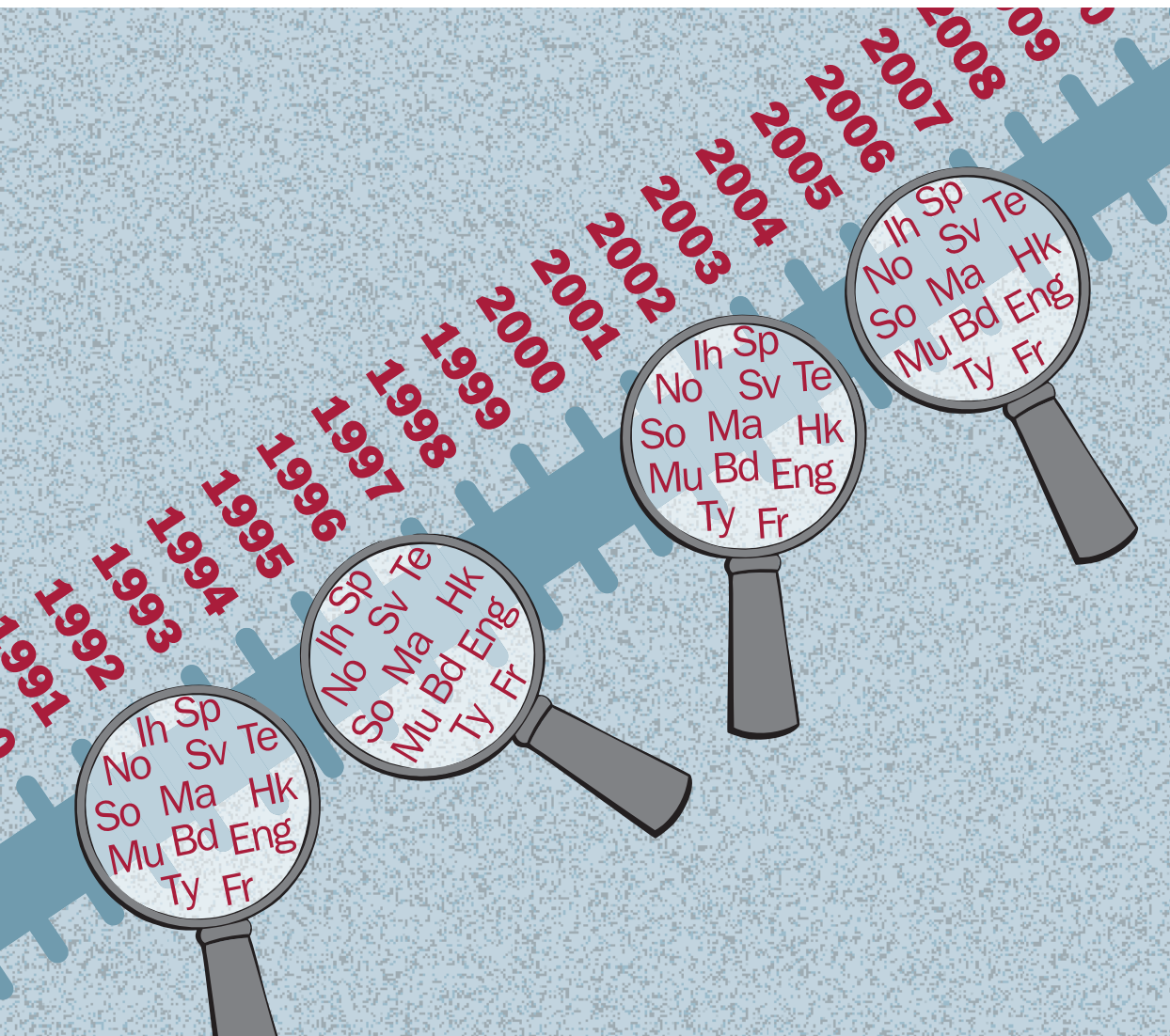


Rullande stickprovsbaserat system för kunskapsutvärdering av grundskolans ämnen



**Rullande stickprovsbaserat
system för kunskapsutvärdering
av grundskolans ämnen**

www.skolverket.se

ISBN: 01-2005:1688

Beställningsnummer: 06:973

ISSN: 1103-2421

ISRN: SKOLV-R-286-SE

Illustration: Marie Kassman/Kassmansbilder
Stockholm 2006

Förord

Skolverket har under en följd av år – 1992, 1995, 1998 och 2003 – genomfört större s.k. nationella utvärderingar av grundskolan och dess olika ämnen. I december 2004 uppdrog regeringen åt verket att utreda förutsättningarna för att införa ett mer systematiskt, rullande stickprovsbaserat system för kunskapsutvärdering för samtliga ämnen i grundskolan. Tanken i uppdraget var att ett begränsat antal ämnen utvärderas med viss regelbundet enligt ett rullande system.

Skolverket har utrett denna fråga och redovisar i föreliggande rapport sina bedömningar och förslag vad gäller ett framtida rullande stickprovsbaserat system för kunskapsutvärdering av grundskolans ämnen i Sverige. Rapporten redovisas som bilaga till Skolverkets redovisning av regeringens uppdrag.

Med anledning av regeringens uppdrag gav verket 2005-05-19 i uppdrag till professor Jan-Eric Gustafsson vid Institutionen för pedagogik och didaktik, Göteborgs Universitet, att utarbeta ett förslag till övergripande ramverk för en rullande uppföljning/utvärdering av elevgruppers kunskapsutveckling, med särskilt beaktande av internationella erfarenheter av denna typ av utvärderings-system. Gustafsson avrapporterade 2006-05-16 sitt uppdrag i form av en särskild underlagsrapport *Ramverk för ett system för uppföljning av kunskapsutvecklingen i grundskolan*. Gustafssons underlagsrapport innehåller bl.a. en gedigen genomgång av internationella, metodologiska och andra erfarenheter på området, vilka kan anses ha ett mer allmänt intresse än som ett internt underlag för Skolverket. Den redovisas därför som appendix till föreliggande rapport.

Skolverkets rapport har tagits fram i en projektgrupp bestående av undervisningsråden Sten Söderberg (projektledare), enhetschefen Kristian Ramstedt samt undervisningsråden Bo Palaszewski, Camilla Thinsz Fjellström, Tommy Lagergren och Anita Wester.

Stockholm den 18 september 2006

Per Thullberg
Generaldirektör

Sten Söderberg
Undervisningsråd, projektledare

Innehåll

1 Inledning och sammanfattande slutsatser	6
2 Bakgrund	10
2.1 Uppdraget.....	10
2.2 Skolverkets tidigare och pågående verksamheter med avseende på nationell kunskapsbedömning och nationell utvärdering	12
3 Skolverkets uppdrag till professor Jan-Eric Gustafsson	34
3.1 Skolverkets hantering och avgränsning av regeringens uppdrag.....	34
3.2 Sammanfattning av Jan-Eric Gustafssons underlag och förslag	36
4 Skolverkets överväganden	48
4.1 Kostnadsuppskattningar.....	48
4.2 Måluppfyllelsebedömning eller trendmätning – olika ämnens olika förutsättningar	50
4.3 Koppling till internationella studier – möjligheter och begränsningar	56
4.4 Periodicitet – hur ofta bör mätningar genomföras och i vilka årskurser?	62
4.5 Ämneskombinationer och risker för oönskade effekter	63
4.6 Hur ska ett nationellt system komma till lokal användning?	65
Referenser	68
Bilaga	72

Inledning och sammanfattande slutsatser

1 Inledning och sammanfattande slutsatser

Regeringen uppdrog 2004-12-22 åt Skolverket att utreda förutsättningarna för att införa ett rullande stickprovsbaserat system för kunskapsutvärdering för samtliga ämnen i grundskolan. Med anledning av regeringens uppdrag gav verket 2005-05-19 i uppdrag till professor Jan-Eric Gustafsson vid Institutionen för pedagogik och didaktik, Göteborgs Universitet, att utarbeta ett förslag till övergripande ramverk för en rullande uppföljning/utvärdering av elevgruppers kunskapsutveckling, med särskilt beaktande av internationella erfarenheter av denna typ av utvärderingssystem. Gustafsson har 2006-05-16 avrapporterat sitt uppdrag i form av en särskild underlagsrapport *Ramverk för ett system för uppföljning av kunskapsutvecklingen i grundskolan*.¹ Gustafssons underlagsrapport och förslag har behandlats såväl inom Skolverket som vid två brett sammansatta forskarsymposier. I föreliggande rapport redovisas Skolverkets överväganden och slutsatser med anledning av dels Gustafssons underlagsrapport, dels det övriga och fortsatta utredningsarbete som föranletts av denna.

Sammanfattningsvis kan konstateras att Skolverkets samlade verksamhet inom områdena Uppföljning, Utvärdering och Inspektion är omfattande, men f.n. saknas en systematisk verksamhet vars primära syfte och metodik är inriktad på nationell bedömning av elevers kunskapsresultat och hur dessa utvecklas över tid. Det är heller knappast ett realistiskt alternativ att fortsätta upprepa de tidigare stora satsningarna i form av Skolverkets ”nationella utvärderingar” (1992, 1995 och 2003), vilka har haft utgångspunkt i tidigare läroplaner och kursplaner och i ett tidigare styrsystem.² Dessutom utgör sådana utvärderingsmodeller avsevärda kraftsamlingar, och de glea intervallen skulle vid varje enskilt tillfälle förutsätta omfattande investeringar, inte bara rent ekonomiskt utan också uppbyggnad av kompetens. Skolverkets tätare återkommande men såväl innehållsligt som metodologiskt avgränsade attitydundersökningar är i så fall – utan jämförelser i övrigt – en mer rimlig ”förebild”.

Samtidigt visar sig en modell för rullande kunskapsutvärdering av alla grundskolans ämnen i enlighet med regeringens uppdrag, och av den typ som Jan-Eric Gustafsson på Skolverkets uppdrag har utrett och föreslagit – och som i sin tur delvis bygger på beprövade erfarenheter och modeller från andra länder – alltför kostsam för ett relativt litet land som Sverige. Detta gäller både i uppbyggnadsskedet och i en tänkt mer regelbunden driftsituation.

¹ Jan-Eric Gustafssons underlagsrapport biläggs föreliggande avrapportering som appendix.

² Utgångspunkten för genomförandet av NU 03 var att genomföra en s.k. repeat³ på NU 92, även om en del förändringar efter hand befanns nödvändiga att göra i såväl kunskapsbedömningsinstrument som enkäter, jämfört med 1992 års instrument.

Skolverkets avsikt är istället att – under förutsättning att verkets ekonomiska ramar så medger – genomföra en strategi för nationella kunskapsutvärderingar med följande inslag:

De internationella kunskapsmätningarna föreslås bli en integrerad del i en nationell strategi och helt eller delvis ersätta nationella mätningar inom valda kunskaps- eller ämnesområden. Utgångspunkten blir då att, i de fall där en internationell studie står i rimlig överensstämmelse med intentioner och ambitioner i de nationella måldokumenterna, de internationella studierna med fördel kan ersätta de rent nationella mätningarna. De blir med andra ord en integrerad och mer eller mindre löpande del i det nationella systemet, snarare än – som hittills – en tämligen fristående verksamhet. Därmed ökar kravet på att även de internationella studierna genomförs med viss regelbundenhet och de kan således inte prövas förutsättningslöst vid varje nytt tillfälle. Inom ramen för en nationell strategi blir det också angeläget att Sverige deltar så tidigt som möjligt – förutsatt att Sverige ska delta – för att säkerställa att mätinstrumenten i så hög grad som möjligt avspeglar de svenska måldokumentens målskrivningar. Vidare är det önskvärt att i de enskilda undersökningarna utvärdera möjligheterna att lägga till svenska moduler för att mäta eventuella områden (domäner) som inte täcks av de internationella mätinstrumenten i syfte att öka studiens värde från nationell synpunkt.

Skolverket avser härutöver att så snart möjligt – och som nämnts under förutsättning att de ekonomiska ramarna så medger – implementera ett nationellt kunskapsutvärderingssystem (NKU) omfattande ett fåtal strategiskt valda ämnen där det nationella kunskapsbehovet, åtminstone på kort sikt, inte täcks av internationella studier, eller av andra skäl bedöms som särskilt strategiskt angeläget. Skolverket menar att detta system kan avgränsas till fyra strategiskt viktiga grundskoleämnen vilka förslagsvis utvärderas vart femte år och två åtgångar, i årskurserna 4/5 och 8/9. Exakt vilka ämnen som ska ingå och hur omfattande domäntäckning inom dessa ämnen som är lämplig behöver dock utredas vidare. Preliminärt kan dock sägas att svenska och valda delar av SO-ämnena bör ingå; vidare bör engelska ingå i planeringsarbetet i avvaktan på hur EU-kommissionens utveckling av en s.k. språkindikator utfaller. Ett rimligt startår för planeringsarbetet är 2007 med sikte på en första datainsamling 2009 eller senast 2010.

