, liksom resonemang kring skilda företeelser. Det är också viktigt att återigen påminna om det hot mot validitet som har att göra med irrelevanta faktorer; elever måste ges möjlighet att visa kunnande av skilda slag utan att detta kräver skriftlig – eller muntlig – språkproduktion. Viktigt att påpeka är dock att såväl gedigen kompetens kring provkonstruktion som grundliga utprovningar krävs för att konstruera flervalsuppgifter. Det är därför ett format som i regel inte lämpar sig för lärares egna prov men väl för storskaliga provsammanhang.

Den mer kvantitativa skalorienterade ansatsen kan också inkludera sådana komplexa uppgifter som den uppgiftsbaserade ansatsen är fokuserad på. Men eftersom ett stort antal uppgifter krävs är det inte realistiskt att enbart använda komplexa uppgifter.

Som nämnts i översikten över nationella kunskapsbedömningssystem ovan, redovisas resultaten från den skalorienterade ansatsen i allmänhet i form av poäng på en skala av likartat slag som de skalor som används i de internationella studierna. Som förklaras i ruta *Modern mätteori* (s. 48) är det också möjligt att utveckla kompetensskalor där olika poängnivåer beskrivs i termer av vilka slag av uppgifter eleverna har god sannolikhet att bemästra. Sådana kompetensskalor gör resultaten enklare att tolka och att använda. De kan också göra det möjligt att mer konkret knyta resultat på skalorna till exempelvis kunskapskraven för olika betygssteg i det målrelaterade betygssystemet, eller till de särskilda behörighetskraven för olika utbildningar inom högskolan.

De uppgifts- och skalorienterade ansatserna har både för- och nackdelar och de är i olika grad användbara för olika ämnen och olika mål inom ämnen. Vi vill också understryka att de två ansatserna inte ska ses som artskilda, eftersom båda måste beakta de grundläggande principer som gäller för utveckling av högkvalitativa prov. Exempel på detta är validitets- och reliabilitetsrelaterade grundregler vid utveckling av uppgifter i form av tydliga specifikationer och successiva utprovningar där resultat kring uppgifter och elevers och lärares uppfattningar samlas in, liksom mångfacetterade analyser av data såväl från utprovningar som från färdiga prov.

Någon form av matrissamplingsteknik är nödvändig för båda ansatserna, och då framför allt för den uppgiftsorienterade ansatsen. För att det ska vara möjligt att få tillräckligt god täckning av de olika ämnenas innehåll behövs en stor mängd uppgifter med olika utformning och svarssätt, och det är orimligt att varje elev besvarar varje uppgift. Men det föreligger stora skillnader mellan hur matrissampling kan implementeras i den uppgiftsorienterade och den skalorienterade ansatsen.

I den skalorienterade ansatsen är det nödvändigt att se till att grupper av uppgifter som besvaras av olika elever överlappar. I den uppgiftsorienterade ansatsen finns inte något sådant behov, om inte uppgifterna också ska inkluderas i skalor. I den skalorienterade modellen krävs därför att det finns en direkt eller indirekt länk mellan alla uppgifter, vilket kan åstadkommas på flera olika sätt. Dock innebär detta restriktioner på uppgifternas utformning och framför allt är det viktigt att alla block av uppgifter ställer samma tidskrav och att instruktioner och genomförandeformer är enkla och enhetliga för alla uppgifter.

I den uppgiftsorienterade ansatsen kan en uppgift utformas friare, utan hänsyn till andra uppgifters krav och egenskaper, och det finns inget som hindrar att varje enskild elev endast gör en enda uppgift. Det är därför möj-

ligt att använda mer komplexa och omfattande uppgifter, där eleverna själva eller i grupper producerar mer omfattande svar. Detta är knappast möjligt inom ramen för den skalorienterade ansatsen.

Det viktigaste syftet med det nationella kunskapsbedömningssystemet skulle vara att följa eventuella förändringar i nivån av kunskaper och färdigheter över tid. Den uppgiftsorienterade ansatsen till trendmätning innebär med nödvändighet att vissa uppgifter som använts tidigare administreras återigen. Detta är en enkel teknik, men den har också vissa begränsningar. Ett problem är att de uppgifter som ska användas för att etablera trendlinjer måste hållas hemliga, och den uppgiftsorienterade ansatsen förlorar en del av sin kraft och mening om det inte är möjligt att presentera och diskutera resultat i anslutning till en presentation av uppgiften. Ett svenskt nationellt kunskapsbedömningssystem kräver också en sådan öppenhet att det inte är rimligt att bygga trendstudier på att samtliga uppgifter ska hållas hemliga, även om det naturligtvis är möjligt att undanta vissa uppgifter från publiceringskravet. Ett annat problem är att den uppgiftsorienterade ansatsen till trendmätning i första hand lämpar sig för redovisning på uppgiftsnivå. På grund av mångfalden av uppgifter ger dock inte detta en överskådlig information om resultaten på den nationella nivån. Denna uppgiftsnära redovisning lämpar sig inte heller för beskrivning av skillnader i resultat och resultatfördelningar för olika undergrupper, och inte heller för förklaringsinriktade analyser.

I den skalorienterade ansatsen genomförs trendmätning på så sätt att en viss andel av uppgifterna från en viss mätomgång (30–50 procent) hålls hemliga för användning i nästa mätomgång, varefter en del av dessa görs offentliga och ersätts med nya uppgifter. Denna teknik har nu använts under lång tid både i NAEP och andra nationella kunskapsbedömningssystem, liksom i de internationella undersökningarna, och den framstår som väl beprövad. En viktig förutsättning för att korrekta resultat ska nås är dock att förändringar i uppgifter och procedurer genomförs med försiktighet, och att trendmätningarna genomförs på grundval av ett väl definierat ramverk som i huvudsak är stabilt över tiden. Negativa erfarenheter från kvaliteten i trendmätningen inom PISA:s *minor domains* pekar också på att det är nödvändigt att använda ett stort antal uppgifter.

Av diskussionen ovan framgår att den uppgiftsorienterade ansatsen har begränsningar när syftet är att göra trendmätningar, medan den skalorienterade ansatsen medför begränsningar i vilka typer av uppgifter och bedömningssituationer som kan användas. Det vore därför olyckligt att utforma ett nationellt kunskapsbedömningssystem antingen som ett renodlat skal-

orienterat system eller som ett renodlat uppgiftsorienterat system. Dessa ansatser bör i stället kombineras på ett sådant sätt att fördelarna hos de olika ansatserna kan tas till vara. Den föreslagna modellen för rullande kunskapsbedömningar innebär därför att dessa två ansatser ska kombineras till en integrerad helhet. Båda ansatserna kan förväntas förekomma inom samtliga ämnesområden, även om tyngdpunkten också kan förväntas vara olika för olika ämnen.

Det av Gustafsson (2006a) formulerade förslaget till rullande kunskapsbedömningar fokuserar i enlighet med uppdraget helt på grundskolan, men även för gymnasieskolan finns behov av att på nationell nivå följa förändringar i kunskapsnivåer. Detta gäller i synnerhet som det på gymnasienivån endast finns en internationell undersökning, nämligen TIMSS Advanced som avser matematik och fysik i sista året i gymnasiet. Det finns dock inget som hindrar att ett nationellt system för kunskapsbedömning på gymnasienivån utformas enligt samma grundläggande principer som det system som här skisseras för grundskolan.

ANALYS OCH TOLKNING

De internationella erfarenheterna visar att det är nödvändigt att ha möjlighet till fördjupad analys och tolkning av de observerade förändringarna, eftersom de olika intressenter som tar del av resultaten också förväntar sig förklaringar. Även om svårigheterna att nå fram till entydiga slutsatser om kausala relationer är stora, kan ett empiriskt underlag ge bättre förutsättningar för en fruktbar diskussion. Eftersom undersökningarna i allmänhet är upplagda som tvärsnittsundersökningar är det svårt att dra slutsatser om kausala effekter (Gustafsson, 2008, 2013a), så i den mån det är möjligt att använda longitudinell undersökningsdesign bör detta prövas.

Gustafsson (2006a) pekade på att den uppgiftsorienterade ansatsen kan möjliggöra kvalitativa analyser som visar förändring över tid, vilka kan utgöra värdefulla tolkningsunderlag i jakten på förklaringar. Samtidigt ger den skalorienterade ansatsen bättre möjligheter att med statistisk metodik undersöka effekter av förklaringsfaktorer på olika nivåer inom och utom utbildningssystemet.

Enkäter till elever, lärare, skolledare och eventuellt även till föräldrar är den viktigaste källan till information om olika tänkbara förklaringsfaktorer, och om undervisningens inriktning och uppläggning. Med hjälp av enkäter insamlas också information om viktiga utfall, exempelvis elevers egna bedömningar av sina kunskaper och färdigheter, attityder till olika ämnen och ämnesdelar, och motivation att lära.

SKOLVERKETS VÄRDERING AV FÖRSLAGET

Efter beräkning av förväntade kostnader vid full implementering av Gustafssons (2006a) förslag till rullande kunskapsbedömningar i samtliga grundskolans ämnen fann Skolverket (2006) att såväl implementerings- som driftskostnader skulle bli alltför höga. I sitt förslag till regeringen föreslog därför Skolverket (2006) en partiell implementering av förslaget i en grupp av ämnen där bristen på information är mest påtaglig, som svenska, språk och samhällsorienterande ämnen. I övrigt menade Skolverket att en systematisk inkorporering av de internationella studierna i det nationella utvärderingsprogrammet skulle täcka behoven i flera ämnen, eventuellt med utnyttjande av möjligheten att göra nationella tillägg. Något förslag från regeringen om uppbyggnad av ett svenskt nationellt kunskapsbedömnings-system existerar dock ännu inte.

NATIONELLA PROV SOM INDIKATORER PÅ FÖRÄNDRINGAR I KUNSKAPSNIVÅER

Tanken att de nationella proven skulle kunna ge information om förändringar i kunskapsnivåer över tid har väckts vid flera tillfällen. Ett av de uttalade syftena med de nationella proven är att de ska ge underlag för att analysera i vilken utsträckning kunskapskraven uppfylls på skolnivå, huvudmannanivå och nationell nivå. Det finns dock flera omständigheter som gör att detta inte är en särskilt framkomlig väg att vinna kunskap om utvecklingen av kunskapsnivåer på nationell nivå. Ett problem är att de nationella proven ofta innehåller ett alltför begränsat antal uppgifter för att en acceptabel täckning av innehållet ska kunna uppnås. I de nationella proven är det inte heller möjligt att använda matrissampling eftersom rättvisekrav medför att alla elever bör få samma prov.

Ett annat problem är att det innebär stora praktiska utmaningar att utveckla och länka uppgifterna i de nationella proven. Detta har framför allt sin grund i att alla elever ska delta i proven, samtidigt som det är nödvändigt att använda ett urval av elever ur samma population för att utveckla och länka provuppgifterna.

Intressant nog finns det en lösning på detta problem, om man inför ett urvalsbaserat nationellt kunskapsbedömningsystem av det slag som diskuteras ovan. Det är rimligt att anta att det inom varje ämne finns åtminstone något eller några delområden som skulle lämpa sig för en skalorienterad ansats. I det nationella kunskapsbedömningsystemet bör dessa då representeras av ett stort antal uppgifter som används i en matrissamlingsdesign. En del av uppgifterna ska släppas fria, men innan så sker används vissa av

dem i det nationella provet. Om det finns tillräckligt med länkade uppgifter i det nationella kunskapsbedömningssystemet som kan användas till nationella prov under några påföljande år kommer alla prov att automatiskt vara länkade med varandra, vilket innebär att de ger jämförbara resultat uttryckta på samma skala. Eftersom det nationella kunskapsbedömningssystemet ger stabil information om kunskapsutvecklingen ger de nationella proven inte något egentligt tillskott av information om detta, förutom att de ger information om alla elever, klasser, skolor, huvudmän och kommuner i landet. Men poängen är att vi på ett enkelt sätt har skapat en komponent i det nationella provsystemet som kan användas för att stödja en likvärdig betygssättning som är stabil över tid. Genom att använda sådana länkade delprov tillsammans med komplexa, autentiska, uppgifter kan vi skapa ett provsystem som kan tillgodose krav på både reliabilitet och validitet.

En förutsättning för att de nationella proven ska kunna bidra till en likvärdig betygssättning är att resultat på provet är explicit kopplade till de kunskapskrav som återfinns i styrdokumentet. För närvarande görs detta genom en process som kallas *standard-setting* där för varje ny omgång prov en grupp bedömare söker fastställa vilka poänggränser som svarar mot de olika kunskapskraven. Detta är en svår uppgift som kan leda till påtagligt varierande resultat (Koretz, 2008). Som vi redan konstaterat är detta sannolikt en bidragande orsak till stabilitetsbristerna i det nuvarande nationella provsystemet.

Med den föreslagna modellen med koppling mellan nationell kunskapsbedömning och nationella prov är det möjligt att genomföra anknytningen mellan prov och kunskapskrav på ett annat sätt. Eftersom varje nationellt prov inom ett visst område är baserat på samma skala är det möjligt att här i stället ta utgångspunkt i kompetensskalan och identifiera de nivåer på skalan som svarar mot de kunskapskrav som är formulerade för de olika betygsstegen. Eftersom informationen här är betydligt mer omfattande och empiriskt välgrundad än vid vanlig *standard-setting* finns en mer stabil grund för identifiering av skalnivåer som motsvarar de olika betygsstegen. Eftersom detta bara behöver göras en enda gång är det möjligt att lägga mer omsorg och resurser på arbetet än om det ska upprepas varje år. I fortsättningen behöver endast en omräkning ske av poänggränserna för betygsstegen med utgångspunkt i det nya provets egenskaper, vilka kan bestämmas från den information som är tillgänglig för uppgifterna. De instabilitetsproblem som orsakas av den årliga ombestämningen av betygsgränserna i det nuvarande systemet försvinner sålunda. Det är dock viktigt att påpeka att detta arbetssätt förutsätter explicita provspecifikationer som

tillåter en över åren standardiserad provsammansättning. Det bör också betonas att det kollaborativa arbetssätt som kännetecknar utvecklingen av de nationella proven, under medverkan av flera kategorier av experter, inte minst lärare och elever, är av fortsatt central betydelse för upprätthållande av validitet i systemet.

Diskussionerna kring nationell kunskapsbedömning har i första hand rört grundskolan men även i gymnasieskolan finns stora behov av att kunna följa kunskapsutvecklingen. Med tanke på gymnasiebetygens stora betydelse för individerna är behovet av stöd för likvärdig betygssättning stort också på gymnasiet. Gymnasieskolans kursutformning ger dock något annorlunda förutsättningar än de som gäller för grundskolan. Centralt innehåll, mål och kunskapskrav är i Lgy 11 formulerade för varje kurs för sig. Läroplanen anger dock för alla de gymnasiegemensamma ämnena syfte och de övergripande kunskaper och färdigheter som undervisningen ska ge eleverna förutsättningar att utveckla. Kunskapskraven för de olika kursbetygsstegen anges i samma bra/bättre/bäst-form som för grundskolans ämnesbetyg, men formuleringarna för de olika kurserna anger också progression mellan kurserna. Dessa gemensamma och övergripande delar gör det sannolikt möjligt att skapa integrerade ramverk för nationell kunskapsbedömning på ämnesnivå inom gymnasieskolan, men där också de olika kurserna inom ämnet är identifierbara enheter. Sådana ramverk skulle ge en god grund både för nationell kunskapsbedömning och för vidareutveckling av nationella prov för olika kurser.

Kunskapsbedömning på skolnivå

Medan staten har ansvaret för utvärdering på nationell nivå är ansvaret för uppföljning och utvärdering på skolnivå delat mellan staten och huvudmännen. För närvarande finns det framför allt fyra informationskällor för skolors arbete och resultat.

Den första är den officiella statistiken över slutbetygen från årskurs 9 (till exempel meritvärde, jämförelsetal, andel elever som når godkänt resultat och provbetyg), vilken i relativt lätt tillgängliga databaser (SIRIS, SALSA) kan aggregeras till skol- och kommunnivå. Här finns även möjlighet att göra statistiska kontroller för elevsammansättning med avseende på utbildningsbakgrund, migrationsbakgrund och kön, för att i någon mån söka bestämma skolans bidrag till slutbetygen.

Den andra informationskällan är Skolinspektionen, som ungefär vart

fjärde år genomför reguljära inspektionsbesök i alla landets skolor. Granskningen avser såväl legala aspekter i form av hur regler efterlevs, som kvalitetsaspekter i undervisning och annan verksamhet. Varje skola får återkoppling i form av ett inspektionsprotokoll, vilket i många fall föreskriver olika förbättringsåtgärder som måste vidtas inom en viss tid. Det bör dock påpekas att Skolinspektionen inte gör någon egen värdering av skolornas uppnådda resultat, utan baserar sina bedömningar av dessa på statistik över betygs- och provresultat.

Den tredje informationskällan är de självvärderingar som skolorna genomför. Fram till för några år sedan fanns för kommunala skolor ett obligatoriskt system för kvalitetsvärderingar men detta är nu nedlagt. Kravet på skolor och kommuner att redovisa sitt kvalitetsarbete kvarstår dock.

Den fjärde informationskällan är enkätundersökningar riktade till bland annat föräldrar och elever om hur de upplever skolans verksamhet. Flera olika mer eller mindre standardiserade undersökningssystem erbjuds av olika företag. Även staten och olika institutioner har utvecklat system för att mäta exempelvis skolklimat.

Nusche m.fl. (2011) påpekar att de svenska kommunerna har mycket olika förutsättningar att genomföra kvalitetsarbete och systematiskt utvecklingsarbete, bland annat på grund av olikheterna i storlek och den varierande tillgången på kompetens. Det faktum att stora delar av ansvaret för utvärdering på skolnivå är decentraliserat till huvudmännen kan därför förväntas leda till ökande skillnader mellan skolor. Det kan också konstateras att det framför allt är kommuner i storstadsområdena som har förbättrat utbildningsresultaten i grundskolan sedan 1998, mätt med meritvärde (Gustafsson & Yang Hansen, 2011).

Det svenska systemet för redovisning av skolresultat i form av rankinglistor är i hög grad ett *high-stakes*-system, och jämfört med andra system för att fastställa skolans bidrag till elevens resultat är det svenska SALSA-systemet (se ordlista) primitivt. Nusche m.fl. (2011) reser därför frågan vilka negativa konsekvenser som detta system har medfört, bland annat hur lärarna påverkas, dock utan att själva ge något svar.