Övriga ämnen i grundskolan – vilka inte ingår i internationella kunskapsmätningar eller ett nationellt NKU – behöver också utvärderas med viss regelbundenhet. Detta gäller inte minst de praktisk-estetiskt inriktade ämnena. I många fall saknas dock här vedertagna mät- och utvärderingsmetoder som också är lämpade för jämförelser av kunskapsresultat över tid. Beslut om såda-

na särskilda ämnesutvärderingar får fattas i särskild ordning. Det kan vara värt att här överväga mera tematiska eller processinriktade studier, vilka eventuellt skulle kunna samordnas med utbildningsinspektionens granskningar.

Motsvarande gäller mer (ämnes)övergripande läroplansmål som t.ex. problemlösningsförmåga, kreativitet eller samarbetsförmåga. Inget hindrar dock att sådana övergripande kompetensområden i tillämplig utsträckning kan beaktas inom NKU, eller för den delen i anslutning till vissa av de internationella studierna.

Bakgrund

2 Bakgrund

2.1 Uppdraget

2.1.1 Skolverkets tidigare förslag och bedömningar kring ett rullande utvärderingssystem

Bakgrunden till föreliggande utredning är Skolverkets rapport 2003-10-30 Det nationella provsystemet – vad, varför och varthän?³ och som var ett av flera underlag som Skolverkets lämnade som svar på ett tidigare regeringsuppdrag att göra en ”bedömning av dagens provsystem med avseende på kvalitet och kostnadseffektivitet”.⁴ I denna rapport konstaterade Skolverket att det nuvarande provsystemet ej lämpar sig för uppföljning och utvärdering av kunskapsnivåer och utveckling av kunskap över tid, vare sig det gäller grundskolan eller gymnasieskolan. De nuvarande nationella proven, menade verket, är inte konstruerade för att resultaten ska aggregeras och t.ex. jämföras mellan olika år. Stoffvalets betydelse medför också att de inte lämpar sig särskilt väl för jämförelser mellan klasser eller skolor. Verket förordade istället ett system där nationell uppföljning eller utvärdering genomförs genom att nationella stickprov deltar i utvärderingar inom olika skolämnen. Dessa utvärderingar skulle gälla fler ämnen än de nuvarande provämnena och skulle kunna genomföras enligt ett rullande schema med vissa intervall för respektive ämne. I Skolverkets förslag angavs att ett sådant system skulle tillgodose behovet av nationell och lokal uppföljning av olika skolämnen inom såväl grundskolan som utbildning på gymnasial nivå. Ett sådant system skulle också kunna ersätta Skolverkets tidigare ”nationella utvärderingar” som NU-92 och NU-03. Systemet skulle dock behöva utredas närmare.

Vidare menade Skolverket⁵ att de internationella systemen för kunskapsutvärdering (PISA, TIMSS, PIRLS osv.) i ett nationellt perspektiv mycket väl kan fungera som utvärderingar av kunskapsläget inom avgränsade ämnesområden i det svenska skolsystemet, med referens till såväl andra länder som till tidigare resultat.

Slutligen kan påpekas att Skolverket i sin rapportering diskuterade en rad andra mer eller mindre radikala förändringar av det nationella provsystemet, såsom t.ex. att totalinsamling av resultat från nationella prov överges till förmån för stickprovsbaserade insamlingar; liksom en övergång till ett elektroniskt distributionsförfarande för delar av det nationella provsystemet. En

³ Skolverket (2003).

⁴ Utbildningsdepartementet (2003)

⁵ *ibid.*, s. 123.

omläggning av hela provsystemet enligt dessa och övriga förslag – inklusive ett rullande stickprovsbaserat system för nationell kunskapsutvärdering istället för sporadiskt återkommande ”nationella utvärderingar” – bedömdes kunna bli kostnadsneutralt⁶ även om kostnaderna enligt verket skulle behöva utredas närmare.

2.1.2 Regeringens uppdrag till Skolverket

Regeringens uppdrag 2004-12-22 åt Skolverket avseende det nationella provsystemet innefattar bl.a. att utreda förutsättningarna för att införa ett rullande stickprovsbaserat system för kunskapsutvärdering. Den del av regeringens uppdrag som här avses lyder:

Skolverket skall prioritera en lösning med rullande undersökningar vid utvecklingen av det föreslagna stickprovsbaserade systemet för kunskapsutvärdering (motsvarande NU 03).

I bilaga till regeringsbeslutet⁷ anges, vad gäller systemets inriktning, att

Skolverket har föreslagit införandet av ett stickprovsbaserat system för kunskapsutvärdering för samtliga ämnen i grundskolan och för ämnen i gymnasieskolan (motsvarande rullande NU 03). Detta skulle bidra till måluppfyllelsen i fler ämnen än de för vilka det idag finns nationella prov, dvs. svenskalsvenska som andraspråk, engelska och matematik.

Vidare avgränsas i bilagan Skolverkets uppdrag till att omfatta *grundskolan*:

Skolverket bör vid utvecklingen av det föreslagna stickprovsbaserade systemet för kunskapsutvärdering för samtliga ämnen i grundskolan prioritera lösningen med rullande undersökningar (motsvarande de undersökningar som genomförts i NU 03).

Provsystemet bör så långt som möjligt utnyttja såväl de erfarenheter som följer av Sveriges deltagande i internationella uppföljningar som resultaten från dessa studier.⁸

⁶ *ibid.*, s. 26.

⁷ Promemoria, Bilaga till regeringsbeslut 2004-12-22, nr 21.

⁸ Sammanhanget i regeringens skrivelse är sådant att Skolverket har tolkat ordet ”Provsystemet” i den sista meningen såsom syftande på ”den nationella kunskapsutvärderingen”.

2.2 Skolverkets tidigare och pågående verksamheter med avseende på nationell kunskapsbedömning och nationell utvärdering

2.2.1 Inledning

Inledningsvis vill Skolverket här uppmärksamma att begreppet ”Utvärdering” i olika sammanhang ges mycket olika innebörd, såväl inom Sverige som i olika länder. Dessa olika typer av aktiviteter är ofta föremål för en intensiv ideologiskt färgad diskussion. Skolverket tar dock inte ställning till huruvida vissa typer av utvärdering är att föredra framför andra, utan vill här främst konstatera att de normalt har olika syften, riktar sig till skilda målgrupper och ger svar på helt olika typer av frågor.

Nedanstående aktiviteter går ofta under beteckningen ”Utvärdering” och är – i varierande grad – vanliga inom olika delar av skol- och utbildningssektorn i åtskilliga länder:

Uppföljning (motsvaras närmast av det engelska begreppet ”*monitoring*”) är en i huvudsak kvantitativ, statistikbaserad, aktivitet innehållande bl.a. statistik över betygs- och i förekommande fall provresultat. Sådana system finns i ett flertal länder – Skolverkets nationella uppföljningssystem, vilket numera också innehåller data på skolnivå som redovisas i bl.a. SIRIS-systemet – är ett ambitiöst exempel. Likaså kan OECD:s s.k. INES-projekt⁹ vilket resulterar i de årliga EAG –publikationerna (Education at a Glance) ses som ett exempel på denna typ av aktivitet.

Kunskapsutvärdering (motsvaras närmast av det engelska begreppet ”*assessment*”) baseras på prov eller tester av elevers kunskaper och baseras på vetenskapligt avancerad testteori (psykometri eller pedagogiska mätningar). Metoden används både på *individnivå* (prov och annat material för stöd i bedömning och diagnos av enskilda elevers kunskapsutveckling) och på *systemnivå*, oftast baserad på stickprovsmetodik och med primärt syfte att på nationell nivå kontrollera kunskapsresultat och utgöra underlag för politiska beslut. ”Nationella provsystem” av det slag som finns i Sverige men också i andra länder (t.ex. Norge, England, Frankrike, Nederländerna, USA och som nyligen har introducerats i Danmark¹⁰) är av den förstnämnda typen (individnivå). Dessa ”individsystem” är egentligen inte avsedda att ge underlag för nationell kunskapsbedömning över tid, bl.a. eftersom testuppgifterna varierar över tid och resultaten normalt bedöms/rättas lokalt, av berörda lärare. Statistik över provresultat samlas dock ofta in inom ramen för uppföljningssystem. De internationella mätningarna (PISA, TIMSS och PIRLS m.fl.) tillhör den andra typen av kunskapsutvärdering (systemnivå). Det senare kännetecknar också de

⁹ Indicators of Educational Systems.

¹⁰ Den s.k. ”Fremme af evalueringkulturen i folkeskolen”.

nationella ”assessment”-systemen i USA, Nederländerna och Nya Zeeland som redovisas i Jan-Eric Gustafssons inledningsvis nämnda underlagsrapport.

Inspektion (motsvaras närmast av de engelska begreppen ”*inspection*” eller ”*oversight*”) förekommer traditionellt i de flesta länder på grundskolenivå och gymnasial nivå, numera åter i Sverige i form av Skolverkets utbildningsinspektion. Dess fokus är på (oftast statliga) besök och kontroll på enskilda skolor. Kunskapsbedömning genomförs inte utan man stödjer sig, vad gäller resultatbedömning, i huvudsak på betygs- och provstatistik på skolnivå. Etablerade och väl utbyggda inspektorat finns förutom i Sverige också i bl.a. England, Skottland, Frankrike och Nederländerna, men inte i övriga nordiska länder.

Kollegial bedömning baserad på *självevaluering* (motsvaras närmast av det engelska begreppet ”*peer review*”) är en vedertagen metod på högskolenivå. Metoden grundas på att expertteam besöker och bedömer kvaliteten på utbildningsinstitutioner inom de ämnesområden de är experter på, varvid s.k. självbedömningar ofta ligger till grund för expertbedömningen. Det svenska Högskoleverkets utvärderingsprogram baseras på denna metodik, liksom det danska Evalueringsinstitutet (EVA). Den senaste händelseutvecklingen i Danmark indikerar dock att denna metod kan ha sina begränsningar om syftet är bedömning av elevers kunskapsresultat inom grundläggande utbildning och gymnasial utbildning.

Evaluering eller *utvärdering* (motsvaras närmast av det engelska begreppet ”*evaluation*”) används ofta som samlingsbegrepp för ovan nämnda aktiviteter, men reserveras ibland också för en forskningsbaserad och ambitiös helhetsbedömning av såväl förutsättningar, processer som resultat inom skol- eller utbildningsverksamhet på olika nivåer (såväl skolnivå som nationell nivå). Ibland ingår kunskapsmätningar, ibland inte. Metoden har eller har haft en stark ställning i i synnerhet Sverige och Norge men också främst på lokal nivå i USA. Skolverkets ovan nämnda ”nationella utvärderingar” kan sägas vara exempel på denna typ av aktivitet, likaså de norska utvärderingarna av senare års läroplansreformer (”Reform 97” och ”Kunskapslöftet”).

Skolverket har, inom sitt verksamhetsområde, löpande och väletablerad verksamhet inom områdena Uppföljning (inklusive statistik över betyg och provresultat), Kunskapsbedömning på individnivå (det nationella provsystemet), Kunskapsbedömning på systemnivå i form av deltagande i internationella jämförande studier samt Inspektion. Dessutom kan Skolverkets tidigare nationella utvärderingar sägas vara exempel på Evaluering/Utvärdering men på ad-hoc-mässig snarare än systematisk och planerad basis.¹¹ Föreliggande

¹¹ Skolverket har nyligen (2006-09-13) beslutat om en strategi för verkets utvärderingsverksamhet och långsiktiga utvärderingsplan.

avrapportering syftar till att undersöka möjligheterna att komplettera metodarsenalen med Kunskapsbedömning på nationell nivå. Innan vi går in på denna fråga, vill vi dock ge en kort översikt över i vilken grad Skolverkets övriga verksamheter ger, eller kan väntas ge, underlag för bedömning av elevers kunskaper och den svenska grundskolans resultat.

2.2.2 Skolverkets tidigare "nationella utvärderingar"

Skolverket har tidigare genomfört s.k. nationella utvärderingar år 1992 (Nationella utvärderingen av grundskolan 1992; NU 92), 1995 (Utvärdering av grundskolan 1995; UG 95), 1998 (Utvärdering av skolan avseende läroplanernas mål; US 98) och 2003 (Nationella utvärderingen av grundskolan; NU 03). 1992 års nationella utvärdering, NU92, byggde i sin tur på den försöksomgång som Skolöverstyrelsen (SÖ) genomförde år 1989 i grundskolans årskurser 2 och 5 och som avsåg ämnena svenska, matematik, engelska, "omvärldskunskap", bild och musik. NU 92, UG 95 och NU 03 var stora riksrepresentativa undersökningar av svenska, engelska och matematik i grundskolans årskurs 5 och de allra flesta av grundskolans ämnen i årskurs 9: Ett riksrepresentativt urval av ett hundratal grundskolor på vilka alla mätningar gjordes. Likheten i urvalsmetodik och innehåll avsåg att dels möjliggöra jämförelser över tid i bedömningen av elevers kunskaper och färdigheter, dels att kunna göra uttalanden och ge feedback till de enskilda deltagande skolorna.

NU 92, UG 95 och NU 03 var i första hand inriktade mot mer traditionella mätningar av kunskaper och färdigheter i de enskilda skolämnena i grundskolan. I vissa avseenden introducerades dock nya bedömningsmetoder, såsom den studie av undersökande arbetssätt (problemlösning i grupp) som första gången genomfördes inom ramen för NU 92 och sedan har vidareutvecklats inom ramen för US 98 och NU 03. I alla tre utvärderingsprojekten studerades också frågor som bakgrundsvariabler, attityder till ämnena, motivation, skolans sociala klimat, lärarbehörighet m.m. med hjälp av bl.a. omfattande elev-, lärar- och skolenkäter. Dessa enkäter var i allt väsentligt desamma i NU 92 och UG 95, medan det i NU 03 befanns nödvändigt att göra omfattande revideringar och kompletteringar av enkätmaterial, inte minst mot bakgrund av de stora förändringar i såväl styrdokument som samhällsklimat som inträffat sedan 1995.

US 98 hade ett delvis annorlunda upplägg och var ett försök att bryta ny mark med utgångspunkt från den då relativt nya läroplanen (Lpo 94), varvid fem ämnesövergripande läroplansmål/kompetensområden studerades, i flera fall med kvalitativ metodik vid ett mindre antal skolor. Varken undersökning-

ens uppläggnings eller metodiken i kunskapsmätningarna var utformade med tanke på att de skulle medge trendmätningar av kunskapsnivåer.¹²

Det uttalade syftet med såväl UG 95 som NU 03 var att utgöra upprepningar, s.k. ”repeat”, av NU 92, både vad gäller kunskapsmätningarnas innehåll och undersökningsmetodik – även om detta syfte i NU 03 inte kunde fullföljas helt och hållet, vilket utvecklas nedan. NU 92 har på så vis kommit att stå som sinnebild och modell för begreppet ”nationell utvärdering” inom Skolverket. Det förtjänar därför att påpekas att denna ”tradition” har sitt ursprung i en tidigare statlig skoladministration (SÖ) och en tidigare läroplan (Lgr 80). Det var regeringen som initierade NU 89 mot bakgrund av den mot slutet av 1980-talet upplevda bristen på kunskap om grundskolans resultat. Samtliga Skolverkets nationella utvärderingar från NU 92 och framåt har varit egeninitierade och inte regeringsuppdrag (och har som sådana finansierats inom förvaltningsanslaget). Besluten att genomföra utvärderingarna har varje gång fattats (relativt ad-hoc-mässigt) internt inom Skolverket, och det har inte funnits någon långsiktig plan – varken från Skolverkets eller statsmakternas sida – att genomföra denna typ av nationella studier med vissa intervall.

Ytterligare ett gemensamt drag för Skolverkets ”nationella utvärderingar” har varit att utvecklingen av ramverk och instrument, liksom analysen av insamlade data på ämnesnivå, av Skolverket har lagts ut på olika ämnesdidaktiska universitets- och högskoleinstitutioner runt om i landet. Dessa institutioner och forskargrupper har i huvudsak varit desamma sedan 1989 och 1992. Även valet av forsknings- och utvärderingsmiljöer emanerar således i allt väsentligt från en tidsepok innan Skolverket bildades. Institutionerna har haft relativt fria händer att tolka och utforma mätningarna utifrån dels gällande läro- och kursplaner, dels (i UG 95 och i synnerhet NU 03) det uttalade ”repeat”-syftet. Något samlat och genomtänkt ramverksarbete av den typ som genomförts i exempelvis USA och Nederländerna, och som utgör utgångspunkten för Skolverkets uppdrag till Jan-Eric Gustafsson, har egentligen inte använts under hela denna period, dvs. sedan andra hälften av 1980-talet.

Utvärderingarna har i genomsnitt kunnat avrapporteras ca 18–20 månader efter genomförd datainsamling, vilket är ungefär samma tidsutdräkt som för de internationella jämförande kunskapsmätningar vilka har genomförts under senare år. De nationella utvärderingarna har enligt Skolverkets egen uppfattning fått oerhört stort genomslag och påverkat den utbildningspolitiska diskussionen i hög grad. Samtidigt kan konstateras att de i samtliga fall varit

¹² Möjligen med undantag för vissa inslag i delprojekten ”Tema tillståndet i världen” (SO/NO) och ”Kommunikativ förmåga i främmande språk” (engelska, franska och tyska) där vissa instrument som använts i NU 92 respektive UG 95 återanvändes. Se vidare Skolverket (1999b och 1999c).

fleråriga och mycket kostnadskrävande kraftsamlingar vilka har fått finansieras inom ramen för Skolverkets ordinarie förvaltningsanslag och i resursmässig konkurrens med andra verksamhetsgrenar inom verket. Exempelvis kan den sammanlagda kostnaden för US 98 uppskattas till ca 11 Mkr och för NU 03 till ca 30 Mkr.

Några fortsatta planer på ”nationell utvärdering” av angiven typ finns inte. Skolverket betraktar regeringens uppdrag om rullande nationell utvärdering som ett uttryck för ett önskemål om en mer systematisk planering och genomförande av kunskapsutvärderingar. I avvaktan på slutsatserna av föreliggande utredning har dock verket satt igång planering för en kompletterande nationell utvärdering, avsedd att genomföras våren 2008, av de grundskoleämnena som av olika skäl inte ingick i NU 03, dvs. Språkvalet med fokus på moderna språk (franska, spanska, tyska) och teknik – en studie som går under arbetsnamnet NU 08. Spanska och teknik vilka introducerades som nya ämnen i och med införandet av Lpo 94 har aldrig varit föremål för någon nationell utvärdering över huvud taget.

Nationella utvärderingar av elevers kunskaper bygger på att två viktiga aspekter beaktas så långt möjligt: *Trend och måluppfyllelse* – ett resonemang som utvecklas närmare i avsnitt 4.2 nedan. Bedömning av *trender* i kunskapsutveckling förutsätter att det i samtliga skolämnena som ska utvärderas finns ett överenskommet ramverk och en ”kärna” av kunskaper som är någorlunda stabil och därmed låter sig bedömas över tid.¹³ En bedömning av *maluuppfyllelse* å andra sidan förutsätter att de nationella målen i olika ämnen och avseenden är så pass tydliga att de låter sig konkretiseras och operationaliseras i mätinstrument och provuppgifter som kan vinna allmän acceptans. Samtidigt kan målen i ämnena helt naturligt behöva anpassas och förändras över tid, vilket då kan komma att stå i konflikt med trendmätningens modellen.

I verkligheten kan båda dessa modeller vara svåra att realisera. Erfarenheterna från NU 03 är tydliga på denna punkt. Även om det uttalade syftet med NU 03 var att göra en ”repeat” på NU 03, visade det sig att de båda kursplaneförändringarna 1994 och 2000 i vissa ämnen, inte minst i de praktisk-estetiska ämnena, varit så pass omfattande att en direkt upprepning av tidigare mätningar framstod som mindre meningsfull. Vidare befanns i vissa ämnen att kursplaneeinnehållet är så pass allmänt hållet att långtgående operationaliseringar måste göras av de anlidade forskargrupperna, operationaliseringar som inte självklart eller nödvändigtvis svarar mot de operationaliseringar som görs lokalt av

¹³ Som närmare utvecklas nedan i föreliggande rapport är detta ett antagande och en förutsättning som har styrts (och styr) de internationella jämförande studierna och de nationella kunskapsutvärderingssystemen, t.ex. det amerikanska NAEP.

verksamma lärare. Inte minst gäller detta SO-ämnena. I vilken grad NU 03:s resultat ger besked om förändring över tid och/eller graden av måluppfyllelse varierar därför mellan de olika ämnena. Dessa erfarenheter leder Skolverket till uppfattningen att en ytterligare upprepning av NU 03 om – exempelvis – tio år framstår som mindre meningsfull.

Enligt verkets uppfattning förtar dock detta inte värdet av nationella utvärderingar av den typ som NU 03 representerar. Den samlade styrkan har således snarast legat i *dels* den samlade och jämförande ”bild” av grundskolan och dess olika ämnen som utvärderingen har givit, *dels* i den kring- och processinformation som har givits av NU 03:s olika enkäter och som kan tjäna som förklaringsunderlag. Åtminstone den senare aspekten – vikten av enkätdata – kan och bör tas tillvara i fortsatta nationella kunskapsutvärderingar.

Det förtjänar slutligen att här påpekas att gymnasieskolan aldrig har varit föremål för liknande nationella utvärderingar med tonvikt på kunskapsbedömningar och elevresultat.¹⁴ De två större ”nationella utvärderingar” av gymnasieskolan som har genomförts: Den första utvärderingen av gymnasiereformen 1995 den s.k. reformutvärderingen, samt utvärderingen av de då relativt nya sexton nationella programmen 1997–1999 (den s.k. programutvärderingen) 1998. Dessa utvärderingsprojekt var upplagda på ett annat sätt än ämnesinriktade kunskapsmätningar: Fokus var på reformgenomförande, förutsättningar och processer snarare än på ämneskunskaper och måluppfyllelse. Den nationella bilden av gymnasieskolans resultat är således svag. Regeringens uppdrag om rullande nationell utvärdering omfattar dock inte gymnasieskolan, varför denna skolform inte har inkluderats i föreliggande utredning och förslag.

2.2.3 Skolverkets deltagande i internationella jämförande studier

Sverige har varit med att utveckla och har deltagit i internationella jämförande studier av elevers kunskap alltsedan de startade för drygt 40 år sedan.¹⁵ De första internationella jämförande studierna av elevers kunskap började med bildandet av IEA¹⁶ på slutet av 1950-talet.

De ämnesområden i grundskolan där IEA har genomfört studier av elevers kunskaper och färdigheter mer eller mindre regelbundet är framför allt inom följande områden: läsförståelse¹⁷ (Reading Literacy; åren 1970 och 1990/91

¹⁴ Den gymnasiala nivån har dock ingått i ett fåtal internationella studier i vilka Sverige har deltagit, exempelvis ”Civic Education in ten Countries” 1970–74, TIMSS 1995 och CIVED 1999/2000. Vidare avser Sverige, enligt ett beslut fattat av Skolverket 2006-08-21, att delta i TIMSS Advanced 2008 som också avser gymnasial nivå.

¹⁵ För en mer detaljerad beskrivning av de internationella kunskapsstudierna, se Skolverket (2004c).

¹⁶ International Association for the Evaluation of Educational Achievement.

¹⁷ Se. Skolverket (1995).

och PIRLS 2001¹⁸ samt PIRLS 2006), matematik och naturvetenskap (TIMSS 95,¹⁹ TIMSS 2003²⁰ och TIMSS 2007) samt elevers kunskaper, engagemang och attityder i samhälls- och demokratifrågor (Civic Education Study 1970–74; CIVED 1999/2000). Samtliga IEA-studier har framför allt berört grundskolan samt i viss utsträckning även gymnasieskolan

Sverige har deltagit i samtliga av IEA:s läsförståelsestudier, alltsedan början på 1970-talet. IEA:s läsförståelsestudie PIRLS är en fristående fortsättning på Reading Literacy 1991, vilken var den andra internationella läsförståelsestudien – den första, den s.k. ”Six Subjects Survey”, genomfördes som nämnts redan år 1970. Sverige deltog i den första PIRLS-studien, PIRLS 2001, och deltar också i den pågående studien, PIRLS 2006. PIRLS studerar läsfärdighet hos de yngre eleverna (elever i årskurs 4). Studien är upplagd för att studera förändring i kunskaper och färdigheter över tid. PIRLS är tänkt att upprepas vart femte år.

IEA har i olika studier undersökt elevers kunskaper och färdigheter i matematik och naturvetenskap alltsedan mitten av 1960-talet. Dessa studier genomförs under namnet TIMSS alltsedan 1995. TIMSS har sedan 1995 bedrivits som en studie med uttalat syfte att studera förändring i kunskaper och färdigheter över tid, avseende matematik och naturvetenskap. I studien ingår elever i årskurs 4 och årskurs 8. Mer oregelbundet studeras även elevers kunskaper i matematik och naturvetenskap för elever i gymnasieskolan. Denna del av studien benämns TIMSS Advanced.²¹ Sverige har tidigare valt att inte delta i TIMSS 1999 samt TIMSS-R Video Study. Vad beträffar TIMSS 2003 så deltog Sverige endast i den del av studien som avsåg elever i årskurs 8 och avstod från att delta i den delen av studien som avsåg elever i årskurs 4. I TIMSS 2007 deltar Sverige sedan november 2005 i den del av studien som avser elever i årskurs 4, och Skolverket har alldeles nyligen²² också beslutat att Sverige ska delta i TIMSS 2007 årskurs 8, liksom i TIMSS Advanced 2008 – den senare studien avser matematik och fysik på gymnasial nivå. TIMSS är planerad att upprepas vart fjärde år.

¹⁸ Progress in International Reading Literacy Study.