Sammanfattningsvis kan vi konstatera att kunskapsbedömning på skolnivå inte bidrar med någon ytterligare information om vare sig skillnader i kunskaper och färdigheter mellan olika skolor eller om utvecklingen över tid. Även här är den information som ges av betyg och nationella prov det som används i redovisningarna och rangordningarna av skolor. Även Skolinspektionen tar utgångspunkt i den officiella betygsstatistiken och bidrar inte med några självständiga bedömningar av skolornas resultat. Bristerna i

betygens likvärdighet och stabilitet över tid gör dominansen av denna enda informationskälla problematisk.

Inte heller ger bedömningar på skolnivå systematisk information om hur verksamheten bedrivs vid olika skolor, vilket skulle kunna vara värdefull information både för att inspirera andra skolor och för att förstå skolans resultat. Det finns sålunda behov av att komplettera den lokala utvärderingsinformationen med ytterligare information om både förutsättningar och resultat.

Sammanfattning och slutsatser

Sedan lång tid tillbaka har det gjorts försök att få till stånd ett svenskt nationellt kunskapsbedömningssystem. Skolöverstyrelsen lade grunden till ett sådant system under slutet av 1980-talet men eftersom verket lades ner strax efter att den första undersökningen genomförts försvann möjligheterna att vidareutveckla systemet. Skolverket genomförde visserligen nästa planerade undersökning i frågan, men Skolverkets halvhjärtade och missriktade satsningar under 1990-talet ledde inte till någon vidareutveckling. Den sista nationella kunskapsbedömning som Skolverket genomförde blev NU-03, men denna bidrog endast i begränsad utsträckning till att beskriva förändringar i kunskapsnivåer över tid totalt och för olika undergrupper. Även om Skolverket senare försökt att aktualisera frågan om nationell kunskapsbedömning har inget hänt.

Under tiden har de jämförande internationella undersökningarna kommit att bli den främsta källan till information om utvecklingen av de svenska elevernas kunskapsresultat. Även om denna information är av god kvalitet i många avseenden är den dock begränsad till att gälla ett fåtal ämnesområden och åldersnivåer, och den har de begränsningar som följer med storskalig informationsinsamling med hjälp av papper och penna. Vidare har betygen kommit att få en alltmer stärkt ställning som informationskälla om utvecklingen av kunskaper och färdigheter. Det faktum att den resultatbild som beskrivningar baserade på betygen ger inte är i överensstämmelse med den resultatbild som framträder från de internationella undersökningarna är ett stort bekymmer.

När det gäller kunskapsbedömning på nivåer under den nationella (kommun och skola) framträder återigen betygen som den helt dominerande informationskällan om skillnader i resultat och utveckling.

Det finns uppenbara behov av ett nationellt stickprovsbaserat kunskapsbedömningssystem som har förmåga att fånga trendmässiga förändringar

i nivå på kunskaper och färdigheter och även ge information om i vilken grad de mer svårfångade, kvalitativa målen för skolan uppnås. Den grundläggande konstruktion av det rullande system för kunskapsbedömningar som föreslagits, med en kombination av en skalorienterad ansats och en uppgiftsorienterad ansats framstår som ändamålsenlig för dessa syften. Den mer detaljerade konstruktionen av systemet kräver dock fortsatt utredningsarbete.

I detta arbete är det i hög grad intressant att också knyta an till den idé som skisserats ovan, nämligen att koppla samman den skalorienterade ansatsen inom den nationella kunskapsbedömningen med de komponenter/delprov inom det nationella provsystemet som kan konstrueras i form av skalor, och som kan bidra till att stödja en likvärdig och stabil betygssättning. Detta skulle lösa flera svåra metodproblem vid utveckling och länkning av de nationella proven, och det skulle ge synergieffekter och kostnadsbesparingar. Vi återkommer till denna tanke i den avslutande diskussionen i kapitel 5.

4. Antagningssystemet till högre utbildning

Betyg från gymnasieskolan har sedan mycket länge varit den huvudsakliga grunden för såväl bedömning av behörighet som för urval till högre utbildning, även om Högskoleprovet fick en likvärdig roll som urvalsinstrument under 1990-talet. För att betygen även fortsättningsvis ska kunna spela en central roll i antagningssystemet krävs en tilltro inte bara till betygens prognosförmåga utan även till att de är rättvisa i det att betygssättningen är likvärdig såväl mellan lärare och skolor som över tid. Ytterligare hot mot betygens goda förutsättningar att utgöra ett rättvist och effektivt urvalsinstrument är de förändringar av betygssystem och skalor som genomförts under de senaste decennierna, liksom senare års förändringar av meritvärderingen av betygen som urvalsinstrument.

De kontinuerliga förändringarna av antagningssystemet och dess ökande komplexitet gör det svårt att utvinna kunskap om dess brister och förtjänster. Systematiska utvärderingar och uppföljningar har genomförts endast i mycket begränsad utsträckning på myndighetsnivå, och forskningen om antagningsfrågorna är begränsad. Med tanke på rättssäkerhetskraven och den stora praktiska, ekonomiska och politiska betydelse som antagningssystemet har, framstår det dock som nödvändigt att identifiera brister i såväl enskilda komponenter som i hela systemets funktionssätt, liksom att skissera på möjliga sätt att komma tillrätta med bristerna.

Som utgångspunkt diskuteras först antagningssystemets funktioner och vilka övriga krav det bör uppfylla. I avsikt att åskådliggöra systemets alltmer komplexa utformning ges en kort historik av dess tillkomst och utveckling fram till idag. Därefter förs en kritisk diskussion av systemets enskilda komponenter, strukturerat utifrån de två huvudfunktionerna behörighet och urval, med en tyngdpunkt på den senare. Kapitlet avslutas med en sammanfattning och slutsatser avseende identifierade brister och förslag till förbättringar.

Antagningssystemets funktioner

Antagningssystemet till högre utbildning förändras ständigt för att anpassas till nya omständigheter. Det har på så vis kommit att utformas för att fylla allt fler funktioner och krav. Det finns dock en risk att ju fler funktioner och krav som ska uppfyllas, desto sämre precision. Det är därför viktigt att systemets huvudfunktioner tydliggörs. En sådan är att systemet ska reglera vilka förkunskaper, formulerade som villkor för grundläggande respektive särskild behörighet, som ska gälla för olika utbildningar. Grundläggande behörighet krävs för samtliga utbildningar, medan särskild behörighet innebär krav på särskilda förkunskaper som anses nödvändiga för en specifik utbildning. En andra huvudfunktion är att reglera vad som ska gälla när antalet sökande till en utbildning är högre än antalet tillgängliga platser (7 kap. Högskoleförordningen). Om antalet sökande överstiger antalet utbildningsplatser spelar urvalsreglerna en avgörande roll, medan endast behörighetsvillkoren är avgörande om antalet sökande understiger antalet utbildningsplatser.

Det är också viktigt att göra klart vilka övriga krav som rimligen bör ställas på ett antagningssystem. Ett huvudkrav är att systemet ska vara effektivt, utifrån såväl student- som systemperspektiv. De sökande som har bäst förutsättningar ska därför antas, vilket förutsätter att de urvalsinstrument som används har god prognosförmåga. Ett annat krav är att antagningssystemet ska vara rättssäkert eftersom tillträde eller inte tillträde till en utbildning kan få livsavgörande konsekvenser för enskilda individer. Detta innebär bland annat att det ska vara tydligt, överblickbart och förutsägbart samt hållbart över rimlig tid. Systemet ska också vara rättvist med avseende på olika grupper av sökande och vara fritt från negativa konsekvenser på andra utbildningar. Systemet bör snarare ge positiva signaleffekter. Det vill säga att systemet snarare bör påverka elevernas attityder och beteenden i positiv riktning, bland annat genom att motivera till studier, framför allt inom gymnasieskolan, i det att det ska ”löna sig att plugga”.

I internationell forskning (t.ex. Atkinsson & Geiser, 2010) diskuteras fyra bärande begrepp: prognosförmåga, rättvisa, signaleffekter och kontinuitet mellan nivåerna i utbildningssystemet, vilka de menar är väsentliga att förhålla sig till när antagnings- och urvalssystem till högre utbildning utvärderas och modifieras.

Antagningssystemets tillkomst och utveckling fram till idag

72

Omfattningen av den högre utbildningen har ökat dramatiskt under de senaste 60 åren. År 1947 uppgick antalet studenter till färre än 15 000, fördelade på två universitet och en dryg handfull högskolor. Regler för antagning och urval var vid den tiden inga problem, då högre utbildning var ett privilegium för en begränsad grupp. Situationen förändrades under 1950-talet då högre utbildning började få större betydelse för fler individer. Den första stora expansionen av högskolan inträffade under 1960-talet. Antalet studenter har sedan dess vuxit kraftigt. År 2010 var antalet helårsstudenter det högsta någonsin, drygt 369 000, fördelade på cirka 60 universitet, högskolor och enskilda utbildningsanordnare. Antalet nya sökande har ökat mycket kraftigt, till det högsta antalet någonsin inför höstterminen 2009, för att sedan ligga på ungefär samma nivå under efterföljande år. Antalet antagna har dock minskat under de senaste åren, vilket innebär att det har blivit svårare att bli antagen till högskolestudier.

Alltsedan den första stora expansionen av högskolan har regler för antagning och urval varit föremål för flera utredningar och reformer med en mängd mindre förändringar och justeringar både mellan och efter de större reformerna. Nedan ges en översikt av systemets utveckling fram till idag.

FRÅN INKLUDERING OCH ENHETLIGHET TILL AVREGLERING OCH DECENTRALISERING

Före år 1977 baserades behörighet till högskoleutbildning i huvudsak på studieförberedande gymnasieutbildning och platsbegränsningar fanns bara för en mindre del av utbildningarna. Genom högskolereformen 1977 vidgades definitionen av högskoleutbildning genom att ett stort antal postgymnasiala utbildningar integrerades i högskolan. Vidare etablerades ett antal nya högskolor i avsikt att få en större geografisk spridning av den högre utbildningen. Det nya antagningssystemet kännetecknades av enhetlighet, samordning och formell hantering av de sökandes meriter. I ambition att bredda rekryteringen liberaliserades de allmänna behörighetskraven till att även omfatta två-åriga gymnasielinjer, viss folkhögskoleutbildning, och 25:4-regeln (fyllda 25 år och fyra års arbetslivserfarenhet samt vissa kunskaper i engelska) infördes. Då tillströmningen av antalet sökande blev mycket större än förväntat infördes så småningom platsbegränsningar på de flesta utbildningar.

Som underlag för urval infördes ett kvotgruppssystem kopplat till de olika

behörighetsgrunderna, där kvotgrupperna tilldelades platser proportionellt till antalet sökande. För sökande med behörighet enligt 25:4-regeln infördes Högskoleprovet som urvalsgrund. Arbetslivserfarenhet gavs i denna kvotgrupp lika stor vikt som provresultatet. Inom systemet infördes möjligheter till extra poäng för arbetslivserfarenhet och för fackligt och ideellt arbete att lägga till betygspoängen. På grund av oväntade negativa effekter fattades redan efter några år beslut om att revidera reglerna för dessa extra poäng.

Inför höstterminen 1991 gjordes Högskoleprovet tillgängligt som en "andra chans" även för sökande med gymnasiekompetens. Betyg och Högskoleprov skulle vara de generella urvalsgrunderna där 60 procent skulle antas på betyg och 40 procent på Högskoleprovet. Provurvalet delades upp i två grupper där det i den ena gruppen endast togs hänsyn till provresultatet medan det i den andra gruppen även togs hänsyn till den sökandes arbetslivserfarenhet.

År 1993 infördes en reform som innebar en betydande avreglering och decentralisering till de enskilda lärosätena att själva besluta om behörighet och urval. Den lokala friheten ledde till att de särskilda behörighetskraven och urvalsreglerna kom att variera över såväl utbildningar och lärosäten som över tid. Den allmänna behörigheten förblev dock centralt reglerad.

ÅTERSÖKANDE AV ETT CENTRALT SYSTEM

Avsikten med att införa ett nytt antagningssystem 1997 var att återskapa en nationell ordning som var överblickbar för de sökande. Syftet var att tillgodose studenternas rättssäkerhet i antagningsprocessen. Regelverket skulle också anpassas till sökande från den nya gymnasieskolan och ett betygssystem med fyra skalsteg och kursbetyg. Centralt reglerade standardbehörigheter för särskild behörighet infördes, vilka fastställdes för de utbildningar som gav yrkesexamen, medan högskolorna själva kunde fastställa vilka standardbehörigheter de ville tillämpa för övriga utbildningar.

Betygen fick fortsatt vara en viktig urvalsgrund. Samtidigt ansågs det viktigt med ett kompletterande urvalsinstrument för dem som av olika anledningar inte lyckats så bra i gymnasieskolan. Vid den tidpunkten var forskning om Högskoleprovets prognosförmåga i det närmaste obefintlig, men det fanns en föreställning om att Högskoleprovet uppvisade ett nästan lika högt samband med framgång i högre studier som betygen. Därför ansågs Högskoleprovet fortsatt lämpligt som kompletterande urvalsinstrument, och att det kunde användas för dem som saknade formell gymnasieutbildning. Minst en tredjedel skulle antas på betyg och minst en tredjedel på resultat på Högskoleprovet. Regeln att hälften av högskoleplatserna i hög-

skoleprovgruppen skulle reserveras för sökande med tilläggspoäng för minst fem års arbetslivserfarenhet kvarstod. Kompletteringsbetyg fick medräknas vid meritvärderingen av betygen och sökande från den gamla respektive den nya gymnasieskolan skulle komma att behandlas i olika kvotgrupper.

Den sedan 1980-talet pågående försöksverksamheten med särskilda urvalsmetoder reglerades genom att de endast fick användas vid urval till sådana yrkesutbildningar som ställer särskilda krav på vissa personliga egenskaper eller kompetenser. Endast en andel av platserna på en utbildning fick, efter tillstånd från Högskoleverket, tillsättas genom särskilt urval.

NUVARANDE UTFORMNING AV ANTAGNINGSSYSTEMET

Den senaste stora förändringen av antagningssystemet beslutades år 2006, medan de flesta regeländringar kom att gälla från och med höstterminen 2010. Sedan dess har nya reformer av betydelse för antagningssystemet införts, till exempel en ny gymnasieskola (GY11) och en sexgradig betygsskala. Därutöver har systemet justerats på en rad punkter av andra skäl.

BEHÖRIGHET ENLIGT NU GÄLLANDE REGLER

Reglerna för grundläggande behörighet har stramats upp genom att till exempel samlat betygsdokument från kommunal vuxenutbildning (komvux, se ordlista) togs bort som behörighetsgivande dokument. De nya reglerna innebär också att möjligheterna att ”toppa” jämförelsetalet (se ordlista) genom reducerat program eller att slutföra gymnasiestudierna inom komvux, med ett samlat betygsdokument försvann. Grundläggande behörighet enligt 25:4-regeln togs bort.

Enligt tidigare gymnasireform (Lpf 94) gav samtliga gymnasieprogram grundläggande högskolebehörighet, vilket inte är fallet i och med införandet av den nya gymnasieskolan 2011, då kraven för behörighet till högskolan dessutom ytterligare skärpts. På cirka hälften av yrkesprogrammen finns möjlighet att läsa in grundläggande behörighet inom programmets ordinarie poäng genom att utnyttja det individuella valet. En förutsättning för detta är att huvudmännen har valt att erbjuda sådana kurser och att eleven utnyttjar sitt individuella val. På vissa gymnasieskolor begränsas dock denna möjlighet, då det måste finnas tillräckligt många elever som väljer en kurs för att den ska ges inom ramen för det fria valet. För de övriga yrkesprogrammen krävs att eleverna läser utökat program för att kunna nå grundläggande behörighet. Detta ändrades dock helt nyligen, och det ska även inom dessa program vara möjligt att läsa grundläggande behörighetskurser inom ramen för den ordinarie programpoängen.

För den särskilda behörigheten har så kallade områdesbehörigheter införts. Dessa beslutas av Universitets- och högskolerådet för program på grundnivå som leder till yrkesexamen, medan högskolorna själva beslutar vilka områdesbehörigheter som ska gälla för övriga utbildningar. Högskolorna kan sänka kraven på särskild behörighet inom vald områdesbehörighet men inte höja kraven utan tillstånd från Universitets- och högskolerådet.

URVAL ENLIGT NU GÄLLANDE REGLER

I avsikt att stimulera elever att välja vissa kurser i gymnasieskolan infördes möjligheten att tillgodoräkna sig extra poäng (meritpoäng) att lägga till jämförelsetalet vid ansökan till högskolan. Initialt kunde enbart vissa grupper av sökande, beroende på avgångsår från gymnasieskolan, få extra poäng för godkända kurser på en nivå över de behörighetsgivande kurserna i språk och matematik. Meritpoäng kunde också få tillgodoräknas för särskilt angivna kurser inom det område som eleven valt att studera på högskolan. De nya reglerna skapade dock orättvisor, framför allt för sökande med utländska betyg som på grund av meritvärderingssystemet tilldelades en egen kvotgrupp, och för sökande med äldre betyg, som inte hade möjlighet att tillgodoräkna sig meritpoäng. Årliga regeländringar har genomförts för att i någon mån korrigera för detta. På så vis har den särskilda kvotgruppen för utländska betyg kunnat tas bort från och med antagning till utbildningar som påbörjades höstterminen 2012. De med betyg enligt 1992 års kursplaner eller senare har delvis blivit gottgjorda för de orättvisor som meritpoängen ger upphov till, medan sökande med äldre betyg inte kommer att få tillgodoräkna sig meritpoäng. Därutöver gäller ytterligare nya meritvärderingsregler för dem som läser gymnasieskolan enligt GY11. Reglerna för meritpoäng för så kallade områdeskurser, kopplade till områdesbehörigheterna, avskaffas och reglerna för meritpoäng för vissa kurser i språk och matematik ändras. Detta innebär att de olika förutsättningar som gäller för olika grupper att uppnå maximalt antal meritpoäng (2,5 p) kommer att kvarstå under överskådlig tid.

Därutöver avskaffades möjligheten att få extra poäng för arbetslivserfarenhet att lägga till högskoleprovpoängen och en ny kvotgrupp infördes för sökande med kompletterade betyg, oavsett om det rör sig om behörighets- eller konkurrenskomplettering (se ordlista). Antalet platser i kvotgruppen reducerades med en tredjedel, som fördes över till gruppen sökande med betyg från gymnasieskolan. Med den nya gymnasieskolan infördes även en ny sexgradig betygsskala (A–F). Enligt det ursprungliga förslaget skulle sökande med den nya skalan konkurrera i samma betygskvotgrupp som övriga

sökande. Då det kommer att bli svårare att uppnå högsta betyg med en sexgradig skala, kommer sökande från den nya gymnasieskolan att ha svårare att konkurrera om platserna på högskolan om samtliga antas i samma kvotgrupp. I förhoppningen att lösa detta problem införs, under en övergångsperiod om tre år, en egen kvotgrupp för sökande med en sexgradig betygsskala. Sökande från den nya gymnasieskolan kommer dock också att konkurrera i den ordinarie betygskvoten.

Från och med höstterminen 2003 har högskolorna rätt att själva besluta vilka urvalsregler som ska tillämpas vid antagning till 10 procent av utbildningsplatserna på en utbildning. Andelen ökades senare till 20 procent och från och med höstterminen 2008 kan högskolorna besluta om urvalsgrunder genom så kallade alternativa urvalsformer till högst en tredjedel av platserna inom en utbildning. Högskolornas möjlighet att bredda rekryteringen var det argument som angavs för införandet av den så kallade fria kvoten höstterminen 2003. Argumentet för att utöka den fria kvoten till en tredjedel höstterminen 2008 var att man ville värna högskolornas frihet och ökade möjligheter att ge sina utbildningar lokala profiler. Därutöver kvarstår regelverket från 1997 om att, efter tillstånd från Universitets- och högskolerådet, få använda särskilda prov vid urval till 50 procent av platserna till vissa utbildningar.

Från och med höstterminen 2011 infördes ett nytt Högskoleprov. Huvudinriktningen på utvecklingsarbetet har varit att öka antalet kvantitativa uppgifter i provet, så att dessa uppgår till samma antal som de verbala. Bakgrunden till detta var forskningsresultat som visade att Högskoleprovet har en svag prognosförmåga, särskilt för utbildningar inom det naturvetenskapliga området. En grupp internationella utvärderare hade också rekommenderat en uppdelning i poäng för en verbal och en matematisk del. Genom att de båda delpoängen skulle kunna användas antingen separat eller viktas olika beroende på utbildningarnas särart skulle prognosförmågan kunna höjas. Enligt det regelverk som fortfarande gäller ska dock urvalet grundas på en totalpoäng.

Användningen av Högskoleprovet har ökat, eftersom det också används som första steget i särskilt urval och som särskiljande vid i övrigt lika meriter. Detta innebär i praktiken att Högskoleprovet blivit obligatoriskt vid ansökan till konkurrensutsatta utbildningar.

Den historiska översikten visar att det har skett omfattande förändringar av antagningssystemet över tid. Nya antagnings- och urvalsregler har införts ofta och snabbt, och allt oftare utan att föregås av ordentliga utredningar. Detta har resulterat i ogenomtänkta regler och behov av frekventa juste-

ringar som gjort antagningssystemet alltmer komplext, svårt att överblicka och ineffektivt när det gäller huvuduppgifterna: att reglera förkunskapskrav och göra urval av studerande som har de bästa förutsättningarna att nå goda studieresultat.

Nedan förs en kritisk diskussion om systemets enskilda komponenter, strukturerat utifrån de två huvudfunktionerna behörighet och urval. Inledningsvis ges dock några exempel på den vanligt förekommande sammanblandningen av systemets båda huvudfunktioner, behörighet och urval. Avslutningsvis ges en sammanfattande bild av identifierade brister och förslag till förbättringar av antagningssystemet.

77

Behörighet kontra urval

Att de sökande som bedömts vara behöriga för en viss högskoleutbildning också bedöms ha de grundläggande och särskilda förkunskaper som anses absolut nödvändiga för att tillgodogöra sig utbildningen kan tyckas självklart, likasom att urvalsreglerna endast reglerar vad som ska gälla om antalet behöriga sökande överstiger antalet utbildningsplatser. Likafullt sker ofta sammanblandningar av behörighet och urval.

Ett exempel är avsikten att sätta en nedre gräns för ett acceptabelt resultat på Högskoleprovet. Ursprungsproblemet förefaller vara att sökande till vissa utbildningar, exempelvis lärarutbildningar, inte bedöms ha tillräckliga förkunskaper, trots att de uppfyller behörighetskraven. Problemet orsakade av lågt söktryck kan dock inte lösas genom en nedre gräns för provresultat. Det finns heller ingen empirisk grund för en sådan gränssättning. Detta innebär också att man inför ett särskilt behörighetskrav, och det torde ge förvirrande signaler om en sökande som inte anses uppfylla behörighetskraven inom kvotgruppen för Högskoleprovet blir antagen inom betygsgruppen.

Ett annat exempel gäller utveckling av särskilda prov vid antagning till olika högskoleutbildningar, där ett problem är att det inte alltid är helt klart huruvida provet ska användas för behörighet eller för urval. Ett annat problem är att det ofta uppstår diskussioner om var gränsen för ett acceptabelt resultat ska dras även när det står klart att det är ett urvalsprov som ska utvecklas. En uttryckt nedre gräns för ett accepterat resultat är per definition att betrakta som behörighet och kan därför inte förekomma i urvalsprov.

Behörighet för antagning till högre utbildning

Behörig att antas till högre utbildning är den som bedöms ha förutsättningar att tillgodogöra sig utbildningen. Förutsättningarna kan enligt högskoleförordningen (7 kap.) ha förvärvats genom svensk eller utländsk utbildning, praktisk erfarenhet eller på annat sätt. Den sökandes förutsättningar ska bedömas just utifrån kravet att kunna tillgodogöra sig utbildningen, inte utifrån hur förutsättningarna har förvärvats. Förkunskapskrav till högre utbildning indelas i krav för grundläggande respektive särskild behörighet. Grundläggande behörighet avser de generella kunskaper och förmågor som krävs för samtliga högskoleutbildningar, medan särskild behörighet innebär krav på ytterligare kunskaper som bedöms vara absolut nödvändiga för att studenten ska kunna tillgodogöra sig den specifika utbildningen. Behörighetsbedömning kan göras utifrån den sökandes formella meriter, det vill säga avslutad utbildning på gymnasienivå med betyg i behörighetsgivande kurser, men också utifrån den sökandes reella kompetens, det vill säga att motsvarande kunskaper har förvärvats på annat sätt. När behörighet till högre utbildning anges ingår dock sällan en beskrivning av faktiska förkunskapskrav. Det som vanligen anges är särskilda kurser, eller utbildningsvägar, på gymnasienivå genom vilka den sökande förmodas ha tillägnat sig de förkunskaper som krävs.

GRUNDLÄGGANDE BEHÖRIGHET

Betyg är, och har sedan mycket länge varit, den huvudsakliga grunden för bedömning av behörighet till högre utbildning. De centralt reglerade mål- och kunskapsrelaterade betygen⁵ kan i princip ge information om kunskapsnivåer. De är därför särskilt användbara för att fastställa behörighetsnivåer och för att åstadkomma kontinuitet och progression mellan nivåer inom utbildningssystemet.

Universitets- och högskolerådets information om antagningsregler (Bedömningshandboken, för antagning till högskoleutbildning 2013–2014) anger olika sätt att styrka grundläggande behörighet. Informationen ges företrädesvis i form av dokumentation från formella utbildningar, men det framgår inte vilka faktiska kunskaper som krävs för att tillgodogöra sig utbildningen. Tydlig information om de faktiska kunskapsnivåer som bedöms nödvändiga för att den sökande ska kunna tillgodogöra sig högskoleutbildning är dock en nödvändig utgångspunkt för arbetet med att utveckla och kvalitetssäkra

⁵ Fortsättningsvis kommer benämningen målrelaterade betyg att användas.

former för bedömning av sökandes förutsättningar, och då särskilt om dessa har förvärvats genom praktisk erfarenhet eller på annat sätt.

Det har fortlöpande framhållits av företrädare för den högre utbildningen, liksom av Högskoleverkets utvärdering (Högskoleverket 2006), att studenternas förkunskaper har försämrats över tid och inte alltid motsvarar förväntade krav. Frågan är hur väl avmätt gränsen för de behörighetsgivande betygen är.

Bland sökande till högre utbildning finns studenter med olika skolbakgrund. För närvarande kan två stora grupper urskiljas: de som avgått från gymnasieskolan före år 1997 och som har normrelaterade ämnesbetyg från gymnasieskolans linjesystem, och de som har avgått från och med vårterminen 1997 och som har fyrgradiga målrelaterade kursbetyg från gymnasieskolans programsystem. Normrelaterade betyg varierar mellan betygen 1 och 5, där lägst betyget 3 i det behörighetsgivande ämnet satts som gräns för behörighet. I det målrelaterade betygssystemet har sifferbetygen ersatts av de fyra stegen Icke godkänt (IG), Godkänt (G), Väl godkänt (VG) och Mycket väl godkänt (MVG). Lägst betyget G är den gräns som satts för att den betygssatta kursen ska anses vara behörighetsgivande. Från och med höstterminen 2014 tillkommer ytterligare en grupp sökande, med betyg som satts efter den sexgradiga betygsskalan (A–F), där lägst betyget E kommer att utgöra gränsen för behörighet.

Thunberg, Filipsson och Cronhjort (2006) diskuterar mot bakgrund av rapporteringen om successivt försämrade kunskaper i matematik bland nyantagna studenter vid KTH, om betyget 3 i det gamla systemet och betyget G på den fyrgradiga skalan motsvarar jämförbara kunskaper. De menar att betyget G ofta bedöms innefatta även betyget 2 i det normrelaterade systemet, vilket skulle innebära en sänkning av behörighetskraven. Det är också oklart vilken kunskapsnivå betyget E på den sexgradiga skalan står för. Införandet av den mera differentierade betygsskalan skulle kunna erbjuda möjligheter till en nyansering av behörighetskraven.

Ett annat problem som är relaterat till behörighet är elevernas bristande möjligheter att läsa behörighetsgrundande kurser inom gymnasieskolans yrkesprogram. Inom 1994 års gymnasieskola gav samtliga program grundläggande behörighet till högskoleutbildning. Dessa förutsättningar återgick dock i stort till det gamla när 2011 års gymnasieskola infördes. På cirka hälften av yrkesprogrammen krävdes tills helt nyligen, att eleverna läste utökad program för att kunna nå grundläggande behörighet. Från och med läsåret 2013/14 har dock reglerna ändrats så att alla elever som valt ett yrkesprogram har rätt att läsa in grundläggande högskolebehörighet under utbildningens gång.

De krav som får ställas på särskild behörighet ska vara helt nödvändiga för att kunna tillgodogöra sig den sökta utbildningen (7 kap. Högskoleförordningen). De förkunskaper som krävs ska ha en direkt koppling till målen i de kurser/den högskoleutbildning som avses. Formulering av särskilda behörighetskrav bör därför idealt innebära att nödvändiga förkunskaper identifieras och formuleras, och därefter prövas mot kurser i gymnasieskolan, varefter det fastställs vilken eller vilka kurser i gymnasieskolan som bäst svarar mot kraven. Sökanden som tillägnat sig dessa kunskaper på annat sätt än genom angivna behörighetskurser ges därigenom en rimlig chans att återropa dessa kunskaper i sin ansökan. Lärosätet ges också rimligare möjligheter att bedöma dokument som styrker efterfrågade kunskaper, då kunskapskraven på förhand är kända. Lärosätet skulle härigenom kunna erbjuda prövning av återropade kunskaper och/eller utveckla samverkan med andra aktörer som kan erbjuda sådan prövning. Det ligger också ett värde i att kunna kommunicera förkunskapskrav med sökande som har formella meriter från utbildning i ett annat sammanhang, exempelvis från annat land eller studier som ligger långt bak i tiden.

Lustig (SOU 2004:29) poängterar att den information kraven förmedlar inte enbart riktar sig till presumtiva studenter, utan att den också på ett tydligt sätt ska återspegla kopplingen mellan gymnasieskolan och högskolan. Han menar vidare att det är ett gemensamt ansvar för de två utbildningsnivåerna att utforma behörighetskraven och att de ska ange högskoleutbildningens faktiska startnivå. Detta kräver en fungerande dialog mellan högskolan och gymnasieskolan så att högskolan har kännedom om de kunskaper och färdigheter gymnasieskolan faktiskt ger och att gymnasieskolan håller sig informerad om högskolans förväntningar på de behörighetsgivande kursernas innehåll och nivå. Bristande kommunikation mellan nivåerna har framförts som en av orsakerna till uttalanden om studenternas alltmer bristande förkunskaper. Thunberg, Filipsson och Cronhjort (2006) identifierade framför allt två slags diskrepanser mellan nivåerna, dels att ett antal stoffområden som traditionellt förutsätts som förkunskaper vid universitet och högskolor inte ingår i gymnasieskolans kursmål, dels att det finns en kulturklyfta mellan gymnasieskolan och högskolan i synen på vad som konstituerar matematikens kunskapsområde. Sammantaget menar de att det har skett förändringar av såväl behörighetskraven, i detta fall för civilingenjörutbildningarna, som i gymnasiets struktur, vilka har samverkat till en påtaglig sänkning av de särskilda behörighetskraven. Det är inte en orimlig tanke att liknande problem också går att finna inom andra ämnesområden.

Från läsåret 2013/14 är det möjligt att läsa in grundläggande behörighet inom gymnasieskolans samtliga yrkesprogram, inom ramen för det fria valet. Ett kvarstående problem är dock att regelverket även fortsättningsvis gör det mycket svårt för dessa elever att uppnå särskild behörighet. Den enda möjligheten är att läsa ett utökat program eller att i efterhand läsa nödvändiga kurser endera genom komplettering inom ramen för vuxenutbildning eller genom prövning. Betygskomplettering via komvux är dock inte ett alternativ för alla elever, eftersom kommunerna inte har någon skyldighet att erbjuda en sådan möjlighet. I avsikt att öka individens möjligheter att gå vidare till högre studier kommer dock regeringen att presentera ett lagförslag som innebär en rättighet till vuxenutbildning för alla som vill läsa in grundläggande och särskild behörighet till högskolan. Tanken är att lagförslaget ska börja gälla från januari 2017.

Värderingen av behörighetskompletterade betyg är dock inte likställd med värderingen av reguljära betyg, utan jämföras med konkurrenskompletterade betyg. Detta innebär att sökande med betyg från ett av gymnasieskolans yrkesprogram måste konkurrera i en särskild kvotgrupp för kompletterade betyg. Avsikten är att stävja strategiska konkurrenskompletteringar, vilket är rimligt eftersom en frekvent konkurrenskomplettering tenderar att medverka till betygsinflation, som i sin tur sänker betygens prognoskraft (Cliffordson, 2004a). Dock är det djupt orättvist att jämföra behörighetskomplettering med konkurrenskomplettering, då det drabbar helt olika grupper på samma sätt. Den ena gruppen består av högpresterande individer som handlar strategiskt för att bli antagna till prestigeutbildningar, medan individerna i den andra gruppen inte planerade för högskolestudier då de stod i begrepp att välja gymnasieprogram, men senare kommer på andra tankar.

Ytterligare ett problem förknippat med särskild behörighet är att högskolans utbildningar är inordnade i 17 områdesbehörigheter, vilka är svåra att förena med nödvändighetsprincipen, det vill säga att kunskapen ska vara absolut nödvändig för att studenten ska kunna tillgodogöra sig utbildningen. Precisionen vid val av vilka kurser på gymnasienivå som kan anges som behörighetsgivande, efter att behovet av förkunskaper analyserats utifrån lärandemålen i de kurser/den högskoleutbildning som avses, försämras av system med fasta kombinationer. Det har också visat sig att det inte är möjligt att på ett ändamålsenligt sätt inordna alla utbildningar inom de 17 områdesbehörigheterna.

Urval för antagning till högre utbildning

Den svenska högskolan har sedan slutet av 1970-talet begränsningar av antalet platser för i stort sett all högskoleutbildning. För de utbildningar där antalet sökande är större än antalet platser är det nödvändigt att göra ett urval. Eftersom individuella utbildnings- och yrkesmöjligheter bestäms av sådana urval, och eftersom utbildningsanordnarna har ett intresse av att anta de studenter som har bäst förutsättningar att framgångsrikt genomföra utbildningarna ställs mycket höga krav på kvaliteten i de system som används för att göra urvalet. Urvalsinstrumenten måste uppfylla högt ställda krav på objektivitet, reliabilitet och validitet, de måste ha bred acceptans och upplevas som rättvisa, och de får inte missgynna vissa grupper av sökande. De måste också vara praktiska och kostnadseffektiva, och de får inte ha negativa konsekvenser på andra delar av utbildningssystemet.

Betyg från gymnasieskolan var fram till år 1991 det dominerande urvalsinstrumentet, varefter betyg och Högskoleprov i princip varit likställda som urvalsinstrument. Genom att Högskoleprovet också alltmer kommit att användas som särskiljningskriterium vid lika betyg, och som första steg i särskilda urval med flerstegsförfarande, har dock provets betydelse alltmer kommit att öka.

I mindre omfattning används också särskilt urval, efter tillstånd från Universitets- och högskolerådet, samt alternativa urvalsmetoder inom ramen för den av högskolans styrelse beslutade urvalskvoten för antagning av maximalt en tredjedel av utbildningsplatserna till en utbildning. Nya urvalskvoter för Högskoleprov inklusive arbetslivserfarenhet, kompletterade betyg, folkhögskolebetyg och utländska betyg har introducerats över tid. Därtill kommer ytterligare en kvotgrupp att införas för de sexgradiga betygen. Med olika, mer eller mindre välövertänkta, argument har kvotgruppen för Högskoleprov inklusive arbetslivserfarenhet samt kvotgruppen för utländska betyg tagits bort. Tillsammans har denna utveckling och modifiering av regelverket inneburit att systemet har blivit oerhört komplext och i det närmaste ogenomträngligt.

I det följande diskuteras urvalssystemets två huvudkomponenter, betyg och Högskoleprov, utifrån aspekter med särskild relevans för urvalsfunktionen.