¹⁹ Third International Mathematics and Science Study. De elevpopulationer som IEA studerade är elever ur vad IEA kallar population 1 (9-åringar), population 2 (13-åringar) samt population 3 (vissa elever ur gymnasieskolans sista år). Här deltog Sverige med elever ur population 2 och 3.

²⁰ Från och med 2003 ändrades betydelsen av TIMSS, vilket nu skall uttydas som ”Trends in International Mathematics and Science Study”. Studien bytte också populationsdefinition, och utnyttjar hädanefter en årskursbaserad definition.

²¹ TIMSS Advanced genomfördes för första gången 1995, och för närvarande planeras studien att upprepas under 2008 (TIMSS Advanced 2008).

²² Verksledningsbeslut 2006-08-21.

IEA planerar även att genomföra en internationell jämförande studie om unga elevers kunskande, attityder och engagemang i demokratifrågor, ICCES.²³ ICCES huvudstudie planeras till år 2009. Målet med ICCES är att undersöka hur unga människor, elever i årskurs 8, är förberedda på sin roll som medborgare, i respektive land. Studien bygger på IEA:s tidigare studie om medborgerlig utbildning CIVED²⁴ 1999/2000, i vilken Sverige deltog. Skolverket har nyligen beslutat att Sverige ska delta också i ICCES.²⁵ Studien syftar till att kunna rapportera om elevers utbildningsresultat i ett test om dels konceptuell förståelse och dels medborgerliga kompetenser. I ICCES finns det en ambition att göra vissa begränsade analyser av förändringar hos elevers kunskande över tid, och därför ska vissa uppgifter från CIVED upprepas. Till skillnad mot exempelvis TIMSS och PIRLS finns ingen planering att ICCES/CIVED ska återkomma regelbundet.

Under 1990-talet tillkom ytterligare en aktör på arenan för internationella jämförande kunskapsmätningar, OECD, som sedan länge arbetat med internationella indikatorer på utbildningsprestationer. Under det senaste decenniet har OECD initierat flera jämförande internationella kunskapsstudier. Under 1990-talets början mättes vuxnas läsfärdigheter²⁶ i ett antal OECD-länder, och OECD har fortsatt driva olika studier av vuxnas kompetenser. OECD har där- efter även drivit studier av 15-åringars läsförståelse och kunskaper inom matematik och naturvetenskap i projektet PISA.²⁷ Första studien genomfördes år 2000. Därefter har PISA upprepats med tre års intervall varvid fokus år 2000 låg på läsfärdighet ("reading literacy"), år 2003 på matematik ("mathematical literacy") och år 2006 på naturvetenskap ("scientific literacy"). Detta medför att det enskilda ämnet, exempelvis matematik, har en periodicitet på vart nion- de år i form av en fullskalig undersökning (de två andra komponenterna ingår dock, i mindre skala vid samtliga undersökningstillfällen). Sverige har deltagit i alla dessa tre PISA-undersökningar.

En annan nytillkommen aktör på arenan för internationella jämförande studier är Europeiska Gemenskapernas Kommission, efter att det vid Euro- peiska rådets möte i Barcelona 2002 konstaterats att det finns ett behov att utveckla kunskaper i främmande språk i medlemsländerna.²⁸ Som en följd av detta kommer planeringen av en omfattande europeisk studie av 15-åringars

²³ International Civic and Citizenship Education Study.

²⁴ Civic Education Study.

²⁵ Verksledningsbeslut 2006-08-21.

²⁶ I IALS (International Adult Literacy Study 1994-96), deltog Sverige samt tio andra deltagande OECD- länder. Sverige har ej deltagit i den efterföljande ALL-studien (Adult Literacy and Life Skills Survey 2006)

²⁷ Programme for International Student Assessment.

²⁸ EU (2005).

kunskaper i två främmande språk att inledas under hösten. Kommissionen har föreslagit att i den första omgången ska man av praktiska skäl testa språkkunskaperna i de fem språk som är de viktigaste undervisningsspråken i unionen som helhet, dvs. engelska, franska, tyska, spanska och italienska.²⁹ Utvecklingsarbete har påbörjats men det är för närvarande svårt att bedöma när en fullskalig studie ska genomföras.

Ett svenskt deltagande i internationella kunskapsutvärderingar är ett möjligt alternativ till att genomföra periodiskt återkommande nationella utvärderingar inom ämnesområden som täcks av internationella utvärderingar, framför allt då inom matematik, naturvetenskap och läsförståelse. Detta är också en möjlighet som nämns i Jan-Eric Gustafssons underlagsrapport vilken bifogas föreliggande rapport. Det finns emellertid flera förutsättningar som ska vara uppfyllda om internationella jämförande kunskapsbedömningsstudier ska kunna ingå som delar i ett rullande nationellt utvärderingssystem. Sådana förutsättningar diskuteras mer i detalj i avsnitt 4.3. En förutsättning för att ett deltagande i de internationella studierna ska kunna ingå som en del i ett nationellt kunskapsutvärderingssystem är naturligtvis att Sverige regelbundet deltar i relevanta internationella studier.

Skolverket har tidigare i ett internt policybeslut³⁰ förespråkat restriktivitet i deltagandet i internationella jämförande studier. I det tidigare policybeslutet sägs också att det nationella behovet av tätt återkommande internationella kunskaps- och kompetensmätningar av samma slag är begränsat, och att fokus bör vara på nationella utvärderingar. Denna bedömning har dock nyligen reviderats, bland annat som en följd av det pågående arbetet med att revidera strategin för kunskapsutvärderingar, dvs. det arbete som redovisas i föreliggande rapport. Skolverket menar att om internationella studier ska kunna ingå som en del i ett nationellt system för periodiskt återkommande kunskapsutvärderingar i syfte att ge tillgång till information om kunskapsförändringar hos eleverna i den svenska grundskolan, krävs en regelbundenhet och långsiktighet i planering av och deltagandet i relevanta internationella kunskapsstudier. Skolverket har därför beslutat att revidera sin policy och sina riktlinjer för deltagande i internationella studier i enlighet med de förslag som redovisas längre fram i föreliggande rapport.³¹ Som en konsekvens härav har verket, som ovan nämnts, samtidigt också omprövat det tidigare beslutet att inte delta i TIMSS 2007 årskurs 8, liksom att även delta i TIMSS Advanced 2008 och ICCES 2009.

²⁹ Det har riktats kritik mot valet av språk i första omgången, och för närvarande är det osäkert vilka språk som skall inkluderas.

³⁰ Skolverket (2004b).

³¹ Verksledningsbeslut 2006-08-21.

2.2.4 Skolverkets insamling och redovisning av betygsstatistik och resultat på nationella prov

I ett mål- och resultatstyrt system som det svenska är det givetvis viktigt att det finns tillförlitliga resultat som anger graden av måluppfyllelse. Utbildningens mål anges i Skollagen (1 kap 2§):

Utbildningen skall ge eleverna kunskaper och färdigheter samt, i samband med hemmen, främja den harmoniska utvecklingen till ansvars-kännande människor och samhällsmedlemmar. I 4 kap 1§ anges att Grundskolans mål är att "Utbildningen i grundskolan skall syfta till att ge eleverna de kunskaper och färdigheter och den skolning i övrigt som de behöver för att delta i samhällslivet. Den skall ligga till grund för fortsatt utbildning i gymnasieskolan."

I Lpo 94 anges skolans mål i form av mål att sträva mot och mål att uppnå med formuleringar av typen "Skolan skall sträva efter att varje elev..." eller "Skolan ansvarar för att varje elev...". Kursplanerna å sin sida uttrycker målen i termer som "Skolan skall i sin undervisning i svenska sträva efter att eleven...", eller beträffande mål att uppnå "Eleven skall...".

Man kan konstatera att Skollagens mål är formulerade på en hög abstraktionsnivå, vilken sedan i större eller mindre grad konkretiseras i läroplan och kursplaner. Man kan vidare notera att målen är formulerade i termer av elevens kunskaper (om man håller sig till kunskapsmålen). Det innebär rimligen att bedömningen av skolans eller skolväsendets måluppfyllelse beträffande kunskapsmålen (vilka är de som främst är av intresse i det här sammanhanget), får anges i statistiska termer som kan gälla andel elever som inte når målen, andel elever som är behöriga till ett nationellt program, andel elever som nått målen i alla ämnen, andel elever med MVG i ett ämne, genomsnittligt meritvärde, genomsnittlig betygspoäng i ett ämne, andel som nått minst betyget Godkänd på det nationella provet, genomsnittlig betygspoäng på det nationella provet etc.

Om man hårdrar målbeskrivningarna i styrdokumenterna skulle resultatuppföljning gälla andelen elever som nått målen, dvs minst fått betyget Godkänt, men inte med vilken kvalitet det skett, dvs. hur stor andel som har betygen VG och MVG. Några sådana formuleringar finns inte i målbeskrivningarna. Därmed skulle andel MVG-betyg, meritvärde och genomsnittlig betygspoäng inte i strikt mening vara att betrakta som mått på måluppfyllelse utan mer på slag av måluppfyllelse. Om man emellertid intar en mindre strikt hållning torde det vara tämligen självklart att även kvaliteten på måluppfyllelsen är in-

tressant ur ett resultatperspektiv och att det därför är relevant att använda t.ex. meritvärde och genomsnittlig betygspoäng som resultatmått.

Betygsstatistik – Skolverkets nationella uppföljningssystem

Det som då blir intressant är att söka svar på frågan hur tillförlitliga de olika måtten på måluppfyllelse är, dvs. betygsresultat och provresultat. Detta är en svår fråga. När det gäller betygssättningen måste minst två tolkningsprocesser genomföras. För det första ska läraren tolka och skapa en innebörd i de verbala utsagor som utgör mål att uppnå och betygskriterier. För det andra ska läraren tolka elevens olika prestationer och utsagor och relatera dessa till sin tolkning av kursplanerna. Vilka kunskaper har eleven visat? Hur ska dessa kunskaper relateras till det som är skrivet i mål och betygskriterier? Har eleven kunnat vissa alla sina kunskaper? Har jag som lärare förstått vad eleven visat? När eleven visat olika grad av måluppfyllelse för olika mål, hur ska jag då sammanfatta de olika måluppfyllelsorna till ett betyg? Ger G-kvaliteter med avseende på ett mål och MVG-kvaliteter med avseende på ett annat VG som betyg? Eller endast G? Hur förhåller det sig med en G-kvalitet och två MVG-kvaliteter? Osv. Hur lärare resonerar i situationer som de beskrivna är en process som inte kan sägas vara helt klarlagd eller genomskinlig.

Empiriska studier av Skolverket och andra – exempelvis US 98,³² Skolverkets nationella kvalitetsgranskning 2000 av betygssättningen³³ och NU 03³⁴ liksom aktuell forskning på området³⁵ – har också sammantaget visat att betygssättningen är en grannlaga uppgift, och att det är tveksamt hur likvärdig den är mellan olika lärare, ämnen, skolor och till och med kommuner. Det är också osäkert hur stabil betygssättningen är över tid. Betygssättningen kan ses som en form av praxisstyrd verksamhet och praxis har en tendens att förändras över tid. Stabiliteten över tid i en betygssättning som baseras på tolkning av mål och betygskriterier kan således betvivlas.

Om man emellertid granskar meritvärdet under den tid som det nuvarande betygssystemet varit i bruk får man den bild som diagram 1 visar.

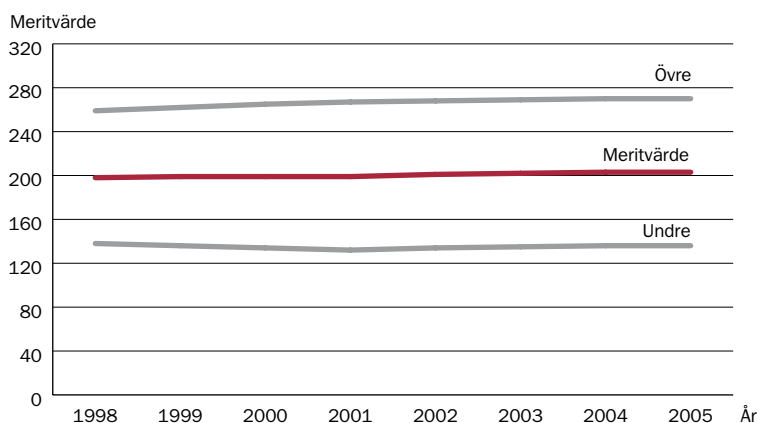
³² Skolverket (1999a), s. 124-125 och 132 och Skolverket (1999c), s. 37–40.

³³ Skolverket (2000), s. 130-178.

³⁴ Se t.ex. Skolverket (2004d), s. 94-102 och 136-137.

³⁵ Se t.ex. Selghed (2004) och Tholin (2006).

Diagram 1 Meritvärde/plus minus en standardavvikelse



Den röda linjen anger medelvärdet för meritvärdet respektive år. De grå linjerna anger meritvärdet plus/minus en standardavvikelse och ger en bild av spridningen. Man kan grovt säga att två tredjedelar av eleverna har meritvärden som ligger mellan de streckade linjerna.