BETYG SOM URVALSINSTRUMENT

Urval baserat på betyg har använts i Sverige sedan början av 1900-talet. Genom att Sverige har en centralt reglerad gymnasieskola med gemensamma kursplaner, vilket inte är fallet i många länder, är förutsättningarna goda för

att använda betyg vid urval till högre utbildning. Trots att de målrelaterade betygen är förknippade med problem som variationer i betygssättning över tid (Cliffordson, 2004a; Vlachos, 2010) samt mellan lärare, skolor, kommuner och huvudmän (Fredriksson & Vlachos, 2011; Klapp Lekholm & Cliffordson, 2008; Wikström, 2005) finns det i dagsläget inget urvalsinstrument som har bättre prognosförmåga (t.ex. Atkinson & Geiser, 2009; Carroll, 1982; Cliffordson, 2004b, 2008), som ger bättre signaleffekter, eller är mera rättvist. Även om förutsättningarna är mycket goda finns det dock flera faktorer som påverkar betygens goda egenskaper i negativ riktning.

Bland dessa återfinns givetvis de brister i betygens likvärdighet som vi redan diskuterat i kapitel 2. Skillnader i generositet i betygssättning mellan olika lärare och skolor, men också skillnader i betygssättning över tid, det vill säga betygsinflation, skapar orättvisor och försämrar betygens prognosförmåga. Ett annat problem som uppmärksammats i anslutning till de målrelaterade betygen är att regelverket för att räkna fram ett sammanfattande betyg, det så kallade jämförelsetalet, kan medföra att de studerande och skolorna utvecklar strategier för att i första hand optimera jämförelsetalet, till förfång för verksamheten i gymnasieskolan, och för betygens jämförbarhet. Dock innebär skärpta regler för exempelvis behörighet till högre utbildning, som gäller från och med höstterminen 2010, en inskränkning i möjligheten att ”toppa” jämförelsetalet genom till exempel reducerat program eller att slutföra studierna inom komvux.

Som en konsekvens av förändringar i betygssystemets utformning har olika kvotgrupper införts, vilket i sin tur leder till ineffektivitet och orättvisor. Det är också möjligt att genom konkurrenskomplettering läsa upp betyg i det målrelaterade systemet, vilket bidrar till betygsinflationen och sänker betygens prognosförmåga. Ytterligare ett problem är att sökande sedan några år tillbaka har möjlighet att tillgodoräkna sig meritpoäng vid val av vissa kurser inom gymnasieskolan, vilket ger upphov till orättvisor mellan olika grupper av studerande. Det finns också goda skäl att anta att meritpoäng försämrar betygens prognosförmåga. Nedan behandlas de faktorer som är förknippade med de största problemen.

KVOTGRUPPSSYSTEMET

Betygen är inte jämförbara över olika betygssystem och betygsskalor. Detta problem har vanligen hanterats genom att sökande med betyg satta enligt olika principer förts till olika kvotgrupper, där antalet platser är proportionellt mot antalet behöriga sökande i gruppen. Ett problem med detta system är att antalet platser i vissa kvotgrupper kan bli mycket lågt eller

till och med noll. I synnerhet missgynnas sökande i nyinrättade kvotgrupper där det ännu inte skett något urval. Detta har sin grund i att de högst meriterade tenderar att söka omedelbart efter avslutad gymnasieutbildning, och dessa kommer då att konkurrera med varandra om ett litet antal platser. Varje förändring i betygssystemet leder till orättvisor, och försök att lösa detta med ytterligare kvotgrupper förvärrar snarare än löser problemet.

Detta kan åskådliggöras genom att studera andelen antagna till högre utbildning inom olika avgångskohorter från gymnasieskolan, över en tidsperiod då ett nytt betygssystem införts. Cliffordson och Gustafsson (2007) visade att andelen individer inom den första åldersgruppen med målrelaterade gymnasiebetyg, som antogs till högskolestudier inom tre år efter avslutat gymnasium, var märkbart mindre jämfört med kringliggande ålderskohorter. En förklaring till detta är just att de konkurrerade i en egen liten kvotgrupp. Samtidigt medför förändringar av betygen orättvisor även om de värderas i samma kvotgrupp. Det blir sannolikt betydligt svårare att erhålla högsta betyg inom ramen för den sexgradiga målrelaterade skalan än i den fyrgradiga. För att minimera orättvisor av det här slaget kommer de med betyg enligt den sexgradiga skalan under en treårsperiod att konkurrera både i en egen kvotgrupp och i den ordinarie betygskvoten. Ett annat, och förmodligen bättre, sätt att hantera problemet med de olika betygsskalorna, vore att ekvivalera skalorna så att betyg satta på olika skalor går att jämföra. Det skulle dock vara betydligt enklare att i möjligaste mån undvika förändringar av betygssystem eller antal skalsteg.

KOMPLETTERING AV BETYG

Regelverket kring de målrelaterade betygen tillåter så kallad konkurrenskomplettering, vilket innebär att betyg kan tenteras upp efter att gymnasiestudierna avslutats, antingen genom formella studier inom kommunal vuxenutbildning, eller efter självstudier och prövning. Detta utnyttjades flitigt av de sökande efter att det målrelaterade betygssystemet införts 1997 och omfattningen av studier på komvux åren efter avslutat gymnasium ökade kraftigt.

Forskare inom området (Gustafsson, 2000; Svensson, 1996; Wedman, 2000) uttryckte farhågor inför möjligheten att komplettera, vilka också till stora delar besannades. En farhåga var risken för betygsinflation, en annan att möjligheten att läsa upp betygen på komvux skulle leda till ett "fjärde gymnasieår" som inte bara medför ökade kostnader för samhället, utan också till orättvisor på grund av kommunala skillnader. En tredje farhåga var att unga sökande skulle konkurreras ut av dem som kompletterat. Den

sistnämnda effekten förefaller dock att åtminstone delvis ha neutraliserats av betygsinflationen.

Ytterligare en fråga som restes var om de gymnasiebetyg som höjts på detta sätt också svarar mot bättre studieresultat i den högre utbildningen, det vill säga hade motsvarande prognosförmåga som reguljära betyg. Förutom att vara resurskrävande, med en hög andel antagna som kompletterat sina betyg, visar forskningen att konkurrenskomplettering försämrar betygs prognosförmåga och ökar betygsinflationen (Cliffordson, 2004a). Sedan möjligheterna att konkurrenskomplettera betygen minskats och en särskild kvotgrupp införts för sökande med kompletterade betyg har omfattningen av konkurrenskomplettering minskat. Ett problem är dock, som vi redan pekat på, att antagningssystemet med dess nuvarande utformning av kvotgrupper inte skiljer på behörighetskomplettering och konkurrenskomplettering.

85

MERITPOÄNG

Sedan höstterminen 2010 finns ett allvarligt hot mot användningen av betyg som urvalsinstrument, nämligen de meritpoäng som gör det möjligt för sökande till högskolan att öka betygs meritvärde genom val av vissa kurser (till exempel främmande språk eller högre matematikkurser) inom gymnasieskolan. Meritpoängens införande har skapat stora och svårslösliga problem i urvalssystemet, sett utifrån såväl individ-, grupp- som systemperspektiv. Reglerna innebär att summan möjliga meritpoäng är avhängig den utbildning som ansökan till högskolan avser, vilket missgynnar elever som inte har tidiga planer att studera på högskolan, liksom elever som inte tidigt har bestämt vad de ska läsa på högskolan. En maximal utdelning av meritpoäng (2,5 p) kräver planering som kan sträcka sig ända ner till val av språk i årskurs 6 i grundskolan. En ytterligare källa till orättvisa är att möjligheten att välja kurser ser olika ut i olika skolor/kommuner, såväl i grundskolan som i gymnasieskolan, och det kan vara avgörande för elevernas möjligheter att kunna tillgodogöra sig meritpoäng. Därtill kommer olikheter i utrymmet för tillval av meritkurser mellan olika gymnasieprogram. Det är också orättvist för sökande med utländska betyg och sökande med äldre betyg, som inte haft möjlighet att tillskansa sig meritpoäng. Detta har skapat behov av ad hoc-betonade förändringar av meritvärderingsreglerna. Efter problematiska bedömningar och svårgenomträngliga beräkningar har vissa grupper tilldelats extra poäng i compensation för uteblivna meritpoäng eller retroaktiva extra poäng att jämföra med meritpoäng. Dessa nära nog årliga regeländringar som har genomförts för att i någon mån korrigera för

uppkomna orättvisor, innebär att olika regler gäller för sökande med betyg från olika kontexter och tidsperioder. Därutöver gäller ytterligare nya meritvärderingsregler för dem som läser gymnasieskolan enligt 2011 års gymnasieskola. Beräkningsgrunderna för meritpoängen är komplicerade och att de dessutom varierar över åldersgrupper gör att regelsystemet är svårt att förstå, svårt att överblicka och svårt att förhålla sig till vid planering av framtida studier.

De negativa konsekvenserna av att införa meritpoäng är omfattande också i andra avseenden. Ett allvarligt problem är att det finns goda skäl att anta att meritpoäng försämrar betygens prognosförmåga. Såväl svensk forskning (Cliffordson, 2004b, 2008) som en omfattande internationell forskning (t.ex. Atkinsson & Geiser, 2010), har visat att betyg från gymnasieskolan är det urvalsinstrument som bäst prognostiserar framgång i högre utbildning. Betygens prognoskraft ligger till stor del i de prestationer som krävs för att höja betyget ett eller flera steg på betygsskalan (t.ex. Cliffordson & Berndtsson, 2007). Det krävs en betydligt större prestationsförmåga att höja jämförelsetalet 2,5 poäng genom högre betyg i ett antal kurser än att välja meritkurser som ger 2,5 extra poäng enbart genom att prestera godkänt betyg i ett mindre antal kurser. Visad prestationsförmåga, i form av motivation, intresse och idoghet, är signifikanta egenskaper som bidrar till betygens prognoskraft, bortom tillägnade kunskaper och färdigheter (Thorsen & Cliffordson, 2012; Thorsen, 2013). Meritpoäng ger däremot inte motsvarande information om förväntade studieprestationer eftersom det räcker att prestera ett godkänt betyg på en meritkurs för att få extrapoäng. Detta innebär att urvalssystemets effektivitet försämras. Det vill säga att det blir mera osäkert om det är de sökande som har bäst förutsättningar att framgångsrikt genomföra sökt utbildning som antas. Eftersom det räcker att prestera ett godkänt betyg på en meritkurs för att få extrapoäng, leder reglerna också till negativa signaleffekter.

HÖGSKOLEPROVET

Genom att möjligheten att använda Högskoleprovet har utvidgats till samtliga sökande har det kommit att spela en annan roll i urvalssystemet än den ursprungligen tänkta, att bredda rekryteringen till högskolan. Det har blivit ett väletablerat urvalsinstrument som åtnjuter stort förtroende, det är lätt och snabbt att genomföra, och ger alla behöriga sökande en ytterligare chans att antas till konkurrensutsatta utbildningar. Dock är Högskoleprovet förknippat med flera problem.

Trots de dokumenterade bristerna i betygen har Högskoleprovet säm-

re prognosförmåga än gymnasiebetyg (Björklund m.fl., 2010; Cliffordson, 2004b, 2008). Regelverket medger upprepat provtagande och det bästa resultatet räknas, vilket bidrar till gruppskillnader i provresultat och till provets svaga prognosförmåga (Cliffordson, 2004b). Reglerna medger också att eleverna kan ta provet redan under gymnasietiden, vilket kan ge negativa signaleffekter på prestationer i gymnasieskolan. Ett ytterligare problem är att Högskoleprovets användningsområden successivt har utökats över åren, vilket innebär att det har fått en alltmer dominerande roll i urvalssystemet. Minst en tredjedel av platserna ska fördelas utifrån resultat på Högskoleprovet. Högskoleförordningen (7 kap.) ger möjlighet att använda Högskoleprovet som särskiljande instrument vid i övrigt lika meriter. Därutöver utgör ofta Högskoleprovet det första steget vid så kallat särskilt urval. Detta innebär att provet används i hela urvalet, vilket innebär att det i praktiken blir obligatoriskt för antagning till de mest konkurrensutsatta utbildningarna.

Hösten 2011 implementerades ett nytt Högskoleprov, med delvis annat innehåll och en annorlunda konstruktion.

KORT BESKRIVNING AV HÖGSKOLEPROVET

Förebilden för Högskoleprovet var det amerikanska "Scholastic Aptitude Test", numera benämnt SAT. Högskoleprovet ges två gånger per år och en provtagare kan genomföra provet vid vilken ålder som helst och hur många gånger som helst. Ett provresultat gällde initialt i två år, vilket senare kom att utökas till fem år. För den som har genomfört provet flera gånger räknas det bästa resultatet vid urvalet. Provet rättas så att varje korrekt svar ger en poäng, och den summerade råpoängen (se ordlista) överförs sedan till en normerad (se ordlista) poäng som kan variera från 0,0 till 2,0. Avsikten med normeringen är att erhållna provresultat ska vara jämförbara över tid även om provets svårighetsgrad varierar (Stage & Ögren, 2009). År 1982 infördes en avgift för att få genomföra provet.

Högskoleprovet har genomgått ett antal förändringar sedan den första versionen år 1977 (Stage & Ögren, 2001). Sedan 1996, och fram till höstterminen 2011, har provet bestått av fem delprov varav tre mäter verbal förmåga, ORD, LÅS och ELF, och två delprov som mäter kvantitativ/analytisk förmåga, NOG och DTK. Totalt bestod Högskoleprovet av 122 uppgifter, av vilka 80 mätte verbal förmåga och 42 mätte kvantitativ förmåga (Stage & Ögren, 2009).

Huvudinriktningen på utvecklingsarbetet av det nya provet, har varit att öka antalet kvantitativa uppgifter, så att dessa uppgår till samma antal som

de verbala uppgifterna. Detta har resulterat i att provet kan indelas i en verbal och en kvantitativ del om vardera 80 uppgifter, vilket innebär att antalet uppgifter totalt har ökat till 160 uppgifter. Inom den verbala uppgiftskategorin har ORD-uppgifterna minskat till förmån för en ny uppgiftstyp, MEK, där en eller flera luckor i en mening ska fyllas i genom val mellan fyra svarsalternativ. Två nya kvantitativa uppgiftstyper, benämnda XYZ och KVA, har också introducerats. Delprovet XYZ utgörs av mer renodlat matematiska uppgifter som motsvarar de kunskapskrav som gäller för gymnasieskolans A-kurs i matematik. Det andra delprovet, KVA, innebär att två matematiskt bestämda kvantiteter som ska jämföras med varandra presenteras, och där uppgiften är att fastställa vilken av kvantiteterna som är störst, om de är lika eller om informationen är otillräcklig för att kunna avgöra detta.

Resultatet på det nya Högskoleprovet uttrycks, på samma sätt som för det tidigare provet, i en normerad totalpoäng. I enlighet med Högskoleverkets direktiv inför revideringen av provet, beräknas också en normerad delpoäng för den verbala delen (ORD, LÄS, ELF & MEK) respektive den kvantitativa delen (NOG, DTK, XYZ & KVA) av provet, vilka är möjliga att vikta relativt utbildningens innehåll. Då regelverket för provets användning inte har reviderats i samma anda utnyttjas dock inte den fulla potentialen att höja provets prognosförmåga inom ramen för det reguljära urvalet. Användningen av delpoängen är endast möjlig genom att ytterligare utöka provets användning, exempelvis genom att vikta de båda delpoängen eller att använda det verbala eller det matematiska delprovet i ett särskilt eller alternativt urval, ensamt eller i kombination med andra urvalsinstrument. Detta är mycket anmärkningsvärt då en av utgångspunkterna för revideringen av provet var att just förbättra provets prognosförmåga genom att skapa möjligheten att på ett flexibelt sätt använda de två delpoängen riktat mot utbildningarnas innehållsliga karaktär.

Det nya provets prognosförmåga har inte studerats, medan en del är känt om gruppkillnader i provresultat (Lyrén & Ögren, 2013). I det följande redovisas resultat baserade på det gamla provet. I den mån motsvarande resultat finns för det nya provet redovisas de i direkt anslutning under respektive rubrik.

HÖGSKOLEPROVETS MÄTEGENSKAPER

I likhet med de utländska förebilderna (till exempel det amerikanska SAT) saknar Högskoleprovet en uttalad teoretisk grund. Ursprunget är dock forskningen om mätning av kognitiva (se ordlista) förmågor och prognos

av studie- och yrkesframgång. Carlstedt och Gustafsson (2005) undersökte hur Högskoleprovets olika delprov är relaterade till väletablerade kognitiva förmågedimensioner. Resultaten talar för slutsatsen att det i Högskoleprovet är möjligt att identifiera kognitiva dimensioner som är väletablerade i den psykologiska forskningen, men att variationen i provresultaten framför allt bestäms av verbala kunskaper och färdigheter.

GRUPPSKILLNADER

89

Högskoleprovet har alltsedan det introducerades varit behäftat med systematiska skillnader i resultat för olika grupper av provtagare. Män lyckas generellt sett bättre än kvinnor och provtagare med högre social bakgrund lyckas bättre än de med lägre social bakgrund. För de prov som gavs i början av 1990-talet var skillnaderna i provresultat något större för kön än för social bakgrund. På delprovsnivå var skillnaderna mellan olika socialgrupper tämligen stabila oberoende av delprov. För kön däremot fanns stora skillnader mellan delproven och den största skillnaden noterades för de två mera kvantitativt inriktade proven (NOG och DTK) och den minsta för det verbala ORD-provet (Stage, 2004, 2005). Åberg-Bengtsson (1999, 2005) identifierade en kvantitativ faktor som påverkar resultat på uppgifter som kräver beräkningar, likaså att den kvantitativa faktorn förklarar en stor del av könsskillnaderna till männens fördel. Inslaget av uppgifter som kräver räknemässiga manipulationer har reducerats i de prov som givits sedan 1996, vilket medfört en viss minskning av könsskillnaderna på de kvantitativa delarna av provet. Detta innebar att prestationsskillnaden på totalprovsnivå blev ungefär lika stor för kön och social bakgrund (Stage, 2004, 2005). För social bakgrund var skillnaden alltså ungefär lika stor för samtliga delprov. Huruvida utländsk bakgrund har betydelse för prestationer på provet har också studerats (Reuterberg & Hansen, 2001; Reuterberg, 2003). Resultaten visar att det finns skillnader till svenskfödda provtagares fördel, vilket talar för att Högskoleprovet kräver goda färdigheter i svenska språket och god förtrogenhet med den svenska kulturen. Sammanfattningsvis kan sägas att prestationer på det gamla Högskoleprovet innebar en fördel i konkurrensen om högskoleplatser för svenska män med högre social bakgrund, vilket ska visa sig gälla även för det nya provet. En bidragande orsak till dessa gruppskillnader är att benägenheten att ta provet flera gånger varierar mellan olika grupper.

Resultat baserade på de två första provomgångarna av det nya provet, som introducerats under 2011, pekar på att skillnaderna till männens fördel har ökat i jämförelse med det gamla provet, och att skillnaderna är större

på den kvantitativa delen än på den verbala (Lyrén & Ögren, 2013). Högre utbildning hos föräldrarna har samband med bättre provresultat, och skillnaderna är större för den kvantitativa delen än för den verbala. Vidare har också utländsk bakgrund betydelse även för det nya provet. Svenskfödda provdeltagare presterar bättre än utlandsfödda, och skillnaderna är större på den verbala delen än på den kvantitativa. Slutsatsen i rapporten är dock att gruppskillnaderna i provresultat för det nya provet inte är mer anmärkningsvärda än för det gamla provet.

Det är viktigt att notera att även gruppskillnader, som enligt traditionella värderingsmodeller är att betrakta som små, på ett dramatiskt sätt kan påverka möjligheterna för olika grupper att bli antagna till konkurrensutsatta högskoleutbildningar. Detta har sin grund i att genomslaget av en medelvärdesskillnad ökar ju lägre andel av de sökande som blir antagna (Reuterberg, 1994). Det är i detta sammanhang också värt att notera att högre social bakgrund/utbildningsbakgrund är förknippad med både högre betyg och högre provresultat. För kön gäller däremot att kvinnor har högre betyg medan män erhåller högre provresultat. Utlandsfödda provdeltagare rangordnas lägre av Högskoleprovet än av betyg.

UPPREPAD PROVTAGNING

Resultatet på Högskoleprovet påverkas starkt av upprepat provtagande, det vill säga provtagare presterar generellt bättre ju fler prov de tar (Björklund m.fl., 2010; Cliffordson, 2004c; Henriksson & Törnkvist, 2002; Svensson & Nielsen, 2005a; Törnkvist & Henriksson, 2004). Möjligheten att göra provet flera gånger utnyttjas tämligen flitigt (Stage & Ögren, 2009), även om benägenheten varierar mellan olika grupper. Män är mera benägna än kvinnor att upprepa provet och yngre upprepar provet flera gånger än äldre. Högpresterande personer är mer benägna än andra att ta provet i tidig ålder och att ta provet flera gånger. Personer med högre social respektive svensk bakgrund är också mera benägna att ta Högskoleprovet och de upprepar dessutom provet i större utsträckning (Björklund m.fl., 2010; Cliffordson, 2004c; Gustafsson, Andersson & Hansen, 2000; Henriksson & Törnkvist, 2002; Stage & Ögren, 2009; Reuterberg & Hansen, 2001; Reuterberg, 2003; Svensson & Nielsen, 2005a; Törnkvist & Henriksson, 2004).

För personer födda 1973–1982 var det minst lika vanligt med högsta poäng på Högskoleprovet som högsta betyg (Svensson & Nielsen, 2005). Cirka tre fjärdedelar av de provtagare som erhöll högsta poäng hade tagit provet flera gånger och knappt en tredjedel behövde fyra eller flera provtillfällen för att nå högsta poäng. Regeln att bästa resultat räknas innebär att de

slumpvariationer i resultat som alltid förekommer medför att endast provresultat med slumpeffekter i positiv riktning beaktas, vilket är en förklaring till att så många erhåller höga/högsta poäng efter upprepat provtagande.

Den samlade bilden visar att upprepat provtagande medför försämrade prognosförmåga och skapar orättvisor. Möjligheten utnyttjas flitigt av framför allt män, personer med högre social bakgrund och personer med svensk bakgrund. Dessa skillnader i benägenheten att upprepa provtagandet, i kombination med det starka sambandet mellan antal genomförda prov och provresultat, utgör viktiga förklaringar till gruppskillnaderna i provresultat.

STUDIEFÄRDIGHETSPROV ELLER KUNSKAPSPROV? EN INTERNATIONELL UTBLICK

Allmänna studiefärdighetsprov, av samma typ som Högskoleprovet, utan tydlig koppling till skolämnen förefaller att vara tämligen ovanliga i andra länder. Förutom i Sverige finns denna typ av prov i USA, Israel och Singapore (McDonald m.fl., 2001). Många länder har kunskapsprov kopplade till skolämnen, antingen som examinationsprov för inträde till högre utbildning eller som examensprov för att erhålla avgångsexamen, motsvarande vår gamla studentexamen.

Det amerikanska SAT-provet har utgjort förebild för Högskoleprovet. Eftersom utbildning, läroplaner och kursplaner i USA är ett delstatligt eller lokalt ansvar finns inga gemensamma examensprov utan examinationsformerna bestäms lokalt, vilket skapat behov av ett gemensamt prov. Det första provet var ett kunskapsprov, vilket senare kom att ändras till att pröva mera allmänna verbala och kvantitativa förmågor. Det kom därmed att i högre grad likna de intelligenstest som varit framgångsrika vid rekryteringen av militär personal under första världskriget. Provet kom under de följande åren att genomgå en rad förändringar (Lawrence et al., 2003). Inriktningen mot det som kallats en verbal del och en kvantitativ del har dock bibehållits från 1930 fram till idag, även om de olika delproven och uppgiftsformerna inom respektive förmågeområde har varierat. Genom uppdelningen kunde resultaten rapporteras som verbala respektive kvantitativa poäng och tilldelas olika vikt vid antagningen beroende på den sökta utbildningens karaktär.

SAT har sedan början 1990-talet förändrats i riktning mot en tydligare koppling till de förmågor och kunskapsområden som ingår i skolornas kursplaner. Provet har trots det kritiserats för att ha låg prognosförmåga i jämförelse med betygen från *high school*. I olika studier har SAT också visat sig öka snedrekryteringen i förhållande till betyg: vita män med hög social bakgrund gynnades mest medan kvinnor, provdeltagare med afro-

amerikansk eller latinamerikansk bakgrund samt deltagare från grupper med låg social bakgrund missgynnades (Geiser & Santelices, 2007; Geiser & Studley, 2002; Morgan, 1989; Rothstein, 2004), förhållanden som känns igen från det svenska Högskoleprovet.

Efter att University of California riktat allvarlig kritik mot SAT I (det generella provet) och förordat användning av SAT II, som utgörs av kunskapsprov i olika ämnen, ändrades SAT genom att ett essäprov lades till. Provet innefattar nu tre delar: kritiskt resonerande, matematik och essäskrivning (se College Board, 2010), och efter denna förändring har skillnaden i prognosförmåga gentemot betyg minskat (Kobrin m.fl., 2008). Om förändringen förmått minska de systematiska olikheterna i resultat som beror på kön, utländsk och social bakgrund är dock oklart än så länge.

En annan typ av kritik mot SAT är att det inte ger adekvata signaleffekter till dem som studerar på gymnasienivå, därför att det inte är kopplat till kursplaner och till den progression i lärande som det sammantagna skolsystemet bör främja. Trenden går mot en tilltagande kritik av generella antagningsprov och ett förordande av ämnesanpassade kunskapsprov. Ett problem i USA är dock att man saknar en gemensam nationell läroplan, vilket försvårar anknytningen av provet till *high school*. Sverige har genom sin gemensamma läroplan däremot goda förutsättningar att utnyttja gymnasiets prov- och betygssystem som grund för antagning och urval till högre utbildning, men har trots detta valt att i ökande utsträckning utnyttja Högskoleprovet för urval.

I USA är man mycket noga med att betona att inga väsentliga beslut bör fattas på basis av enskilda provresultat. Det framhålls att resultat från SAT och andra prov endast ska användas som komplement, eller i kombination med andra urvalsinstrument, som betyg. Samtidigt har Högskoleprovet i Sverige en starkare roll än vad SAT har i USA. Högskoleprovet används för det första i mycket stor utsträckning som det enda underlaget för urval. För det andra har Högskoleprovets betydelse efterhand ökat, medan utvecklingen varit den motsatta i USA.

BETYGENS OCH HÖGSKOLEPROVETS PROGNOSEFÖRMÅGA

Ett betydande antal studier, både svenska och internationella, har undersökt prognosförmågan hos såväl betyg som Högskoleprov. En undersökning har rapporterats av Cliffordson (2004b), där sambandet mellan betyg från de båda betygssystemen jämförts med resultat på Högskoleprovet. Resultaten visade att sambandet var 0,46 för de normrelaterade betygen, och 0,44 för de målrelaterade betygen. De praktiskt taget identiska sambanden pekar på

att de målrelaterade betygens förmåga att rangordna sökande inte är sämre än de normrelaterade betygens.

Ett annat sätt att undersöka betygens rangordnings- och prognosförmåga är att jämföra sambandet mellan betyg från gymnasieskolan och högskoleprovresultat med prestationer på högskolan. I samma studie visade resultaten att sambandet var högre för målrelaterade betyg (0,47) än för normrelaterade betyg (0,41) och att sambandet för Högskoleprovet var betydligt lägre (0,16 resp. 0,11). Detta talar för att målrelaterade betyg fungerar minst lika bra som urvalsinstrument som normrelaterade betyg, samt att betyg har betydligt mycket bättre prognosförmåga än Högskoleprovet. Eftersom denna studie endast fokuserar en högskoleutbildning (civilingenjörsutbildningen) inom ett område är risken uppenbar att resultaten inte är generaliserbara till andra högskoleutbildningar.

Cliffordson (2005) studerade därför också variationen i prognosförmåga för betyg och Högskoleprov över samtliga högskolans program med data från en åttaårsperiod. Resultaten visade på betydande skillnader i prognosförmåga för de olika urvalsinstrumenten och för vissa av dem en stor variation över programmen. De målrelaterade betygen hade positiv prognosförmåga för samtliga program, medan den var lika med noll för drygt fyra procent av programmen för de normrelaterade betygen. För Högskoleprovet var prognosförmågan lika med noll för en fjärdedel av programmen. Skillnaden i graden av prognosförmåga mellan olika ämnesområden visade sig vara betydligt större för Högskoleprovet än för betyg. Sammantaget kan sägas att resultaten talar för att de målrelaterade betygen hade något bättre prognosförmåga än de normrelaterade betygen, medan Högskoleprovet hade sämst prognosförmåga, samt att de målrelaterade betygen hade hög och jämn prognosförmåga över samtliga program. I den mån det förelåg några skillnader var prognosförmågan högre för teknisk-naturvetenskapliga program än för övriga program. Högskoleprovet hade sin bästa prognosförmåga för program inom de humanistiska och samhällsvetenskapliga områdena, medan det motsatta gällde för normrelaterade betyg. Björklund med flera (2010) visade i en studie av sannolikheten att fullgöra kraven för en kandidatexamen, eller motsvarande fem år efter antagning, att Högskoleprovet inte gav något bidrag till prognosen utöver betyg, medan betyg gav ett starkt prognosbidrag. Ovanstående presenterade resultat sammanfaller med de som framträder i aktuell internationell forskning kring motsvarande urvalsinstrument.

Resultat från andra studier (Cliffordson & Askling, 2006; Svensson, 2004; Svensson & Nielsen, 2005b) har visat att avbrott och uppehåll i studi-

erna på bland annat läkar-, jurist- och civilingenjörsutbildningarna är vanligare bland dem som är antagna via Högskoleprov än bland betygsantagna, samt att betygsantagna har bättre studieresultat än högskoleprovantagna så till vida att de producerar fler poäng och oftare avlägger examen inom rimlig tid. Skillnaderna är störst för civilingenjörsutbildningarna och minst för juristutbildningarna.

Sammanfattningsvis visar all tillgänglig forskning att Högskoleprovet har sämre prognosförmåga än betyg, oavsett betygssystem, och att Högskoleprovets prognosförmåga varierar kraftigt för olika program/ämnesområden.

PROGNOSFÖRMÅGA FÖR SÄRSKILT URVAL

Vid urval till grundläggande högskoleutbildning används huvudsakligen betyg och Högskoleprov men till vissa utbildningar, till exempel vissa av landets läkarutbildningar, används även antagning via särskilt urval. Sådana urval är vanligtvis upplagda som flerstegsurval, med den största selektionen i det första steget, för vilket Högskoleprovet vanligtvis används. Nästa steg innebär oftast ett skriftligt prov och/eller en intervju (Cliffordson, 2006). Resultat från en studie (Cliffordson & Askling, 2006), där studieframgång för antagna via särskilt urval till läkarutbildningarna jämförs med antagna via betyg och högskoleprovresultat, visar att i stort samma förhållanden gällde för särskilt urval och betyg. Dock med den skillnaden att omfattningen av uppehåll och avbrott var något lägre än för betygsantagna. Resultaten påverkas dock av om betyg eller Högskoleprov används i det första steget.

BETYGENS OCH HÖGSKOLEPROVETS SIGNALEFFEKTER SAMT KONTINUITET ÖVER UTBILDNINGSSYSTEMET

Under senare år har graden av kontinuitet mellan nivåerna i utbildningssystemet (Atkinson & Geiser, 2010) betonats som betydelsefull i flera avseenden. I antagningssammanhang gäller det bland annat hur behörighetskraven till högre utbildning preciseras och kommuniceras och hur väl förkunskapskraven överensstämmer med innehåll och nivå i gymnasieskolans behörighetsgivande kurser. Som Lustig (SOU 2004:29) poängterar ska den information förkunskapskraven förmedlar vara ett uttryck för en tydlig kontinuitet mellan gymnasieskolan och högskolan.

Högskoleprovet är, som tidigare nämnts, ett allmänt studiefärdighetsprov utan en explicit teoribas. När det gäller frågan om kontinuitet över utbildningssystemet är det svårt att se Högskoleprovets koppling till gymnasieskolans och högskolans kunskapsinnehåll, med undantag för det nya provets koppling till matematik i gymnasieskolan. Avsaknaden av en sådan

koppling gör att Högskoleprovet ger negativa signaleffekter på prestationer i gymnasieskolan. Exempelvis finns det ingen nedre åldersgräns för när provet får göras och det bästa resultatet från de senaste fem åren får räknas. Detta innebär att om en elev får ett tillfredsställande resultat på Högskoleprovet redan under gymnasieåren kan motivationen för att anstränga sig i gymnasieskolan sjunka, eftersom ett högt resultat på Högskoleprovet räcker för att vara garanterad en studieplats.

Högskoleprovet har heller ingen koppling till kunskapskrav i högskolan. Högskoleprovet kan därmed inte sägas ingå på något naturligt sätt i en utbildningsgång från gymnasieskola till högskola. Tvärtom innebär det att sökande med betydligt sämre förkunskaper kan "gå före" sökande med bättre förkunskaper, eftersom de enda förkunskaper som behövs för sökande med högskoleprovsresultat är de som föreskrivs som grundläggande och eventuella särskilda behörighetskrav. Det faktum att det är möjligt att ta provet redan under gymnasiet försvagar också dess roll att erbjuda en "andra chans".

Av de få länder som använder allmänna studiefärdighetsprov är det framför allt i USA som frågor om signaleffekter och kontinuitet mellan olika delar av utbildningssystemet diskuteras. I Sverige förs nästan ingen debatt om positiva respektive negativa signaleffekter kopplade till betyg respektive Högskoleprov som urvalsinstrument. I den debatt som förs i USA har flera organisationer inom utbildningsområdet starkt signalerat ämneskunskapernas betydelse liksom vikten av att ha en gemensam standard för ämnets innehåll och de förmågor undervisningen ska utveckla, för att det ska vara möjligt att använda betyg som urvalsinstrument till högre studier. Dessa förutsättningar finns inom det svenska utbildningssystemet men saknas i USA.

Sammanfattning och slutsatser

Ett väl fungerande antagningssystem måste ha effektiva huvudfunktioner. Det innebär att behörighetskraven måste vara funktionellt utformade och att urvalssystemet måste vara konstruerat så att en god prognosförmåga uppnås. Rättvisa, rättssäkerhet, positiva signaleffekter och stöd för kontinuitet mellan nivåerna i utbildningssystemet är andra väsentliga egenskaper att ta hänsyn till när man ska utvärdera och modifiera antagningssystemet till högre utbildning.

Innan vi diskuterar konkreta problem och förslag till hur systemet kan förändras måste vi framhålla att de brister i det svenska antagningssyste-

met som framkommit i vår genomgång är så pass omfattande, i såväl stort som smått, att det är nödvändigt att genomföra en total översyn av hela systemet.

BRISTER

Det föreligger en utbredd bristande tilltro till att de sökande som bedömts vara behöriga också har de förkunskaper som anses nödvändiga för att tillgodogöra sig utbildningen. Detta har naturligtvis flera orsaker, varav en kan vara att behörighetsgränsen för betygen i de behörighetsgivande kurserna, G för de fyrgradiga målrelaterade betygen och E för de sexgradiga betygen, inte är optimala. Bristande likvärdighet i betygssättning kan vara ytterligare en orsak. En annan att dialogen brister mellan gymnasieskolan och högskolan, när det gäller förväntningar på förkunskaper och på gymnasieskolans kursinnehåll. Ytterligare ett problem är att inordnandet av förkunskapskrav för särskild behörighet i på förhand definierade områdesbehörigheter är svår att förena med den precision och konkretion av förkunskapskrav som skulle kunna åstadkommas genom analys av högskolans kurs-/utbildningsmål.

Ett problem som tangerar såväl behörighet som urval är att komplettering av kurser för behörighet inte värderas lika som ordinarie kurser i gymnasieskolan. Detta skapar orättvisor eftersom personer som inte tänkt "rätt" redan vid valet av gymnasieprogram straffas genom att vid ansökan till högskolan få konkurrera i en mindre grupp för kompletterade betyg.

Det system som reglerar urval till högre utbildning generellt och de urvalsinstrument som används för urval till högskolans utbildningar betonar i alltför liten grad betydelsen av tillägnade kunskaper som grunden för urvalssystemet, vilket är av betydelse för såväl prognosförmåga, rättvisa, signaleffekter som kontinuitet över utbildningsnivåer.