Av diagrammet framgår att meritvärdet legat mycket nära 200 under hela perioden och standardavvikelsen har ökat marginellt från 60,6 våren 1998 till 67,1 våren 2005. Kan då detta tas som intäkt för att kunskapsnivån och måluppfyllelsen varit i stort sett konstant under den aktuella perioden? Svaret måste bli nej eftersom det saknas jämförelseunderlag. Kunskaperna kan ha gått ner samtidigt som betygspraxis har förändrats i motsatt riktning så att bedömningen mildrats i takt med nedgången i kunskaper. Å andra sidan kan man tänka sig motsatsen, dvs. att kunskaperna förbättrats men att bedömningen blivit strängare i motsvarande grad. Eller som ett tredje alternativ att både kunskaper och bedömning legat fast. Vilket av de tre alternativen som är mest troligt är svårt att avgöra.

En jämförelse med internationella studierna PISA och TIMSS ger delvis motstridiga resultat. Enligt PISA har ingen nämnvärd förändring skett mellan 2000 och 2003, vilket stöder figuren. Å andra sidan visar TIMSS en mycket kraftig nedgång mellan 1995 och 2003 (vilken möjligen kan ha skett 1995–2000). Så hur förstå resultaten?

Andra resultat från uppföljningssystemet gör inte bilden klarare. Andelen elever som når behörighet till nationella program pendlar mellan 89 och 90 procent och andelen elever som når målen i alla ämnen ligger runt 75 procent. Men om detta är uttryck för att måluppfyllelsen är konstant eller beror på kompenserande processer av olika slag är oklart.

Det man kan konstatera är att den *administrativa* eller manifesta (observerbara) betygsnivån i grundskolan inte förändrats nämnvärt mellan 1998 och 2005. Med administrativ betygsnivå menas då den registrerade betygsnivån som framträder i den uppföljande statistiken, men hur denna nivå förhåller sig till elevernas faktiska kunskaper är okänt. Man skulle kunna säga att förändringen av den reala eller latent (ej observerbara) kunskapsnivån är okänd.

Statistik över resultat på nationella prov – det nationella provsystemet

Är då bedömning av måluppfyllelse över tid med hjälp av nationella prov mer stabil och trovärdig än bedömning baserad på uppföljningsdata. Syftena med det nationella provsystemet är enligt regeringens uppdrag från december 2004 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 samt att
- ge underlag för en analys av i vilken utsträckning kunskapsmålen nås på skolnivå, på huvudmannanivå och på nationell nivå.

Den första punkten är av övergripande natur medan punkterna två och tre i huvudsak är tänkta som stöd till lärarna i deras pedagogiska arbete. Punkterna fyra har ett likartat syfte, dvs. att ge lärarna förebilder för hur elevarbeten kan bedömas i relation till mål och betygskriterier. Det nationella provsystemet är alltså inte främst tänkt att tillhandahålla ”mätinstrument” i traditionell mening utan mer att bistå med förebildliga exempel på måluppfyllelse av olika kvalitet. Den femte punkten slutligen är den som har mest relevans i det här sammanhanget.

Det kan vara på sin plats att påpeka att det är skillnad mellan ”det nationella provsystemet” och ”nationella prov”. Med det förstnämnda avses hela systemet med diagnostiska material, provbanksmaterial, kunskapsutvärderande material (av det slag som diskuteras i den här utredningen) samt nationella prov. Med nationella prov avses, för grundskolans del, endast de prov som är obligatoriska i år 5 och år 9. Den fråga som aktualiseras här gäller om de nationella proven kan ”ge underlag för en analys av i vilken utsträckning kunskapsmålen nås på skolnivå, huvudmannanivå och på nationell nivå.” Läs oss börja med skolnivån.

Skolverket har i rapporten *Det nationella provsystemet i den målstyrda skolan*³⁶ visat att relationen mellan betyg på nationella prov och ämnesbetyg

³⁶ Skolverket (2004a).

varierar i hög grad mellan olika skolor. Diagram 2 och 3 kan få illustrera förhållandet i svenska år 9 våren 2004.

Diagram 2 Andel elever med högre ämnesbetyg än provbetyg, Sv 04 (genomsnitt 18 %). Skolor med 1 222 med > 15 elever)

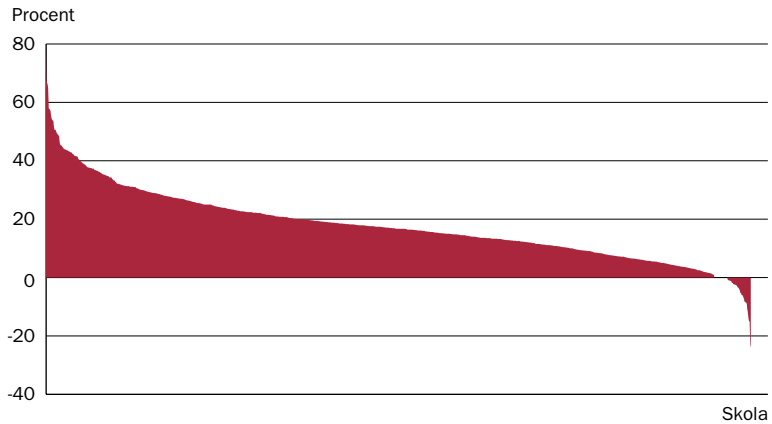


Diagram 3 Sv 04 (97 114 elever)

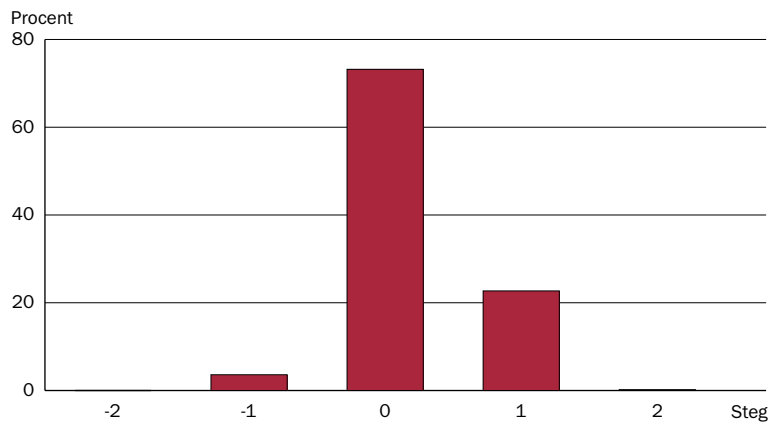
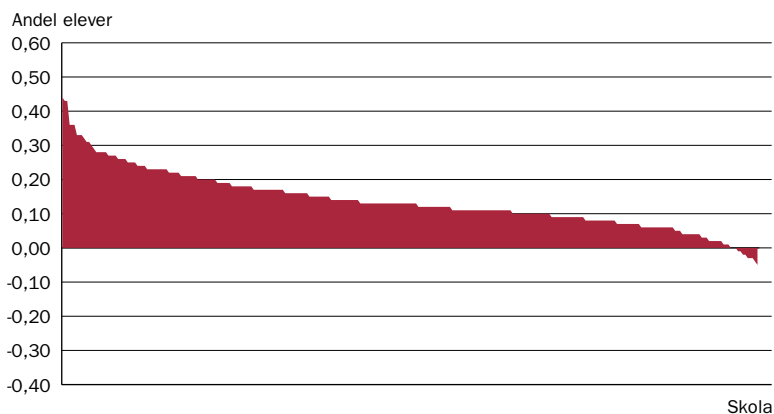


Diagram 3 anger hur stor andel av eleverna som har samma ämnesbetyg som provbetyg (0), ämnesbetyg som ligger ett steg under (-1) eller över (1) provbetyget. Av bilden framgår att drygt 73 procent har samma ämnesbetyg och provbetyg medan drygt 22 procent har högre ämnesbetyg och drygt 3 procent lägre. Man kan säga att nettoeffekten blir att 19 procent av eleverna har ett högre ämnesbetyg än provbetyg på nationell nivå (gäller alla elever som har båda betygen).

Delar man upp resultaten på skolnivå (skolor med fler än 15 elever) blir resultatet som i diagram 2. Den vertikala axeln anger andelen elever (netto) som har högre ämnesbetyg än provbetyg och på den horisontella finns de 1 222 skolor som ingår. Man kan se att variationen mellan olika skolor är betydande. Som beskrivits i ovan nämnda rapport kan det finnas rimliga förklaringar till detta, men det blir ett problem att avgöra om ämnesbetyg eller provbetyg ska användas som mått för att avgöra i vilken utsträckning kunskapsmålen nåtts på skolnivå. Beroende på vilket mått som väljs kommer skolorna att få mycket olika resultat. En del framstår som särskilt framgångsrika i relation till andra om ämnesbetygen används som mått, medan bilden kan bli den omvända om provbetygen används. Det är alltså svårt att på skolnivå avgöra i vilken utsträckning de nationella proven kan användas för att avgöra graden av måluppfyllelse. Det mest rimliga torde vara att se den samlade lärarkårens bedömning som den mest relevanta. Det skulle i sin tur innebära att för det aktuella provet är den korrekta nivån att utgå ifrån den som representerar genomsnittet av avvikelse, dvs, att cirka 18–19 procent av eleverna bör ligga ett betygssteg över det betyg de hade på provet.

Hur ser då bilden ut på huvudmannanivå? Diagram 4 visar samma prov och betyg som ovan men uppdelade på kommuner.

Diagram 4 Andel med högre ämnesbetyg än provbetyg Sv-04, kommuner (m=0,20)

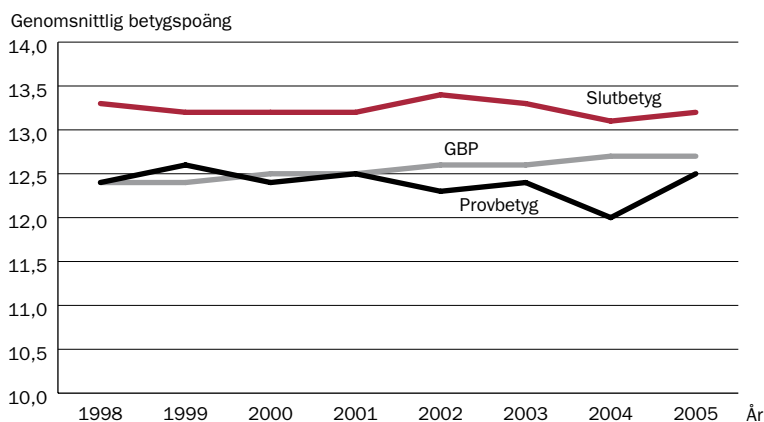


Det som kanske förvånar i diagrammet är att det även på kommunnivå förekommer betydande skillnader mellan olika kommuner. Med samma resonemang som för skolorna kan man ställa sig tveksam till i vilken utsträckning nationella prov kan användas för att följa upp måluppfyllelsen på huvudmannanivå. Även här blir frågan om man vid uppföljning ska tro mer på provbetyg eller ämnesbetyg. I fallet med huvudmännens resultat är de förklaringar som

framstår som rimliga på skolnivå inte lika rimliga på kommunnivå. Slutsatsen både när det gäller skolnivå och huvudmannanivå torde bli att det inte är provresultaten i sig som är intressanta utan hur de används vid betygssättningen och hur de lokala aktörerna resonerar när de sätter de betyg de sätter. Provresultaten lämpar sig således bättre för intern uppföljning och analys än för extern.

Hur är då läget på nationell nivå? Kan de nationella proven i det fallet ge en bättre bild av måluppfyllelsen än lärarnas betyg och ger de en säkrare bild av kunskapsutvecklingen över tid. Ett par diagram kan få illustrera frågorna. Det första visar den genomsnittliga betygspoängen på det nationella provet i svenska, den genomsnittliga betygspoängen för ämnesbetyget (för de elever som har ett provresultat) samt den genomsnittliga betygspoängen för de ämnen som ingår i meritvärdet (kallat GBP i diagrammet).

Diagram 5 Genomsnittlig poäng för svenska

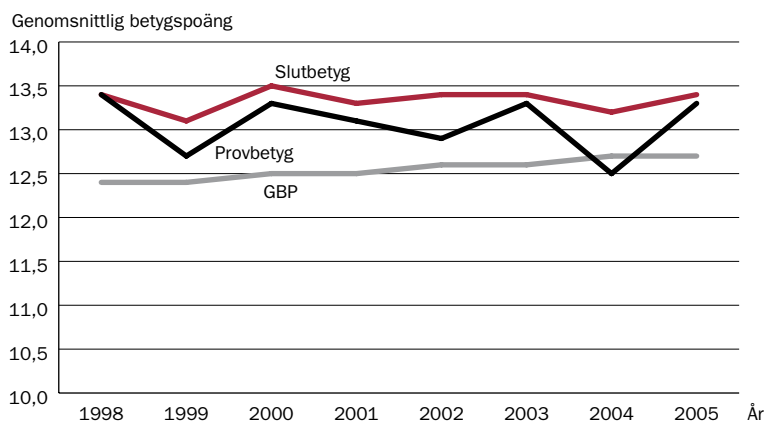


Av diagrammet framgår att ämnesbetyget (slutbetyg) för alla år ligger över provbetyget.³⁷ Man kan också notera att betygen i svenska ligger över den genomsnittliga betygspoängen för samtliga ämnen. Man kan slutligen notera att relationen mellan provbetyg och ämnesbetyg varierar mellan olika år och att det är svårt att se någon tydlig tendens i denna variation.