Det är väl belagt i såväl svensk som internationell prognosforskning att betyg är överlägsna Högskoleprovet som urvalsinstrument. Betygen har förvisso brister då de inte är likvärdiga vare sig över tid eller över lärare och skolor. Både likvärdighetsproblematiken och införandet av meritpoäng, som kanske är det allvarligaste problemet, inverkar menligt på betygens prognosförmåga och skapar orättvisor. Det sistnämnda gäller också införandet av olika kvotgrupper för betyg satta inom olika system och med olika skalor. Möjligheten att tillskansa sig meritpoäng innebär dessutom, utöver försämrade prognosförmåga och orättvisor, negativa signaleffekter.

Högskoleprovet, som har fått en alltmer framträdande plats i urvalssystemet över tid, har en svag prognosförmåga. Huruvida provets prognosförmåga har förbättrats med den reviderade versionen har inte studerats.

Studier baserade på dess amerikanska förebild, SAT, pekar på att införande av uppsatsprov kan bidra till att höja prognosförmågan, men en utveckling av provet i denna riktning är knappast aktuell. Det regelverk som omgärdar provet, såsom möjligheten att upprepa provet hur många gånger som helst och att det bästa resultatet räknas, bidrar till de konstaterade gruppskillnaderna i provresultat. Att det är tillåtet att ta provet redan under gymnasietiden kan förstärka de negativa signaleffekter som provet är behäftat med. Möjligheten att ta provet under gymnasietiden innebär också att det inte utgör en äkta "andra chans", vilket var den ursprungliga tanken. Provets innehåll har få kopplingar till gymnasieskolans innehåll och högskolans förkunskapskrav, vilket innebär att den eftersträvaransvärda kontinuiteten över utbildningsnivåerna saknas. Högskoleprovets negativa effekter överensstämmer väl med resultat från forskning kring dess förebild, det amerikanska SAT-provet.

Ovanstående beskrivna problembild kan förefalla svår att komma till rätta med. Dock kommer vi under nästa rubrik att presentera möjliga förslag som täcker in det mesta av den beskrivna problematiken.

FÖRSLAG TILL FÖRBÄTTRINGAR

Problemen med betygens bristande likvärdighet är sannolikt fullt möjliga att lösa med hjälp av förankring i de nationella proven. Vi kommer att utveckla detta närmare i rapportens avslutande kapitel där vi också pekar på att det är möjligt att med hjälp av nationella prov knyta de olika kursernas betygssteg till innehållsligt definierade nivåer. Detta skulle i sin tur göra det möjligt att relatera kunskaper och färdigheter från gymnasiet till de kunskapskrav som ställs av högskolan. Förkunskapskraven borde vara ett resultat av analyser av högskoleutbildningarnas innehåll så att de kan presenteras i form av de kunskapsnivåer som är nödvändiga för att studenterna ska kunna tillgodogöra sig utbildningen. På detta sätt skulle kraven på ett realistiskt sätt kunna anpassas till respektive utbildning och de trubbigt utformade förkunskapskraven definierade i form av områdesbehörigheter skulle kunna avskaffas. Med hjälp av mer ingående beskrivningar av vilka kunskaper och färdigheter som krävs är det också möjligt att på ett bättre sätt avgöra hur de svarar mot olika kurser och betygsnivåer i gymnasieskolan. Införandet av de sexgradiga betygen gör det också möjligt att överväga anpassning av betygskraven för de behörighetsgivande kurserna.

Även kvotgruppsproblemet skulle eventuellt vara möjligt att lösa genom ekvivalering av betyg satta utifrån olika principer. En betydligt enklare och mera framkomlig väg är dock att i möjligaste mån undvika förändringar av

betygen. Behörighets- och konkurrenskompletterade betyg bör inte föras till samma kvotgrupp.

När det gäller meritpoäng framhåller ansvariga politiker att det har varit framgångsrikt, med hänvisning till att det har påverkat elevernas val av gymnasiekurser i önskad riktning. Eftersom meritpoängssystemet är destruktivt vad gäller antagningssystemets samtliga kvalitetsaspekter bör det dock avskaffas.

98

Högskoleprovet borde få en mindre framträdande, och något annorlunda, roll som urvalsinstrument. Antalet platser som besätts via Högskoleprovet bör minska kraftigt, medan antalet platser som besätts med betygssökande bör ökas i motsvarande omfattning. Det är också angeläget att så mycket som möjligt komma bort från användningen av Högskoleprovet som särskiljningskriterium och som första steg vid särskilt urval. Regelverket bör förändras så att provet kan användas mer flexibelt. Detta bör dock ske inom ramen för den minskade kvoten. Prognosförmågan skulle sannolikt förbättras om provets båda delpoäng kunde viktas eller om antingen den ena eller den andra delpoängen kunde användas beroende på utbildningens ämnesinriktning. I avsikt att minska provets negativa signaleffekter borde det inte vara möjligt att genomföra provet före avslutad gymnasieutbildning. På så vis skulle provet utgöra en äkta "andra chans", vilket inte är fallet idag.

Vår huvudslutsats är att antagningssystemet bör fokusera på kunskaper och färdigheter snarare än på generella förmågor utan direkt koppling till vare sig gymnasieskolans eller högskolans kunskapsinnehåll. Vi har i Sverige också utmärkta förutsättningar att göra detta tack vare våra gemensamma och förhållandevis genomarbetade läro- och kursplaner.

5. Sammanfattande diskussion och förslag

Vi inleder det avslutande kapitlet med att sammanfatta de problem vi identifierat i de olika delarna av den svenska skolans system för kunskapsbedömning och uppföljning, varefter vi diskuterar förslag kring hur dessa problem skulle kunna lösas. Därefter följer en diskussion om hur antagningssystemet till högre utbildning bör reformeras.

Kunskapsbedömning på individ- och systemnivå

Det svenska systemet för bedömning av elevers kunskaper och för uppföljning av skolans kunskapsresultat skulle kunna beskrivas som en koloss på lerbeton. Systemet för summativ bedömning på elevnivå med fokus på betyg och nationella prov i grundskolan och gymnasiet har under senare år byggts ut till att omfatta betyg och nationella prov även i årskurs 6 liksom nationella prov i fler ämnen. De mycket omfattande data som detta system genererar samlas kontinuerligt in och används som underlag för rapportering om uppnådda resultat på nationell nivå och för grupper av kommuner. Data, och då framför allt betygsdata, görs också tillgängliga i databaser för analys och redovisning av resultat på skolnivå och som underlag för bland annat Skolinspektionens granskningar. Men utöver dessa data är det sparsamt med underlag kring den svenska skolans resultatutveckling och vad som påverkar denna. De internationella studierna är visserligen en viktig informationskälla, men denna information är mycket partiell i det att den endast omfattar ett begränsat antal ämnen och årskurser där endast vissa typer av utfall mäts.

Betygsinformationen är behäftad med bristande likvärdighet mellan lärare och skolor, och av betygsinflation. Den senare gör att förändringar i betygsmedeltal inte ger någon pålitlig information om förändringar i kunskapsnivåer. Den bristande likvärdigheten mellan lärare och skolor har bland annat sin grund i att det inom många ämnen är svårt att nå tillräcklig

samsyn mellan olika lärare i tolkningen av de kunskapskrav som kursplanen anger för de olika betygsstegen.

I en del av ämnena/kurserna finns nationella prov vars huvuduppgift är att stödja en likvärdig betygssättning. Av flera anledningar ger proven dock inte tillräckligt stöd. En är att prov endast finns i vissa ämnen och årskurser, och en annan är att det inte är uttalat i vilken utsträckning, och på vilket sätt, proven ska påverka betygssättningen. Andra anledningar är att proven inte är utformade för att vara jämförbara över tid, och att vissa prov brister i tillförlitlighet, bland annat på grund av att de består av ett begränsat antal större och komplexa uppgifter. Detta gör det svårt att generalisera och svårt att nå samstämmighet i bedömning av elevernas svar. För att betygen ska kunna uppfylla kraven på likvärdighet behöver dessa problem lösas.

Betygen har bland annat den mycket viktiga uppgiften att ge underlag för beslut om antagning till nästa nivå i utbildningssystemet. Forskning visar att gymnasiebetygen har bättre prognosförmåga än Högskoleprovet och liknande instrument, detta trots betygens likvärdighetsproblem. Om problemen med betygsinflation och skillnader i betygssättning mellan lärare och skolor kunde bemästras, skulle dock betygens effektivitet och rättvisa som urvalsinstrument förbättras väsentligt.

Betyg ger endast allmän information om de kunskaper och färdigheter som kan förväntas på olika betygsnivåer, vilket försvårar användning av betygen som bas för regler om allmän och särskild behörighet för eftergymnasial utbildning. Om de olika betygsstegen tydligare kunde förankras innehållsligt i mål och kunskapskrav skulle olika utbildningars krav på förkunskaper lättare kunna uttryckas i termer av uppnådda betygssteg inom olika kurser.

Vi har också konstaterat att Sverige saknar ett kunskapsbedömnings-system som på nationell nivå ger information om den långsiktiga utvecklingen av kunskaper och färdigheter, och om hur denna utveckling gestaltar sig för olika grupper av elever.

Det är inte möjligt att i detta sammanhang presentera detaljerade modeller för hur de ovan identifierade problemen skulle kunna lösas. Vi vill ändå ge förhållandevis konkreta förslag om riktlinjer för hur ett integrerat system för kunskapsbedömning i och av den svenska skolan kan utformas.

KUNSKAPSBEDÖMNING PÅ SYSTEMNIVÅ

Den komponent som framstår som den lättaste att hitta lösningar för är kunskapsbedömning på systemnivå. Ett relativt omfattande översikts- och utredningsarbete kring nationell kunskapsutvärdering har genomförts. Skolverket (2006) har på grundval av detta beskrivit ett förenklat förslag till

hur kunskapsbedömning på nationell nivå inom ett begränsat antal ämnesområden skulle kunna utformas.

Vi föreslår i linje med detta att ett rullande, stickprovsbaserat, kunskapsbedömningssystem baserat både på en skalorienterad ansats och på en uppgiftsorienterad ansats successivt byggs upp. Inom samtliga ämnen finns det sannolikt områden som lämpar sig bättre för antingen den ena eller den andra ansatsen, och de två ansatserna kan beskrivas som komplementära, och som delar av en helhet.

Som vi beskrivit i kapitel 3 redovisas resultaten i nationella och internationella kunskapsbedömningssystem ofta i form av uppnådda nivåer på kompetensskalor där de olika nivåerna beskrivs i termer av kunskaper och färdigheter. I den skalorienterade ansatsen är det angeläget att sådana kompetensskalor utvecklas och används, eftersom resultaten blir lättare att tolka och att använda.

Det är också angeläget att utveckla den kvalitativt inriktade uppgiftsbaserade ansatsen, byggd på mer omfattande och komplexa uppgifter, som en viktig del av det nationella kunskapsbedömningssystemet. Som vi pekat på tidigare är det viktigt att vissa av dessa uppgifter släpps fria för att det ska vara möjligt att kommunicera och tolka resultaten. Alla landets skolor skulle då också kunna använda uppgifterna i den egna verksamheten. Dessa väldokumenterade uppgifter med omfattande empirisk underbyggnad skulle kunna bli viktiga hjälpmedel i skolornas summativa utvärderingsarbete och betygssättning, men kanske framför allt i den formativa bedömningen i direkt anslutning till undervisningen.

Diskussionen kring nationella kunskapsbedömningssystem har hittills i första hand avsett grundskolan, men även för gymnasieskolan finns behov av det slags information som ett nationellt kunskapsbedömningssystem kan bidra med. Även inom gymnasieskolan är det rimligt att ha ett ämnesfokus, vilket torde vara möjligt också inom den kursutformade gymnasieskolan där läro- och kursplaner beskriver mål både på ämnes- och kursnivå.

Vi anser att det är angeläget att påbörja uppbyggnaden av ett nationellt kunskapsutvärderingssystem av det slag som skisserats ovan. Ett sådant system bidrar med angelägen information om utvecklingen av nivån på kunskaper och färdigheter inom olika områden och för olika grupper av elever. Genom insamling av kompletterande information från olika aktörer i skolan kan det också ge grund för analyser av möjliga förklaringar till den observerade utvecklingen. Men som vi beskrivit finns det också intressanta möjligheter att mer konkret knyta samman det nationella kunskapsutvärderingssystemet med det nationella provsystemet.

Betygen ska vara likvärdiga och en lång rad verktyg erbjuds för att stödja betygssättningen. Möjligheterna att åstadkomma likvärdighet genom olika former av bedömningsstöd, verbala uttolkningar av mål och kunskapskrav, och utbildningsinsatser av skilda slag är dock begränsade. Mer kraftfulla hjälpmedel behövs därför, och här finns en potential i de nationella proven.

102

De nationella provens utformning

De nationella proven har två huvudsakliga syften: att stödja en likvärdig och rättvis bedömning och att ge underlag för en analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, huvudmannanivå och nationell nivå. De översikter vi presenterat i kapitel 2 visar att de nationella proven med sin nuvarande utformning inte fullt ut förmår att uppfylla dessa syften, vilket har sin grund dels i att proven inte sällan har reliabilitetsbrister, dels i att de inte är jämförbara över tid.

Diskussionerna i kapitel 2 och 3 om de nationella provens framtida uppgifter och utformning kan sammanfattas på följande sätt: det nationella provsystemet bör även fortsättningsvis ha ambitionen att fånga också läroplansmål som avser svårbedömda kompetenser, som skrivförmåga, muntlig förmåga och förmåga till samarbete. Dock bör åtgärder vidtas för att höja graden av samstämmighet mellan olika lärares bedömningar. Vidare bör man analysera när det är ändamålsenligt att använda autentiska uppgifter och när det är lämpligare att använda prov som är uppdelade i ett större antal mer avgränsade uppgifter.

Det nationella provsystemet bör också ha ambitionen att stödja likvärdig och över tid stabil betygssättning. Provsystemet bör därför också inkludera delprov som är tillräckligt tillförlitliga och stabila för att kunna ge ett adekvat stöd för likvärdig betygssättning. Vi har också i kapitel 3 föreslagit att det nationella kunskapsbedömningssystemets skalorienterade ansats kan användas som en grund för att utveckla sådana nationella prov. Utgångspunkten är uppgifter som med hjälp av matrissamplingsteknik länkats samman, och som fortfarande hålls hemliga. Att uppgifterna är länkade innebär att deras svårighetsgrad och övriga karakteristika är kända. Detta gör det möjligt att på förhand beräkna provets svårighetsgrad, och de poänggränser för olika betygssteg som innebär samma kravnivå som för det föregående provet.

Detta arbetssätt förutsätter att det finns tillräckligt många uppgifter i det nationella kunskapsbedömningssystemet för att det ska kunna förse de årliga nationella proven med uppgifter. Nationella prov som är reliabla och

stabila över tid kan dock utvecklas utan att vara knutna till det nationella kunskapsbedömningssystemet, i synnerhet om det där inte finns tillgång till tillräckligt många uppgifter. För detta krävs sannolikt dock likartade tekniker, med en större mängd länkade uppgifter som hålls hemliga innan användningen i det nationella provet.

Den primära uppgiften för de nationella proven är att stödja likvärdig betygssättning. Det finns därför anledning att diskutera hur betygsstödet praktiskt kan utformas. Varje ämnesprov/kursprov kommer att utgöras av flera olika skalor och för var och en av dessa kommer ett betyg att ges. Vidare kommer de provdelar som omfattar autentiska uppgifter, och som inte ingår i skalorna, att betygsättas. Denna information behöver aggregeras till ett sammanfattande provbetyg, vilket bör ske enligt gemensamma, explicit formulerade, principer. Vi förutsätter att lärarens uppdrag att fastställa elevens slutbetyg/kursbetyg på grundval av såväl provbetyget som övrig relevant information som inhämtats om elevens prestationer kommer att kvarstå oförändrat.

Samtidigt är det nödvändigt att de provdelar som ska stödja likvärdig och stabil betygssättning utövar en normerande funktion för betygssättningen i sin helhet. Den exakta formen för hur detta ska genomföras återstår ännu att avgöra men det framstår som rimligt att avvikelserna mellan de normerande provdelarna och slutbetyget på klass-/skolnivå ska hålla sig inom vissa definierade gränser.

Om de nationella proven på detta sätt mer direkt kommer att påverka betygssättningen blir proven mer *high-stakes* än de är idag, och det finns betydande risker med detta. En risk är att det som mäts av de normerande delarna av provsystemet kommer att bli föremål för uppmärksamhet i undervisningen i större utsträckning än det som mäts av de övriga delarna av provsystemet. Vidare är det rimligt att förvänta sig att provuppgifter som ingår i det nationella kunskapsbedömningssystemet kan bli intressanta att kopiera och sprida exempelvis via sociala medier, vilket i så fall skulle göra dessa uppgifter oanvändbara i de nationella proven. Hur dessa risker med ökad *high-stakes* runt de nationella proven, eller delar av dem, ska kunna bemästras bör bli föremål för utredning i en brett sammansatt grupp, bland annat med medverkan av elever. Det förefaller väsentligt att resonemangen knyts till ett rättvise- och likvärdighetsperspektiv och inte reduceras till en fråga enbart om mättekniker.

Informationen från de nationella proven kan förväntas vara av intresse för uppföljnings- och forskningsändamål, och den bör därför samlas in. Även om det är nödvändigt att aggregera all provinformation till ett sam-

manfattande provbetyg är det dock viktigt att insamlingen av provresultat inte begränsas till det sammanfattande provbetyget, och inte heller till de olika delbetygen. Som visats av Koretz (2008) är prestationsmått som sammanfattas i så grova kategorier som fyra till sex betygssteg knappast användbara för uppföljningsändamål, eftersom prestationsförändringar i allmänhet är relativt små. Det är därför nödvändigt att göra storskaliga insamlingar även av de poängtal som eleverna når på de olika uppgifterna och delproven.

Men även om vi anser att det är angeläget att resultaten på de nationella proven samlas in på detaljerad nivå är vi av den uppfattningen att denna information inte bör redovisas på lägre nivå än på nationell eller möjligen kommunal nivå. Den modell vi föreslår skulle i princip göra det möjligt att använda informationen i det nationella provsystemet för redovisning på skolnivå, och i synnerhet som vissa av delproven är länkade över tid skulle det kunna ses som attraktivt att använda informationen för ansvarsutkrävande ändamål på skolnivå. Detta skulle dock få som konsekvens att de nationella proven skulle bli än mer *high-stakes* för skolan än vad som följer av den betygsnormerande funktionen. Detta skulle förstärka de ovan beskrivna riskerna för anpassning av undervisningen efter proven, vilket skulle kunna innebära snedvridning av undervisningen, och det skulle också medföra försämringar av provsystemets förmåga att fungera som indikator på kunskapsnivån. Det är viktigt att inse att det främsta syftet med dessa delprov är att stödja en likvärdig betygssättning och om de används för andra syften kommer detta att deformera provens förmåga att uppnå det primära syftet. Även denna aspekt bör utredas vidare.

Kan provinformation användas formativt?

Sedan en omfattande forskning visat att formativ återkoppling har en betydande potential att förbättra undervisningskvaliteten och elevernas resultat har ett stort intresse riktats mot möjligheten att utvinna formativ information ur prov som utvecklats för summativa ändamål. Forskningen indikerar att det finns vissa sådana möjligheter, men också att det finns anledning till försiktighet, eftersom det finns risk att åsamka skador i både den formativa och den summativa funktionen hos proven (se OECD, 2013, för en översikt). Grundproblemet är att den formativa användningen i allmänhet har andra informationsbehov och genomförs i andra sammanhang än den summativa provanvändningen. I formativ bedömning är syftet att identifiera missuppfattningar och luckor i elevens lärande för att det ska vara möjligt att ge en detaljerad återkoppling som kan influera det fortsatta lärandet. Sådan bedömning görs med fördel i nära anslutning till undervisningen och

i situationer som är befriade från kvalitetsgraderingar av prestationer. Summativa bedömningar i form av nationella prov eller nationella kunskapsbedömningar ger i allmänhet inte tillräckligt detaljerad information för att den ska kunna användas för att ge konstruktiv återkoppling, vilket beror på att provens primära syfte inte är att stödja formativ användning. Ett annat problem är att de summativa bedömningarna ofta genomförs vid tidpunkter som inte ger utrymme för användning av informationen i undervisning, som i slutet av läsåret eller stadiet. Det går också ofta lång tid mellan genomförandet av bedömningen och att eleverna faktiskt får återkoppling, om detta överhuvudtaget sker. Ett ytterligare problem är att resultaten ofta rapporteras i breda betygsliknande kategorier, vilket inte ger tillräckligt precis information för diagnostiska ändamål.

Skolverket har inför budgetåret 2014 gett de universitetsinstitutioner som ansvarar för utvecklingen av de nationella proven i uppdrag att lämna förslag dels på hur provens formativa roll kan stärkas i syfte att ge lärare stöd att utveckla sin undervisning, dels på hur ett analysstöd för rektorer kan utformas i syfte att stödja rektorer i uppföljningen av resultaten i de nationella proven. Detta uppdrag bör kunna utgöra en del av planeringen för utvecklingen av det framtida systemet.

Som visats av Linn (2000) uppstår också problem om man försöker använda formativa bedömningssystem när dessa också har stora konsekvenser för eleverna, lärarna eller skolan eftersom icke avsedda negativa effekter av ”*high-stakes*”-bedömningar då dominerar över de avsedda positiva effekterna av den formativa bedömningen. Enligt OECD (2013) kan dock ”*low-stakes*”-bedömningar, exempelvis i form av förebildliga uppgifter, vara värdefulla genom att de kan konkretisera lärandemål och erbjuda läraren intressanta pedagogiska hjälpmedel. I linje med detta har vi ovan föreslagit att den uppgiftsbaserade ansatsen inom den nationella kunskapsbedömningen borde vara en intressant källa av frisläppta väldokumenterade uppgifter med omfattande empirisk underbyggnad som kan användas formativt.

GENOMFÖRANDE

Något förenklat skulle man kunna beskriva innebörden i vårt förslag som att delar av det föreslagna urvalsbaseade nationella kunskapsbedömningssystemet används för att utveckla vissa delprov inom de nationella proven; och man skulle också kunna säga att förslaget innebär att de nationella proven används för att utveckla delar av det nationella utvärderingssystemet. Genom att utnyttja den moderna mätlärens kraftfullhet är det på detta sätt möjligt att skapa komponenter inom det nationella provsystemet vars egen-

skaper är kända innan provet ges, och som ger resultat som är jämförbara över tid. Genom att utvecklings- och utvärderingsarbetet samordnas mellan det nationella utvärderingssystemet och det nationella provsystemet erhålls också betydande synergieffekter och kostnadsbesparingar.

Även om förslaget bygger på användning av relativt komplex psykometrisk teknik är denna teknik väl beprövad och etablerad internationellt sedan flera decennier. Detta innebär dock inte att det är en enkel uppgift att implementera systemet, eftersom det finns en lång rad innehållsliga, metodologiska och administrativa frågor att ta ställning till. Det finns inte möjlighet att gå in närmre på dem här, men vi skulle vilja föreslå att systemet i första steget provas inom ett par ämnen. Eftersom förslaget i huvudsak håller sig inom ramen för befintliga styrdokument och regelverk borde en sådan försöksverksamhet kunna komma igång relativt snart, vilket skulle ge tillgång till konkreta erfarenheter av värde för det fortsatta planeringsarbetet.

Antagning till högre utbildning

Slutsatserna i kapitel 4 visar att antagningssystemet till högre utbildning har allvarliga problem för rättssäkerhet och effektivitet, och att det bör bli föremål för en total genomgång. Vi föreslår därför att en offentlig utredning, helst med parlamentarisk sammansättning, får i uppdrag att göra detta.

Vi vill här uppmärksamma några frågor som är angelägna att belysa i en sådan utredning och som knyter an till våra övriga förslag. Och vi vill peka på hur dessa kan medverka till att lösa en del av problemen med antagningssystemet.

Ett av de stora problemen med det nuvarande systemet för att fastställa kraven för allmän och särskild behörighet till olika utbildningar är att detta görs utan någon tydlig koppling mellan å ena sidan innehållet i gymnasieskolans kurser, och vad elever med olika betyg faktiskt kan förväntas veta och kunna, och å andra sidan de krav som utbildningarna inom högskolan ställer. Vi rekommenderar att innehållsligt och empiriskt baserade kompetensskalor bestäms för de nationella proven och att dessa kopplas till de olika betygsstegen. Detta innebär att en betydligt bättre grund erhålls för att åstadkomma en tydlig koppling mellan gymnasieskolans innehåll och krav för olika betyg och högskolans förkunskapskrav.

Slutsatserna i kapitel 4 visar att gymnasiebetygen är det mest effektiva urvalsinstrumentet, medan Högskoleprovet inte bidrar något till prog-

nosförmågan utöver betygen. Högskoleprovet infördes under början av 1990-talet som ett generellt urvalsinstrument med i huvudsak samma ställning som betygen, under antagandet att dessa två urvalsinstrument hade i stort sett samma prognosförmåga. Då så inte är fallet bör balansen i användningen mellan betyg och högskoleprov ändras. Det finns förvisso starka argument för varför det bör finnas möjlighet till en ”andra chans” av det slag som Högskoleprovet erbjuder, men antalet platser som erbjuds genom Högskoleprovet bör vara betydligt lägre än vad som för närvarande är fallet. Vi föreslår att andelen sökande som antas via Högskoleprovet inte bör uppgå till mer än 10–15 procent av det totala antalet platser. Härigenom, och i kombination med en nedre åldersgräns för att få genomföra provet, visas tydligare att det faktiskt är en ”andra chans”, och att det är studieresultaten från gymnasieskolan som ska utgöra den primära grunden för urval.

Det främsta argumentet mot en utökad användning av de målrelaterade betygen för urval är att deras bristande likvärdighet, både vad gäller stabilitet över tid och vad gäller skillnader mellan lärare och skolor, skapar orättvisor och ineffektivitet. Även om betygen trots detta har god prognosförmåga är det givetvis utomordentligt angeläget att likvärdigheten i betygssättningen förbättras. Våra förslag till förbättringar av de nationella provens förmåga att ge stöd för likvärdig betygssättning leder förhoppningsvis till att likvärdigheten i betygssättningen förbättras och att förtroendet för betygen som urvalsinstrument därmed kan återupprättas.

Rapportens rekommendationer i korthet

KUNSKAPSBEDÖMNING PÅ ELEV- OCH SYSTEMNIVÅ

- * Inför en försöksverksamhet med ett rullande, stickprovsbaserat, nationellt kunskapsbedömningssystem för grund- och gymnasieskolan inom ett par ämnen. Låt en utvärdering av försöksverksamheten ligga till grund för att utveckla ett kunskapsbedömningssystem för andra ämnen.
- * Det nationella kunskapsbedömningssystemet liksom det nationella provsystemet ska omfatta såväl provdelar där resultaten uttrycks i poäng på skolor, som provdelar som ger information av kvalitativ art.
- * Koppla samman det nationella provsystemet med det nationella kunskapsbedömningssystemet i syfte att de nationella proven bättre ska stödja en likvärdig och över tid stabil betygssättning.

- * Samla in resultat från de nationella proven storskaligt och på uppgiftsnivå för att främja uppföljnings- och forskningsändamål.
- * Formulera om direktiven för betygssättning i styrdokumentet så att ett betyg på en kurs speglar en sammanvägd bedömning av elevens kunskaper på kursens olika områden och inte som nu enbart elevens sämsta område.

108

ANTAGNINGSSYSTEMET TILL HÖGRE UTBILDNING

- * Låt en offentlig utredning, helst med parlamentarisk sammansättning, göra en total översyn av antagningssystemet till högre utbildning med syfte att komma till rätta med de allvarliga rättssäkerhets- och effektivitetsproblem som föreligger.
- * En sådan utredning bör bland annat ge förslag på hur kraven på allmän och särskild behörighet till olika högre utbildningar kan kopplas till gymnasieutbildningarnas innehåll och kravnivåer för olika betyg.
- * Renodla Högskoleprovets användning som urvalsinstrument till att utgöra en ”andra chans”, enligt de ursprungliga intentionerna. Detta bör göras genom att dels minska andelen sökande som antas till högre utbildning via Högskoleprovet till 10–15 procent av det totala antalet platser, dels genom att införa en nedre åldersgräns för att få genomföra provet.

Ordlista

Ankaruppgift

En valid och reliabel uppgift, som används i identisk form för att länka prov över tid och därmed fastställa olika gruppers kunskapsnivå.

Behörighetskomplettering

Behörighetskomplettering innebär att den sökande har läst upp eller läst in ett ämne eller en kurs som han eller hon tidigare hade för lågt betyg i eller som helt saknades.

Formativ bedömning

Bedömning under lärandets gång, med syfte att formera och förbättra lärandet; refereras stundtals till som ”bedömning för lärande”.

High-stakes

Bedömningar som har betydelsefulla konsekvenser för enskilda personer eller system, t.ex. examensprov och betyg på individnivå, internationella studier på systemnivå.

Jämförelsetal

Jämförelsetal är att jämföra med medelbetyg, baserat på gymnasiekursernas omfattning i poäng och kursbetyg. Jämförelsetalet (utan meritpoäng, max 2,5 poäng) kan maximalt bli 20,00. Meritvärdet, dvs. det värde som ligger till grund för urval till högskolan utgörs av jämförelsetalet plus meritpoäng, maximalt 22,5.

Kognitiv förmåga

En samlingsterm för de mentala processer som handlar om kunskap, tänkande och information. Som exempel kan nämnas förmågan att minnas, orientera sig i tid och rum, problemlösningsförmåga, numerisk förmåga och språklig förmåga.

Kohort

Beteckning på en grupp individer med vissa gemensamma kännetecken, normalt sett födelseperiod, oftast ett visst år. En födelsekohort består sålunda av personer födda under samma kalenderår.

Konkurrenskomplettering

Komplettering av betyg som höjer jämförelsetalet genom ytterligare studier eller provning, i syfte att förbättra sina konkurrensmöjligheter vid urval till högskoleutbildning.

Low-stakes

Bedömningar med ringa eller inga konsekvenser för enskilda personer eller system.

Matrissampling

En design som innebar att olika elever besvarar delvis olika uppgifter. Härigenom kan många uppgifter användas utan att tids- och arbetsbelastningen för den enskilda eleven blir alltför stor. Möjliggör jämförelse av resultat för elever som besvarat olika uppgifter.

Meritpoäng

Extra poäng för vissa meriterande kurser som läggs till jämförelsetalet från gymnasieskolan vid ansökan till högskolan, som utan meritpoäng kan bli högst 20,00. Det går att få maximalt 2,5 meritpoäng, vilket innebär att jämförelsetalet kan bli högst 22,5.

Normalfördelning

Normalfördelningen är ett begrepp som används inom sannolikhets teori och statistik. En normalfördelad variabel antar ofta värden som ligger nära medelvärdet och mycket sällan värden som har en stor avvikelse. Därför påminner normalfördelningen om en kulle eller en klocka och i engelskan används ofta beteckningen *bell curve*. Kullens utseende kan variera med medelvärdet och standardavvikelsen. Ett exempel på användning av normalfördelningens principer är det relativa betygssystem som tillämpades i Sverige från 1960-talet fram till 1994 och där fasta procentsatser angavs för respektive betygssteg 1-5.

Normerad poäng

Standardiserad poäng enligt en given skala till vilken uppnådd råpoäng omvandlas, exempelvis 0,00-2,00 som resultat på Höskoleprovet. Det är viktigt att det är lika lätt respektive svårt att få ett visst resultat vid varje enskilt prov. Därför används normerade poäng.

Performance-prov

Performance-prov, eller autentiska prov, är bedömningar som kräver att provtagaren utför en uppgift eller aktivitet, inte enbart besvarar enskilda frågor inom det aktuella kunskapsområdet.

III

Reliabilitet

Refererar till en bedömnings tillförlitlighet med avseende på grad av reducerat slumpinflytande och/eller mätfel. Kan avse mätningens stabilitet över tid liksom intern konsistens, dvs. grad av samstämmighet mellan olika delar av en bedömning som fokuserar på samma fenomen. Reliabilitet är en nödvändig förutsättning för validitet, men hög reliabilitet garanterar inte hög validitet.

Råpoäng

Det resultat i form av verkliga poäng som fås ur en undersökning eller annan mätning med poängresultat. Det statistiska materialet skall alltid gå att härleda tillbaka till råpoängen, dvs. de egentliga resultaten på olika frågor eller mätningar.

Stakeholder

Person eller instans för vilken resultat av bedömningar har betydelse; refereras stundtals till som intressent.

Summativ bedömning

Bedömning efter avslutat moment, med syfte att sammanfatta och fastställa nivån av hittills uppnådda kunskaper; refereras stundtals till som "bedömning av lärande".

Tvårsnittsundersökning

Ger en bild av en population vid en viss tidpunkt eller under ett kort tidsintervall; innebär man studerar ett antal individer vid endast ett tillfälle.

Validitet

Refererar till en bedömnings giltighet med avseende på relevans visavi det aktuella kunskapsobjektet, och framför allt på hur resultatet av bedömningen används med avseende på slutsatser, beslut och åtgärder. Stundtals inkluderas även konsekvenser i vid bemärkelse i denna definition. Hög validitet förutsätter hög reliabilitet. Som främsta hot mot validitet brukar anges *construct irrelevant variance* (att annat är det som avses bedömas tillåts påverka resultaten) samt *construct under-representation* (att för lite av det avsedda kunskapsobjektet innefattas i bedömningen).

Referenser

- Atkinson, R. C., & Geiser, S. (2009). "Reflections on a century of college admissions tests", *Educational Researcher*, 38(9), s. 665–676.
- Bedömningshandboken, för antagning till högskoleutbildning 2013–2014. Stockholm: Universitets- och högskolerådet, UHR.
- Björklund, A., Fredriksson, P., Gustafsson, J.-E., & Öckert, B. (2010). Den svenska utbildningspolitikens arbetsmarknadseffekter: vad säger forskningen? (Rapport 2010:13). Stockholm: Institutet för arbetsmarknadspolitiska utvärdering.
- Black, P. J. (1998). "Learning, league tables and National Assessment: opportunity lost or hope deferred?", *Oxford Review of Education*, 24 (1), s. 57–68.
- Carlstedt, B., & Gustafsson, J.-E. (2005). "Construct validation of the Swedish Scholastic Aptitude Test by means of the Swedish Enlistment Battery", *Scandinavian Journal of Psychology*, 46(1), s. 31–42.
- Carroll, J. B. (1982). "The measurement of intelligence", i R. J. Sternberg (red.). *Handbook of human intelligence*. New York: Cambridge University Press.
- Cliffordson, C. (2004a). "Betygsinflation i det målrelaterade betygsystemet", *Pedagogisk Forskning i Sverige*, 9, s. 1–14.
- Cliffordson, C. (2004b). "De målrelaterade gymnasiebetygens prognosförmåga", *Pedagogisk Forskning i Sverige*, 9(2), s. 129–140.
- Cliffordson, C. (2004c). "Effects of practice and intellectual growth on performance on the Swedish Scholastic Aptitude Test (SweSAT)". *European Journal of Psychological Assessment*, 20(3), s. 192–204.
- Cliffordson, C. (2006). "Selection effects on applications and admissions to medical education with regular and step-wise admission procedures", *Scandinavian Journal of Educational Research*, 50(4), s. 463–482.
- Cliffordson, C. (2008). "Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education". *Educational Assessment*, 13(1), s. 56–75.
- Cliffordson, C. (2010). "Mål- och kunskapsrelaterade betyg – betygs utveckling över de första 12 åren". Studie presenterad på en konferens arrangerad av Skolverket 2010. Stockholm: Skolverket.
- Cliffordson, C., & Askling, B. (2006). "Different grounds for admission: Its effects on recruitment and achievement in medical education". *Scandinavian Journal of Educational Research*, 50 (1), s. 45–62.