³⁷ För åren 2002–2004 sattes inga sammanfattande provbetyg så de använda värdena är medelbetyget av de tre ingående delproven. Samma sak gäller för proven i engelska.

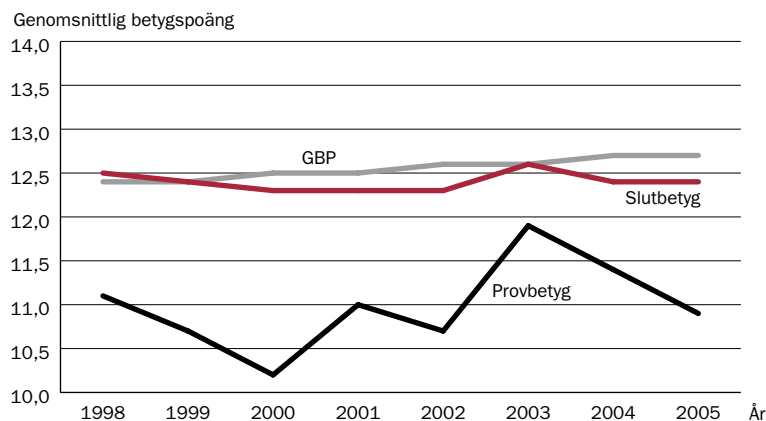
I engelska är bilden enligt diagram 6.

Diagram 6 Genomsnittlig poäng för engelska



Här kan man konstatera att ämnesbetyg och provbetyg ligger närmare varandra och att båda ligger över betygsgenomsnittet för de ämnen som ingår i meritvärdet. Man kan också notera att den genomsnittliga betygs-poängen varierar betydligt mer för provbetyget än för ämnesbetyget även om det finns en viss följsamhet. I matematik slutligen ser bilden ut på följande sätt.

Diagram 7 Genomsnittlig poäng för matematik



Här ligger den genomsnittliga provbetygs-poängen på det nationella provet i matematik långt under den genomsnittliga betygs-poängen för såväl matematik som för de 16 ämnen som ingår i meritvärdet. Det som också är anmärknings-

vårt är den stora variationen mellan olika år för provresultaten.³⁸ Man kan också notera att ämnesbetygen förefaller i stort sett opåverkade av provbetygens variationer. Eftersom det handlar om stora populationer är det inte troligt att kunskaperna variera särskilt mycket mellan olika årskullar. Genomsnittsvärdet för lärarnas betygssättning förefaller därför betydligt tillförlitligare som mått på den nationella måluppfyllelsen än provbetygen. Detta är emellertid resultat på aggregerad nivå och hur det förhåller sig med enskilda lärares betygssättning är en helt annan fråga.

Man skulle kanske vänta sig att de prov som har en mer strikt bedömning via poängsättning skulle visa mer stabila resultat, men bilden är snarast den omvända. Proven i matematik som är poängbaserade visar den största årliga variationen, därefter engelska som har en viss del poängsatta delar, medan proven i svenska, som i störst utsträckning baseras på lärares mer övergripande bedömningar, visar den minsta variationen. Man skulle därmed kunna säga att proven i svenska är de som i störst utsträckning liknar lärares allmänna bedömning av sina elevers prestationer och den bedömning som torde vara minst styrd av strikta bedömningsanvisningar. Därmed är det också rimligt att det är i svenska som provbetyg och ämnesbetyg visar den största följsamheten även om nivåskillnaden är förhållandevis stor. Och fortfarande finns stor variation på skol- och kommunnivå som tidigare visats.

Slutsatser

Slutsatsen av ovanstående redovisning blir att de nationella proven inte lämpar sig för uppföljningsändamål på nationell nivå. Variationen i resultat är alltför stor för att vara trovärdig och har sin grund i de svårigheter som är förknippade med att i förväg på ett tillförlitligt sätt bestämma vilka prestationer i form av poäng eller annat som ska krävas för olika betyg. En betygsgräns som flyttas någon eller några poäng åt ena eller andra hållet kan få stora konsekvenser i genomsnittlig betygs-poäng. Resultaten ska inte tolkas som att de nationella proven inte är välkonstruerade utan de har sin grund i att proven inte har som primärt syfte att mäta kunskapsutveckling över tid och de gör inte heller anspråk på att kunna täcka in alla mål som ingår i ämnet.

Att de nationella proven inte är tillräckliga som underlag för nationell uppföljning av måluppfyllelsen innebär emellertid inte att insamlingen av provresultat inte är meningsfull. Insamlingen behövs för det lokala utvecklingsarbetet och även för att analyser av det slag som diskuteras här ska kunna göras. Även insamling av betyg måste naturligtvis göras i uppföljningssyfte även om det

³⁸ Resultaten för 2003 är osäkra eftersom provet hade spridits på nätet och cirka 40 procent av skolorna avstod från att rapportera resultat. Resultaten tyder på att de som avstod har något lägre betygs-poäng än de som deltog i insamlingen.

också i det fallet finns anledning tvivla på i vilken utsträckning de administrativa betygen representerar den reala kunskapsnivån. På nationell nivå torde dock de betyg lärarna satt vara ett mindre dåligt mått på nationell måluppfyllelse än resultaten på de nationella proven.

Den sammanfattande slutsatsen blir att om det femte syftet i provuppdraget ska kunna uppfyllas krävs en annan typ av instrument än de nationella proven. Det nationella uppföljningssystemet kan inte heller på ett tillförlitligt sätt fylla den uppgiften, utan de möjligheter som finns är att antingen använda de internationella instrument som finns för att mäta kunskapsförändring över tid eller att utveckla egna sådana instrument. Vad det innebär kommer att beskrivas längre fram i föreliggande utredningsrapport. I fallet med de internationella studierna kan man dock redan här konstatera att de inte är utvecklade med svenska måldokument som grund. Bedömningar behöver därför göras av hur väl de internationella mätinstrumenten överensstämmer med svenska måldokument. Dessa belyses längre fram i föreliggande rapport.

2.2.5 Skolverkets utbildningsinspektion

Skolverkets utbildningsinspektion är, vad gäller bedömning av elevers kunskaper och kunskapsutveckling, hänvisad till betygsresultat, resultat på nationella prov samt lokal dokumentation. Underlaget för inspektionens bedömningar hämtas, enligt Skolverkets aktuella frågeguide³⁹ från skolblad, skolans kvalitetsredovisning, arbetsplan och betygs kataloger, sammanställningar av provresultat, andra sammanställningar från t.ex. arbetslag, intervjuer med skolledning, förskollärare och lärare samt i viss mån från lektionsbesök eller besök i den pedagogiska verksamheten. Skolverkets inspektörer inhämtar bl.a. uppgifter om skolans resultat på nationella ämnesprov i årskurserna 5 och 9, betygsresultat i årskurs 9, andelen elever som ej uppnått målen osv. – dvs. i princip samma uppgifter som också samlas in nationellt och som bl.a. redovisas i Skolverkets informationssystem SIRIS. För övriga kunskapsresultat – framför allt före år 8 – liksom för variationer i betygs- och provresultat är inspektörerna enligt frågeguiden hänvisade till ”skolans egna bedömningar”.

Med hjälp av dessa uppgifter används sedan följande bedömningsunderlag för Skolverkets bedömning av resultat vid en enskild grundskola:⁴⁰ ”Skolan kan visa på elevers kunskapsutveckling i de tidigare årskurserna i samtliga ämnen. Alla/flertalet elever når målen att uppnå i årskurs 5 i samtliga ämnen. Alla/flertalet elever når målen för årskurs 9 enligt de nationella ämnesproven.

³⁹ Skolverket (2005).

⁴⁰ Skolverket (2006a; Urval har i föreliggande redovisning gjorts av de kriterier som uttryckligen avser kunskapsresultat).

Alla/flertalet elever når målen att uppnå i årskurs 9 i samtliga ämnen. Alla/flertalet elever med utländsk bakgrund når målen att uppnå i årskurs 9 i samtliga ämnen. Flickor och pojkar når likvärdiga resultat... .. Eleverna utvecklar de kunskapskvaliteter som uttrycks i läroplanens mål att sträva mot”.

Sammanfattningsvis visar ovanstående sammanställning, att det underlag som tas in av Skolverkets utbildningsinspektion är befästat med samma problem som den nationella statistiken över betyg och resultat på nationella prov. Även om värdefull information i form av t.ex. intervjuer och klassrumsobservationer – vilka i mer eller mindre hög grad kan kasta ljus över eller förklara de lokala betygs- och provresultaten – samlas in i samband med Skolverkets inspektioner, ger denna information således knappast underlag för nationella bedömningar av svenska elevers kunskapsresultat och kunskapsutveckling.

2.2.6 Slutsatser

Sammanfattningsvis kan konstateras att Skolverkets samlade verksamhet inom områdena Uppföljning, Utvärdering och Inspektion är omfattande, men f.n. saknar en systematisk verksamhet vars primära syfte och metodik är inriktad på nationell bedömning av elevers kunskapsresultat och hur dessa utvecklas över tid. De tidigare stora satsningarna i form av ”nationella utvärderingar” – NU 92, UG 95 och NU 03 – har haft utgångspunkt i tidigare läroplaner och kursplaner⁴¹ och det är inte realistiskt att fortsätta upprepa dessa i samma form och omfattning. Skolverkets hittillsvarande policy för deltagande i internationella kunskapsmätningar har präglats av en restriktiv hållning, och utgångspunkten för deltagandet har varit att erhålla internationella referenspunkter snarare än att de ses som en viktig del av en nationell strategi för resultatbedömning. Såväl det nationella uppföljningssystemet, det nationella provsystemet som utbildningsinspektionen förlitar sig på statistik i form av insamlade resultat av betyg och från nationella prov. De nationella proven och betygen syftar i första hand till lokal bedömning av individer (elever) vid en given tidpunkt. Flera studier av Skolverket liksom aktuell forskning har sammantagna visat att det är tveksamt hur likvärdig betygssättningen är mellan olika ämnen, lärare, skolor och kommuner. Det är också osäkert hur stabil betygssättningen är över tid. Betygs- och provstatistik har därmed – och som ovan visats – begränsningar i sina möjligheter att som tjäna underlag för nationell jämförelse av elevers kunskaper och kunskapsutveckling över tid liksom jämförelser inom landet mellan t.ex. olika skolor eller typer av skolor.

⁴¹ Utgångspunkten för genomförandet av NU 03 var att genomföra en s.k. repeat” på NU 92, även om en del förändringar efter hand befanns nödvändiga ett göra i såväl kunskapsbedömningsinstrument som enkäter, jämfört med 1992 års instrument

**Skolverkets uppdrag
till professor
Jan-Eric Gustafsson**

3 Skolverkets uppdrag till professor Jan-Eric Gustafsson

3.1 Skolverkets hantering och avgränsning av regeringens uppdrag

En rullande utvärdering av samtliga ämnen i grundskolan, av det slag som avses i regeringens uppdrag, skulle innebära att en samhällelig infrastruktur skulle behöva byggas upp i syfte att fungera i ett långårigt perspektiv. Skolverket ansåg därför det vara väsentligt att frågan utreddes ordentligt som underlag för beslut om genomförande och ett mer operationellt arbete. I detta syfte gav verket 2005-05-19 i uppdrag till professor Jan-Eric Gustafsson vid Institutionen för pedagogik och didaktik, Göteborgs Universitet att utarbeta ett förslag till övergripande ramverk för en rullande uppföljning/utvärdering av elevgruppens kunskapsutveckling.

Skolverkets uppdrag till Jan-Eric Gustafsson avgränsades, i enlighet med regeringens uppdrag 2004-12-22, till grundskolan. Vidare angavs att underlaget skulle ge förslag på lämplig utformning, samt belysa såväl potentialen i som svårigheterna med, ett sådant system. En viktig utgångspunkt har varit att erfarenheter från genomförandet av NU 03 (och tidigare s.k. nationella utvärderingar) skall beaktas vid utvecklingen av förslaget till rullande utvärderingsmodell. Den första etappen av arbetet innebar en genomgång av erfarenheter kring utformning av rullande nationella utvärderingssystem i några andra länder, främst USA men även Nederländerna och Nya Zeeland. I Skolverkets uppdrag preciserades att utvärderingsmodellen skulle ha följande egenskaper:

- Ge beskrivningar av kunskapsnivåer och utveckling av kunnande över tid inom samtliga ämnen i grundskolan, dels för tidigare år, dels för senare år, enligt ett rullande system där olika ämnen eller ämneskombinationer återkommer med viss regelbundenhet.
- Modellen skall möjliggöra (rimligt) stabila jämförelser av elevers kunskaper och färdigheter även efter att läro- och kursplaner reviderats. Det är ett önskemål att det skall vara möjligt att exempelvis efter större kursplaneförändringar göra erforderliga förändringar i de utnyttjade instrumenten med bibehållna möjligheter till jämförelser över tid.
- Vid varje årligt utvärderingstillfälle skall såväl elevers kunskaper och färdigheter i de aktuella ämnena som relevanta bakgrundsfaktorer för elever, lärare och skolor kartläggas.

- Resultaten från utvärderingen skall i första hand presenteras på nationell systemnivå. Det är dock önskvärt att skolor/skolhuvudmän ges möjlighet till att på eget initiativ genomföra lokala utvärderingar vilkas resultat kan relateras till de nationella.
- Översiktliga resultat från varje utvärderingstillfälle skall presenteras inom ett relativt kort tidsspänn efter att utvärderingen genomförts. Denna resultatpresentation skall innehålla nivåer och trend i kunskaper och färdigheter samt relaterade till vissa centrala bakgrundsvariabler. Databasuppbyggnad och rapporteringsrutiner ska dock vara så flexibla att de också medger fortsatta fördjupade och mer förklaringsinriktade analyser av insamlade resultatdata.
- Datainsamling, datahantering och resultatredovisning skall ske på ett så kostnadseffektivt sätt som möjligt.

Jan-Eric Gustafsson har 2006-05-16 avrapporterat sitt uppdrag i form av en särskild underlagsrapport *Ramverk för ett system för uppföljning av kunskapsutvecklingen i grundskolan*. Underlagsrapporten biläggs föreliggande avrapportering. En del av underlagsrapporten omfattar redovisning av liknande system i USA (NAEP⁴²), Nederländerna (PPON⁴³), England samt Nya Zeeland (NEMP⁴⁴), liksom motsvarande erfarenheter från de internationella komparativa studierna av elevers kunskaper och kunskapsutveckling inom vissa områden – OECD:s program PISA samt IEA-undersökningarna TIMSS och PIRLS.

Olika versioner av underlagsrapporten har diskuterats vid två brett upplagda forskarsymposier 2005-09-15 och 2006-01-26. Vid dessa symposier medverkade bl.a. expertis från samtliga övriga nordiska länder samt ett antal framstående svenska pedagogiska forskare, och åtskilliga värdefulla synpunkter, kommentarer och förslag lämnades.⁴⁵ De båda forskarsymposierna har också vid två tillfällen legat till grund för särskilda redovisningar och diskussioner i Skolverkets verksledning.

Slutligen har Skolverket kostnadsuppskattat de förslag som Jan-Eric Gustafsson lämnat. Till grund för dessa kostnadsuppskattningar, vilka redovisas i föreliggande avrapportering, har bl.a. legat verkets erfarenheter av såväl framställningen av nationella prov som deltagande i olika internationella studier.

⁴² National Assessment of Educational Progress.

⁴³ Periodieke Peiling van het Onderwijs Niveau.

⁴⁴ National Education Monitoring Project.

⁴⁵ Symposierna har dokumenterats i form av utförliga minnesanteckningar.

I följande avsnitt (3.2) sammanfattas Gustafssons underlag och förslag utan kommentarer, tillägg eller ändringar från Skolverkets sida. Verkets egna bedömningar, prioriteringar och förslag med utgångspunkt från nämnda underlagsrapport och de synpunkter vi inhämtat på den, redovisas sedan i avsnitt 4.

3.2 Sammanfattning av Jan-Eric Gustafssons underlag och förslag

3.2.1 Inledning och bakgrund

Rapporten redovisar ett uppdrag från Skolverket kring utformning av ett system för rullande utvärdering av elevers kunskapsutveckling. Uppdragets genomförande har styrts av direktiv dels i ett regeringsuppdrag till Skolverket, dels genom uppdragsdirektiv från Skolverket.

3.2.2 Erfarenheter från "National Assessment of Educational Progress" i USA

I slutet av 1960-talet infördes i USA under ledning av Ralph Tyler en modell för nationell utvärdering som var innovativ på många sätt. Medan de standardiserade testsystem som var i bruk vid denna tid fokuserade på att ge säkra resultat för enskilda elever var huvudsyftet med "National Assessment of Educational Progress" (NAEP) att ge en bestämning av nivån på kunskaper och färdigheter på nationell nivå. Eftersom elevresultaten i sig inte var intressanta kunde urval göras både av elever och uppgifter (s.k. matrissampling), vilket innebar att olika elever besvarade delvis olika uppgifter. Härigenom kunde många uppgifter användas utan att den arbetsmässiga belastningen på den enskilda eleven blev alltför stor. Det faktum att urvalen av elever var relativt små gjorde det möjligt att inte endast använda flervalstuppgifter utan även uppgifter med elevproducerade svar som krävde bedömning.

Erfarenheterna av NAEP under 1970-talet visade att programmet kunde ge 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 beslutsfattare. 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 ny design som baserades på metodologiska landvinningar inom mätläran, vilka gjorde det möjligt

att föra över resultat på en och samma mätskala även om eleverna besvarat delvis olika uppgiftsuppsättningar. Härigenom kunde man dra nytta av matrisamplingssteknikens fördelar, samtidigt som nackdelarna förknippade med redovisning på uppgiftsnivå eliminerades. Dessa fördelar hade dock ett pris i form av en betydligt ökad komplexitetsgrad i genomförande av utvärderingarna. För att kunna få ut den information som efterfrågades ur data var det också nödvändigt att utveckla nya, mycket komplexa, analystekniker.

De nyutvecklade metoderna visade sig vara mycket användbara för att skapa överblickbara resultat, och de gjorde det möjligt att undersöka förändringar i nivån av kunskaper och färdigheter över tid. Erfarenheterna visade dock också att den komplicerade tekniken kunde leda till inkorrekta bestämningar av förändringar, varför man valde att som ett parallellt system även fortsättningsvis använda samma uppgifter som användes redan vid 1970-talets början.

Under 1980-talet växte intresset för att följa upp och utvärdera delstatliga utbildningsreformer, liksom att jämföra olika delstaters resultat. Under början av 1990-talet utvidgades NAEP på frivillig bas till delstatsnivån i ett begränsat antal ämnen. I ett senare steg blev utvidgningen till delstatsnivån permanent. Från år 2001 har NAEP ålagts en ny uppgift, nämligen att vara en kontrollinstans för de utvärderingssystem på elevnivå som delstaterna har ålagts att införa. Detta innebär en kraftig förändring av NAEPs roll i det att det får en tydlig plats inom det ansvarsutkrävande systemet. Beslutet innebär också att den tidigare frivilligheten för delstaterna att delta upphör.

En stor mängd utvärderingar har genomförts av NAEP. Dessa är överlag mycket positiva, även om en återkommande kritik är att NAEP under sin nästan 40-åriga existens 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 Tyler's ursprungliga idéer om en "national assessment" har förfuskats genom en insnävning av uppgiftstyper, stark fokusering på vissa ämnesområden, abstrakt rapportering, och en ökad betoning av ansvarsutkrävande funktioner.

3.2.3 Erfarenheter från andra nationella utvärderingssystem

Internationellt komparativa studier av elevers kunskaper och färdigheter, som exempelvis de som genomförs av IEA och OECD, har inte i första hand som syfte att vara nationella utvärderingssystem, men för många länder tjänar sådana studier samma funktioner som NAEP gör i USA. De internationella studierna står också inför likartade metodologiska utmaningar som de nationella utvärderingssystemen, nämligen att beskriva nivå och förändring över tid i kunskaper och färdigheter på utbildningssystemnivå.

Tyngdpunkten i IEAs nuvarande verksamhet ligger på genomförande av regelbundet återkommande studier, dels av läsfärdighet (PIRLS, "Progress in International Reading Literacy Study"), dels av matematik och naturvetenskap (TIMSS, "Trends in Mathematics and Science Study"). Syftet är både att jämföra länders resultat och att studera förändring av resultat inom länder över tid. Både TIMSS och PIRLS bygger på de metoder som utvecklats inom NAEP.

Även OECDs undersökning PISA ("Programme for International Student Assessment"), har det dubbla syftet att jämföra länder och studera förändring över tid. I PISA undersöks vart tredje år läsförståelse, matematik och naturvetenskap. Även PISA använder liknande metoder som de som utvecklats inom NAEP.

År 1986 påbörjades det holländska nationella utvärderingsprojektet PPOON ("Periodieke Peiling van het Onderwijs Niveau").⁴⁶ Det uttalade motivet bakom införandet av PPOON var att ge alla aktörer i det starkt decentraliserade holländska skolsystemet en empirisk grund för sitt agerande.

I sammanfattning kan PPOON sägas ha genomgått fyra utvecklingsfaser. Den första fasen inleddes med en undersökning av aritmetik bland 12-åringar och utvidgades sedan att omfatta fler ämnen och upprepade mätningar. En andra fas inleddes 1997, då mätningarna fokuserades på aritmetik och modersmål. År 2003 inträffade en tredje, kortlivad, fas då mätningar i flera ämnen återupptogs eller planerades. Den senaste tiden har dock systemets legitimitet alltmer minskat. En förklaring till detta är att offentlig redovisning av resultat på andra prov som genomförs på individnivå blivit vanliga. En ytterligare förklaring är att det finns stort behov av förklaringar till funna resultat och diskussion kring möjliga handlingsalternativ. PPOON är dock inte utformat på ett sådant sätt att det ger underlag för mer fördjupade analyser av orsakerna till de erhållna resultaten.

Även England har vad som kan betraktas som ett nationellt utvärderings-system, men det är inte urvalsbaserat utan individfokuserat. Basen för det engelska systemet utgörs av en omfattande nationell läroplan med preciserade målbeskrivningar och prov för att mäta uppnående av dessa mål i olika åldrar (7, 11 och 14 år). Ett syfte med proven är att de ska vara diagnostiska och visa på olika elevers starka och svaga sidor. Resultaten sammanställs dock också i syfte att visa resultat på skol- och distriktsnivå. Det engelska systemet har mött mycket kritik och har ifrågasatts av flera olika grupper. I synnerhet gäller detta proven för 7-åringar.

⁴⁶ Skolverket har inhämtat kompletterande information om PPOON via besök på plats i Nederländerna år 2005.

År 1995 startade på Nya Zeeland ett nationellt utvärderingssystem med beteckningen National Education Monitoring Project (NEMP).⁴⁷ NEMP är på ett tydligt sätt inspirerat av Tylers ursprungliga idéer om utformningen av NAEP.

Analys och redovisning av resultaten sker på uppgiftsnivå. Efter varje tillfälle publiceras ungefär 2/3 av uppgifterna, medan återstoden återanvänds i nästa utvärderingscykel för att studera förändring över tid. Rapportering görs också för olika undergrupper. Ett utmärkande drag hos NEMP är att projektet i stor utsträckning involverar yrkesverksamma lärare, både för att man vill utnyttja den kompetens lärarna besitter, och för att lärarna måste vara delaktiga för att NEMP ska kunna förbättra undervisningen.

De nationella utvärderingssystemen uppvisar stora skillnader. Vissa är ansvarutkrävande (England, och i ökande utsträckning NAEP), medan andra är utvecklingsorienterade (NEMP). Detta återspeglar också skillnader i syften och olika prioriteringar vad avser validiteten i beskrivningarna av kunskaper och färdigheter,

3.2.4 Beskrivning av kunskaper och färdigheter – metodfrågor

Det finns stora metodologiska utmaningar i att beskriva nivån av kunskaper och färdigheter på nationell nivå, och då särskilt att studera förändringar över tid. Dessa avser bland annat validitet, reliabilitet, och generaliserbarhet.

I validitetsbegreppet i vid mening ingår att uppmärksamma även de konsekvenser mätningarna har för individer och grupper, och för vårt sätt att tänka kring de undersökta fenomenen.⁴⁸ Ett annat viktigt validitetsproblem är i vilken utsträckning utvärderingssystemet omfattar hela bredden av mål och verksamheter.

Ett *ramverk* för en utvärdering definierar huvudstrukturen av kunskaper och färdigheter inom ett visst ämnesområde, anger vilka typer av uppgifter som ska användas, hur stor andel av det totala antalet uppgifter en viss uppgiftstyp ska utgöra, osv. Utformningen av ramverket är av fundamental betydelse för validiteten i utvärderingssystemet, varvid fyra aspekter framstår som särskilt viktiga: (1) i vilken utsträckning ramverket har acceptans och legitimitet; (2) i vilken utsträckning ramverket ger signaler som utvecklar skolverksamheten i en önskvärd riktning; (3) i vilken utsträckning ramverket återspeglar undervisningens innehåll och former; och (4) hur väl ramverket fångar upp olika aspekter av ämnesområdet.

⁴⁷ Skolverket bedömer att NEMP-systemet är av relativt stort intresse för svenskt vidkommande och avser att inhämta kompletterande information om och lärdomar av detta system.