- Cliffordson, C., & Berndtsson, Å. (2007). "Samband mellan betyg i gymnasieskolan och prestationer i högskolan" (Högskoleverkets rapportserie 2007:21R). Stockholm: Högskoleverket.
- Cliffordson, C., & Gustafsson, J.-E. (2007). "Effekter av den grundläggande högskoleutbildningens expansion på studerandegruppens sammansättning". I Expansion och kontraktion: Utmaningar för högskolesystemet och utbildningsforskare (Vetenskapsrådets rapportserie 9:2007). Stockholm: Vetenskapsrådet.
- College Board (2010). "Educators handbook for the SAT® and the SAT Subject Tests™" http://www.collegeboard.com/prod_downloads/sat/sat-educators-handbook.pdf
- DES (1988). "Task Group on Assessment and Testing: a report" (London, Department of Education and Science and the Welsh Office).
- Fredriksson, P., & Vlachos, J. (2011). "Reformer och resultat: Kommer regeringens utbildningsreformer att ha någon betydelse?" (Rapport till Finanspolitiska rådet 2011/3).
- Gardner, J. (Ed.) (2012). "Assessment and Learning". (Second edition). London: Sage.
- Geiser, S., & Studley, R. (2002). "UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California". *Educational Assessment*, 8(1), s. 1–26.
- Geiser, S., & Santelices, M. (2007). "Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes". Center for Studies in Higher Education, UC Berkeley: <http://cshe.berkeley.edu/publications/publications.php?id=265>.
- Gipps, C. (1994). "Beyond Testing: Towards a Theory of Educational Assessment". London: The Falmer Press.
- Gustafsson, J.-E., (2000). "Mätetegenskaper hos olika urvalsmetoder". I Antagning till högskolan – erfarenheter och visioner (Högskoleverkets rapportserie 14 R, s. 39–46). Stockholm: Högskoleverket.
- Gustafsson, J.-E. (2006a). "Ramverk för ett system för uppföljning av kunskapsutvecklingen i grundskolan". Appendix till Skolverket (2006) Rullande stickprovsbaserat system för kunskapsutvärdering av grundskolans ämnen. Rapport 286. Stockholm: Skolverket.
- Gustafsson, J.-E. (2006b). "Barns utbildningssituation. Bidrag till ett kommunalt barnindex". Stockholm: Rädda Barnen.
- Gustafsson, J.-E. (2008). "Effects of international comparative studies on educational quality on the quality of educational research". *European Educational Research Journal*, 7(1), s. 1–17.
- Gustafsson, J.-E. (2013a). "Causal inference in educational effectiveness research: A comparison of three methods to investigate effects of homework on student achievement". *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 24(3), s. 275–295.
- Gustafsson, J.-E. (2013b). "Förändringar i kunskapsbedömningar på individ- och systemnivå i den svenska skolan under 25 år". I I. Wernersson & I. Gerrbo (red.). *Differentieringens janusansikte. En antologi från Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet*, s. 45–75. Gothenburg Studies in Educational Sciences, 347. Göteborg: Acta Universitatis Gothoburgensis.

- Gustafsson, J.-E., Andersson, A., & Hansen, M. (2000). "Prestationer och prestationsskillnader i 1990-talets skola". I SOU 2000:39, Valfärd och skola, s. 135–211. Stockholm: Fritzes.
- 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), s. 69–87.
- Gustafsson, J.-E., & Rosén, M. (2005). "Förändringar i läskompetens 1991 – 2001. En jämförelse över tid och länder". Rapport 2005:15, Institutionen för Pedagogik och Didaktik, Göteborgs Universitet
- Gustafsson, J.-E., & Yang Hansen, K. (2009). "Resultatförändringar i svensk grundskola". I *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer*, s. 40–85. Stockholm: Skolverket.
- Gustafsson, J.-E., & Yang Hansen, K. (2011). "Förändringar i kommunskillnader i grundskoleresultat mellan 1998 och 2008". *Pedagogisk Forskning i Sverige*, 16(3), s. 161–178.
- Harlen, W., & Deakin Crick, R. (2002). "A systematic review of the impact of summative assessment and tests on students' motivation for learning" (EPPI-Centre Review, version 1.1*). I: *Research Evidence in Educational Library. Issue 1*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Henriksson, W., & Törnkvist, B. (2002). "The Effects of Repeated Test Taking in Relation to the Test Taker and the Rules for Selection to Higher Education in Sweden". (EM No 41, 2002). Umeå: Umeå University, Department of Educational Measurement.
- Hinnerich, B.T., & J. Vlachos (2012). "Systematiska skillnader mellan interna och externa bedömningar av nationella prov", resultatbilaga till Skolinspektionen (2012) *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*, Skolinspektionen
- Hinnerich, B.T., & J. Vlachos (2013). "Systematiska skillnader mellan interna och externa bedömningar av nationella prov – en uppföljningsrapport". Resultatbilaga till Skolinspektionen (2013) *Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan*.
- 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).
- Högskoleverket (2006). "Förslag till nya föreskrifter om särskilda behörigheter". Stockholm: Högskoleverket.
- Jayasinghe, U.W., Marsh, H.W., & Bond, N. (2001). "Peer Review in the Funding of Research in Higher Education: The Australian Experience". *Educational Evaluation and Policy Analysis*, 23(4), s. 343–364.
- 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.
- Klapp Lekholm, A. (2008). *Grades and grade assignment: effects of student and school characteristics*. Göteborg: Acta Universitatis Gothoburgensis.
- Klapp Lekholm, A., & Cliffordson, C. (2008). "Discrepancies between school grades and test scores at individual and school level: effects of gender and family background". *Educational Research and Evaluation*, 14(2), s. 181–199.

- Kobrin J. L., Patterson B. F., Shaw E. J., Mattern K. D., & Barbuti S. M. (2008). "Validity of the SAT for predicting first-year college grade point average". College Board Research (Report No. 2008-5). New York: College Board.
- Koretz, D. (2008). "Measuring up: What educational testing really tells us", Cambridge, MA: Harvard University Press.
- Lawrence, I., Rigol, G., Van Essen, T., & C. Jackson. (2003). "A historical perspective on the content of the SAT". College Board Research (Report No. 2003-03). New York: College Board.
- Linn, R.L. (2000). "Assessments and accountability", *Educational Researcher*, 29(2), s. 4–16.
- Lundahl, C. (2006). "Viljan att veta vad andra vet: Kunskapsbedömning i tidigmodern, modern och senmodern skola". Stockholm: Arbetslivsinstitutet.
- 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.
- Lyren, P.-E., & Ögren, G. (2013). "Det nya högskoleprovet – Några valideringsstudier". (Rapport 2013-10-01). Umeå: Institutionen för tillämpad utbildningsvetenskap Beteendevetenskapliga mätningar.
- Marklund, S. (1987). "Skolsverige. Del 5: Läroplaner". Stockholm: Liber.
- McDonald A. S., Newton P. E., Whetton C., & Benefield P. (2001). "Aptitude testing for university entrance: A literature review." NFER.
- Morgan, R. (1989). "Analysis of the predictive validity of the SAT and high school grades from 1976 to 1983". College Board (Report No. 89-7). New York: College Board.
- Nusche, D., et al. (2011). "OECD Reviews of Evaluation and Assessment in Education – Sweden". OECD: Paris.
- OECD (2013). "Synergies for better learning. An international perspective on evaluation and assessment". Paris: OECD
- Pettersson, S., & Wallin, E. (1995). "The evaluation programme of the Swedish National Agency for Education". I A. Tuijnman & E. Wallin (Red.) *School research at the crossroads*, s. 36–48. Stockholm: Stockholm Institute of Education Press (HLS),
- Rasch, G. (1960). "Probabilistic models for some intelligence and attainment tests". Oxford, England: Nielsen & Lydiche.
- Reuterberg, S.-E. (2003). "Vilken betydelse har utländsk bakgrund för resultat på Högskoleprovet Del II". (2003:23R). Stockholm: Högskoleverket.
- Reuterberg, S.-E., & Hansen, M. (2001). "Vilken betydelse har utländsk bakgrund för resultat på Högskoleprovet Del I". (2001:3R). Stockholm: Högskoleverket.
- Riksrevisionen (2004). "Betyg med lika värde? – en granskning av statens insatser". RiR 2004:23. Stockholm: Riksrevisionen.
- Riksrevisionen (2011). "Lika betyg, lika kunskap? En uppföljning av statens styrning mot en likvärdig betygssättning i grundskolan". RiR 2011:23. Stockholm: Riksrevisionen.
- Rothstein, J. (2004). "College performance predictions and the SAT". *Journal of Econometrics*, 121, 297–317.
- SFS (1993:100). "Högskoleförordningen". Stockholm: Utbildningsdepartementet.
- Skolinspektionen (2010). "Kontrollrättning av nationella prov i grundskolan och gymnasieskolan, 2010". Stockholm: Skolinspektionen.

- Skolinspektionen (2011). "Lika eller olika? Omrättning av nationella prov i grundskolan och gymnasieskolan, 2011". Stockholm: Skolinspektionen.
- Skolinspektionen (2012). "Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år, 2012". Stockholm: Skolinspektionen.
- Skolinspektionen (2013). "Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan". Stockholm: Skolinspektionen.
- Skolverket (1992). "Den nationella utvärderingen av grundskolan 1992". Stockholm: Skolverket.
- Skolverket (1996). "Grundskola för bildning – kommentarer till läroplan, kursplaner och betygskriterier". Stockholm: Skolverket.
- Skolverket (2004). "Nationella utvärderingen av grundskolan 2003. Sammanfattande huvudrapport". Stockholm: Skolverket.
- Skolverket (2006). "Rullande stickprovsbaserat system för kunskapsutvärdering av grundskolans ämnen". Rapport 286. Stockholm: Skolverket.
- Skolverket (2007). "Provbetyg-Slutbetyg-Likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolan, 1998–2006". (Rapport nr 300). Stockholm: Skolverket.
- Skolverket (2009). "Likvärdig betygssättning i gymnasieskolan? En analys av sambandet mellan nationella prov och kursbetyg". (Rapport 338). Stockholm: Skolverket.
- Skolverket (2009). "Bedömaröverensstämmelse vid bedömning av nationella prov". Bilaga till Redovisning av regeringsuppdrag att ge förslag på hur det nationella provsystemet bör utvecklas och utformas. Dnr U2009/1671/S (2009-04-27). Hämtat 2013-11-11 från www.skolverket.se/om-skolverket/2.3901/2.4245/regeringsuppdrag-1.48931.
- Skolverket (2011). "Skolverkets lägesbedömning 2011. Del 1 – Beskrivande data", Skolverket, Rapport 363.
- Skolverket (2011). "Skillnaden mellan betygsresultat på nationella prov och ämnesbetyg i årskurs 9, läsåret 2010/11". Dnr 71-2011:14.
- Skolverket (2012). "Redovisning av uppdrag om avvikelser mellan provresultat och betyg i grundskolans årskurs 9", Redovisning av regeringsuppdrag, Dnr 75-2012:311.
- Skolverket (2013a). "Provbetyg-Slutbetyg gr 9". Dnr 2013:00164; Dnr U2012/7265/S.
- Skolverket (2013b). "Provbetyg-Kursbetyg gy 2012". Dnr 2013:00164; Dnr U2012/7265/S.
- Skolverket (2013c). "PISA 2012. 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap". Rapport 398. Stockholm: Skolverket.
- Skolöverstyrelsen (1970). "Betygssättning i grundskola och gymnasieskola". Rapport utarbetad av Skolöverstyrelsens arbetsgrupp för betygsfrågor. Stockholm: Skolöverstyrelsen Utbildningsförlaget.
- SOU (1942:11). "Betänkande med utredning och förslag angående betygssättningen i folkskolan". Stockholm: Ecklesiastikdepartementet.
- SOU (1974:21). "Om behörighet och antagning till högskolan. Betänkande av kompetenskommittén". Stockholm: Utbildningsdepartementet.
- SOU (2004:29). "Tre vägar till den öppna högskolan". Betänkande av Tillträdesutredningen. Stockholm: Utbildningsdepartementet.
- SOU (2007:28). "Tydliga mål och kunskapskrav i grundskolan - Förslag till nytt mål- och uppföljningssystem". Stockholm: Utbildningsdepartementet.

- SOU (2013:30). "Det tar tid - om effekter av skolpolitiska reformer". Stockholm: Utbildningsdepartementet.
- Stage, C. (2004). "Gruppskillnader i resultat på Högskoleprovet" (PM nr 192). Umeå: Umeå universitet, Institutionen för pedagogiska mätningar.
- Stage, C. (2005). "Socialgruppskillnader i resultat på Högskoleprovet" (BVM 11:2005). Umeå: Umeå universitet, Institutionen för pedagogiska mätningar.
- Stage, C., & Ögren, G. (2001). "Högskoleprovets utveckling under åren 1977–2000. Provet's sammansättning och provdeltagargruppens sammansättning och resultat" (PM nr 169). Umeå: Umeå universitet, Institutionen för pedagogiska mätningar.
- Stage, C., & Ögren, G. (2009). "Högskoleprovet våren och hösten 2009. Provdeltagargruppens sammansättning och resultat". (BVM 40:2009). Umeå: Umeå universitet, Institutionen för pedagogiska mätningar.
- Stobart, G. (2001). "The validity of National Curriculum assessment". *British Journal of Educational Studies*, 49(1), s. 26–39.
- Svensson, A. (1996). "Kulturfria instrument – finns sådana hybrider?" I Högskoleprovet: Genom elva forskares ögon. (Högskoleverkets rapportserie 22 R) Stockholm: Högskoleverket.
- Svensson, A. (2004). "Gymnasiebetyg eller högskoleprov som urvalsinstrument? Fallet civilingenjörutbildningarna". *Pedagogisk Forskning i Sverige*, 9(1), s. 15–36.
- Svensson, A., & Nielsen, B. (2005a). "Toppresultat - men ej antagen till högskolan. Studier av takeffekter hos gymnasiebetyg och högskoleprov". (IPD-rapport nr 2005:08). Göteborg: Göteborgs universitet, Institutionen för pedagogik och didaktik.
- Svensson, A., & Nielsen, B. (2005b). "Vilka kommer in på juristutbildningen och hur klarar man studierna?" *Pedagogisk Forskning i Sverige*, 10(1), s. 1–23.
- Thorsen, C. (2012). "Dimensions of Norm-Referenced Compulsory School Grades and their Relative Importance for the Prediction of Upper Secondary School Grades", *Scandinavian Journal of Educational Research*, DOI:10.1080/00313831.2012.705322.
- Thorsen, C., & Cliffordson, C. (2012). "Teachers' grade assignment and the predictive validity of criterion-referenced grades". *Educational Research and Evaluation*, 18, 2, s. 153–172.
- Thunberg, H., Filipsson, L., & Cronhjort, M. (2006). "Gymnasiets mål och högskolans förväntningar". *Nämnamn*, 2, s. 10–15.
- Tymms, P. (2004). Are standards rising in primary schools? *British Educational Research Journal*, 30(4), s. 477–494.
- Törnkvist, B., & Henriksson, W. (2004). "Repeated test taking. Differences between social groups" (Em No 47). Umeå: Umeå University, Department of Educational Measurement.
- Vlachos, J. (2010). "Betygets värde – en analys av hur konkurrens påverkar betygssättningen vid svenska skolor", *Konkurrensverket uppdragsforskningsrapport 2010:6*.
- Vlachos, J. (2011), "Friskolor i förändring", i Hartman, L. (red.) *Konkurrensens konsekvenser*, SNS Förlag.
- Von Greiff, C. (2009). "Lika skola med olika resurser? En ESO-rapport om likvärdighet och resursfördelning". *Rapport till Expertgruppen för studier i offentlig ekonomi*, 2009:5. Finansdepartementet.

- Waldow, F. (2011). "Bedömningens roll i fördelning av livschanser i Sverige och Tyskland". I C. Lundahl & M. Folke-Fichtelius (red.). *Bedömning i och av skolan – praktik, principer och politik*, (s. 111–125). Lund: Studentlitteratur AB.
- Wedman, I. (2000). "Behörighet, rekrytering och urval: Om övergången från gymnasieskola till högskola". (Högskoleverkets arbetsrapporter 2000:6 AR) Stockholm: Högskoleverket.
- Wikström, C. (2005). "Grade Stability in a Criterion-Referenced Grading System: the Swedish example". *Assessment in Education: principles, policy and practice*, 12(2), s. 125–144.
- Wilson, M. (2005). "Constructing measures: An item response modeling approach". Mahwah, NJ: Erlbaum.
- Wilson, M., & Draney, K. (2013). "A strategy for the assessment of competencies in higher education". In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Eds). *Modeling and measuring competencies in higher education. Tasks and challenges*, s. 61–80. Rotterdam: Sense Publishers.
- Åberg-Bengtsson, L. (1999). "Dimensions of performance in the interpretation of diagrams, tables and maps: Some gender differences in the Swedish Scholastic Aptitude Test". *Journal of Research in Science Teaching*, 36, s. 565–582.
- Åberg-Bengtsson, L. (2005). "Separating the quantitative and analytic dimensions of the SweSAT". *Scandinavian Journal of Educational Research*, 49, s. 359–383.
- Öquist, O. (2010). "Säkra och bilda kunskap – inspektion respektive utvärdering". I C. Lundahl & M. Folke-Fichtelius (red.) *Bedömning i och av skolan – praktik, principer, politik*. Lund: Studentlitteratur.

Hur elevers kunskaper bedöms med hjälp av betyg och nationella prov har stor betydelse både för eleverna och för kunskapsbedömningar på nationell nivå. Det är därför viktigt att bedömningarna görs på ett genomtänkt sätt.

Men hur tillförlitliga är egentligen betygen och de nationella proven som mått på elevernas kunskapsinhämtning? Vad säger de om elevers kunskaper på nationell nivå? Vet vi om dagens svenska elever lär sig mer eller mindre än gårdagens? Det är några av de frågor som diskuteras i *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter*. Rapporten fokuserar på grundskolan och gymnasieskolan, men även antagningssystemet till högre utbildning diskuteras eftersom gymnasiebetygen har en avgörande betydelse i samband med antagning till högre utbildning.

Resultaten ger anledning till oro. Sverige saknar idag ett sammanhängande och välfungerande kunskapsutvärderingssystem för skolan och antagningssystemet till högre utbildning behöver förändras i grunden. Rapporten ger förslag på hur bristerna kan åtgärdas.

Författare till rapporten är JAN-ERIC GUSTAFSSON, professor i pedagogik, Göteborgs universitet, CHRISTINA CLIFFORDSON, professor i pedagogik, Högskolan Väst och GUDRUN ERICKSON, docent i pedagogik, Göteborgs universitet.