⁴⁸ Detta vida validitetsbegrepp är framför allt förknippat med den amerikanske forskaren Samuel Messick ("consequential validity").

De uppgifter som används ska konstrueras i enlighet med ramverkets specifikation. Ofta är det dock svårt att konstruera uppgifter och bedömningsanvisningar som förmår att fånga upp djupare aspekter av elevernas förståelse. Andra erfarenheter pekar också på att det är svårt att få en mer fördjupad och nyanserad bild av elevers tänkande och förståelse om man endast förlitar sig på de metoder som är tillgängliga i storskaliga undersökningar.

En nationell utvärdering står och faller med kvaliteten i data. Frågorna kring datakvalitet är i hög grad kopplade till validitetsfrågorna, och i många fall finns ett omvänt samband mellan höga ambitioner vad gäller validitet och kvaliteten i de insamlade data. En anledning till detta är att uppgifter som kräver mer omfattande elevproduktion också tenderar att lämnas obesvarade i större utsträckning än andra uppgifter.

Trendmätning kan åstadkommas både genom att inom ramen för väldefinierade och konstanta ramverk successivt byta ut uppgifter, och genom att använda samma uppgifter i samtliga mätningar. Erfarenheterna visar dock också att resultaten är mycket känsliga för ändringar i de procedurer som används vid datainsamling varför det är angeläget att följa principen ”When measuring change, do not change the measure” oberoende av vilken av de två ansatserna man väljer.

3.2.5 Fördjupad analys, tolkning och förklaring

Ett argument för att tolkning och förklaring ska vara ett centralt syfte i ett nationellt utvärderingssystem är att beskrivningarna i sig är ointressanta om undersökningarna inte också bidrar med tolkningar och förklaringar som gör det möjligt att vidta åtgärder som leder till förbättringar.

Svårigheterna att dra slutsatser om orsaker på grundval av tvärsnittsstudier är dock stora. Under förutsättning att adekvata variabler har mätts, kan det dock vara möjligt att med hjälp av statistisk metodik kontrollera för inverkan av störande faktorer. Ett annat problem är att den typ av statistisk analys som är nödvändig för att dra kausala slutsatser från tvärsnittsdata är svår att genomföra, och i synnerhet då komplexa urvals- och mätmodeller används.

Bakgrundsvariabler har en central roll i det fördjupade analysarbetet, och information om sådana inhämtas från elever, lärare och skolläda, och ibland även från föräldrar. Sådana variabler kan användas som oberoende variabler i fördjupade analyser, och som föremål för analys och beskrivning. Många av de mest intressanta resultaten från nationella utvärderingar har avsett den bild av skola och undervisning som framträtt genom beskrivningar av olika bakgrundsvariabler. Trots den stora betydelse som tillmätts fördjupad analys och förklaring har dock utveckling av instrument för att mäta bakgrundsvariabler fått liten uppmärksamhet.

3.2.6 Rapportering och värdering av resultaten

Under förutsättning att uppgifterna är offentliga ger resultatredovisning på uppgiftsnivå en mycket informationsrik beskrivning. När man använder samma uppgift över flera tillfällen ger denna ansats också god information om förändring över tid. Redovisningar av resultat på de skalor som skapas med hjälp av den moderna mättekniken har stora fördelar ur precisionssynpunkt, och tillåter beskrivning av fördelningar. De siffervärden som presenteras kan dock vara svåra att tolka. För att underlätta tolkningen har olika tekniker utvecklats, t ex så kallade uppgiftskartor, vilka illustrerar vilka uppgifter elever på olika poängnivå kan förväntas klara.

I ökande utsträckning tillhandahålls data, resultat och uppgifter via Internet, där det med hjälp av olika verktyg också är möjligt att skapa egna analyser och rapporter.

Valet av beskrivnings- och rapporteringsmodell är nära kopplat till utvärderingens syften och till vilka typer av uppgifter som används. Det framstår dock som nödvändigt att använda sig av flera olika beskrivningsmodeller samtidigt. Utan sammanfattande, abstraherade, mått är det svårt att göra beskrivningar av förändringar över tid, eller göra fördjupade analyser av exempelvis skillnader i resultat mellan olika grupper av individer. Utan de mer innehållsorienterade uppgiftsnära beskrivningarna är det svårt att mer direkt dra nytta av resultaten i utveckling av praktisk verksamhet.

3.2.7 Överväganden och förslag avseende utformning av ett nationellt kunskapsbedömningssystem

I detta avsnitt redovisas och diskuteras utformningen av ett övergripande ramverk för ett nationellt kunskapsbedömningssystem för den svenska grundskolan.

Av Skolverkets uppdragsbeskrivning följer att systemets främsta syfte ska vara att ge information om utvecklingen över tid av kunskaper och färdigheter, varvid både nivå och spridning ska uppmärksammas. Ett andra syfte med det nationella kunskapsbedömningssystemet är att ge underlag för analyser av orsaker till förändringar i kunskapsnivåerna.

Grundläggande designfrågor

Enligt direktiven ska samtliga grundskolans ämnen omfattas av den nationella kunskapsbedömningen. Uppdraget bör också tolkas så att inte endast de lätt utvärderingsbara ämnesdelarna ska fokuseras, vilket utgör en betydande utmaning.

Det finns vissa praktiska fördelar med att fokusera kunskapsbedömningen på ett ämne i taget, men ur innehålls- och analysynpunkter finns fördelar med

att snarare strukturera kunskapsbedömningen kring kluster av ämnen som är innehållsligt närliggande. Kunskapsbedömningen bör huvudsakligen genomföras i åk 4 och åk 8, men mycket talar också för att anpassa valet av årskurs efter olika ämnes förutsättningar (t.ex. förekomst av andra prov, och antal år som ämnet studerats). En grundregel bör vara att kunskapsbedömningarna ska upprepas vart femte år, men med utrymme för viss flexibilitet.

Ett konkret, om dock tentativt, förslag är att TIMSS regelmässigt används i åk 4 och åk 8 vart fjärde år för att delvis täcka av områdena matematik och naturvetenskap. De ramverk för matematik och naturvetenskap som TIMSS bygger på sammanfaller dock inte fullt ut med de ramverk som skulle konstrueras om en motsvarande svensk undersökning skulle byggas från grunden. Jämförelser bör därför göras med ett svenskt ramverk. För de områden som är underrepresenterade i TIMSS kan då ytterligare uppgifter tillföras utvärderingen.

Mot bakgrund av diskussionen är följande gruppering av ämnen tänkbar.

Praktisk-estetiska	Främmande språk	Modersmål	Matematik/NO	SO
Bild	Engelska	Svenska	Matematik	Samhällskunskap
Hem- och konsumentkunskap	Franska	Svenska 2	Biologi	Historia
Idrott och hälsa	Spanska	Modersmål	Fysik	Religion
Musik	Tyska	Teckenspråk	Kemi	Geografi
Slöjd			Teknik	Integrerat SO
			Integrerat NO	

Denna gruppering är tentativ och placeringen av enskilda ämnen kan givetvis diskuteras.

Utformning av kunskapsbedömningarna

En distinktion görs mellan två olika ansatser till nationell kunskapsbedömning: en som betecknas som *uppgiftsorienterad*, och en som betecknas som *skalorienterad*. I den uppgiftsorienterade ansatsen ligger fokus i undersökningsdesign, uppgiftskonstruktion, analys, rapportering och den praktiska användningen av resultaten på en innehållsnära nivå, och den naturliga enheten är den enskilda uppgiften eller en grupp av likartade uppgifter. I den skalorienterade ansatsen ligger fokus på utveckling, analys och tolkning av skalor som representerar nivåer av prestationer inom olika innehållsliga domäner.

Den uppgiftsorienterade ansatsen har begränsningar när syftet är att göra trendmätningar, medan den skalorienterade ansatsen har begränsningar i vilka typer av uppgifter och bedömningssituationer som kan användas. Det är därför inte möjligt att fastslå att systemet ska utformas antingen som ett renodlat

skalorienterat eller ett renodlat uppgiftsorienterat system, utan dessa ansatser bör kombineras. Fördelarna med den skalorienterade ansatsen när det gäller trendmätning och beskrivningar av fördelningar för olika undergrupper kan då utnyttjas, och fördelarna med de mer komplexa uppgiftstyperna vad gäller mer ingående och mångfacetterade beskrivningar av elevernas kunskaper och färdigheter kan tas till vara. Dessa två ansatser bör därför kombineras till en integrerad helhet. Båda ansatserna kan förväntas förekomma inom samtliga ämnesområden, även om tyngdpunkten också kan förväntas vara olika.

Ramverket för ett visst område ska definiera de kunskaper och färdigheter som ska bedömas och ska precisera hur detta ska gå till. Ramverket ska också ange fördelningen och arten av uppgifter som ska användas för konstruktion av skalor, respektive användas på uppgiftsnivå. Ramverket bör utvecklas i en brett sammansatt grupp bestående av lärare, elever, föräldrar, skolledare, och ämnes- och läroplansexperten, där olika synsätt finns representerade och där frågor kring innehåll och uppläggning av kunskapsbedömningen utsätts för en omfattande och mångsidig diskussion.

Erfarenheter från tidigare undersökningar pekar på att det är nödvändigt att med god sannolikhet kunna upptäcka förändringar med effektstorlekar som uppgår till ca 0,10. Detta innebär att det behövs en effektiv stickprovsstorlek om 1570 elever vid var och en av de två mätningarna. På grund av att urvalet är av klustertyp, och att matrissamplingsdesign används behöver den faktiska stickprovsstorleken dock vara ungefär två till tre gånger så stor.

Ett sätt att öka precisionen i urvalsdesignen är att låta varje elev besvara fler uppgifter. Ett annat sätt är att utnyttja information om bakgrundsvariabler, som föräldrarnas utbildning och nationella bakgrund. Möjligheter finns att för de elever som valts ut att ingå i undersökningen skapa en databas genom att information från olika register läggs samman med information som samlas in med hjälp av frågeformulär och prov. Denna information kan användas för att förbättra precisionen i skattningarna, framförallt genom att den ger information om bortfallets omfattning och art.

Som urvalsmodell föreslås ett flerstegsurval. I första steget väljs skolor, eventuellt inom strata. Därefter görs urval av en, några eller alla klasser inom skolan. Urval av elever inom klasser kan också vara aktuellt, men i de fall en matrissamplingsdesign används är det lämpligt att låta alla elever i klassen ingå, och slumpmässigt distribuera de olika häftena till klassens elever.

Fördjupad analys och förklaring

Även om svårigheterna att nå fram till entydiga slutsatser om kausala relationer är stora, kan ett empiriskt underlag ge bättre förutsättningar för en informerad diskussion om förklaringar till de erhållna resultaten.

Den uppgiftsorienterade ansatsen kan ge möjlighet till kvalitativa analyser av arten av 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 determinanter på olika nivåer inom och utom utbildningssystemet.

Även för att förstärka analysmöjligheterna är databaser skapade genom sammanläggning av registerinformation med information insamlad inom den nationella kunskapsbedömningen av stort intresse. Detta kan exempelvis göra det möjligt att samanalysera resultat på nationella prov med resultaten från den nationella kunskapsbedömningen.

Enkäter till elever, lärare och skolledare 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, som exempelvis elevers egna bedömningar av kunskaper och färdigheter, attityder till olika ämnen och ämnesdelar, och motivation att lära.

Former för redovisning av resultat

Den information som ett nationellt kunskapsbedömningssystem skapar är av intresse för ett stort antal grupper som exempelvis allmänheten, elever, lärare, skolledare, beslutsfattare på lokal och nationell nivå, läroplansutvecklare, läroboksförfattare, och lärarutbildare. De olika informationsbehov som dessa grupper har pekar på att det är nödvändigt att på ett optimalt sätt utnyttja informationen både från den skalorienterade och den uppgiftsorienterade ansatsen.

De siffermässiga resultaten från den skalorienterade ansatsen kan vara abstrakta och svåra att tolka. Ansträngningar bör göras att förse presentationerna av resultatförändringar med tolkningsanvisningar som gör siffrorna tolkbara och meningsfulla även för grupper som inte har omfattande erfarenhet av användning av kvantitativa metoder.

Möjligheterna att använda webbaserad rapportering bör utredas. En stor fördel med denna är att användarna själva har möjlighet att med enkla metoder genomföra även komplexa analyser. Dessa analyser kan avse såväl resultat på enskilda uppgifter som resultat på olika skalor.

Ett aktivt analysarbete är en nödvändig förutsättning både för ett framgångsrikt sökande efter förklaringar till de observerade trenderna, och för den fortsatta utvecklingen av instrumenten. Det är därför nödvändigt att den nationella kunskapsbedömningens data används i olika forskningssammanhang, och att resultat presenteras i olika vetenskapliga fora.

Lokal utvärdering

Skolor/skolhuvudmän ska ges möjlighet till att på eget initiativ genomföra lokala utvärderingar vilkas resultat kan relateras till de nationella. Ett sätt att göra detta är att ge stöd för att på lokal nivå använda det uppgiftsmaterial som släpps fritt efter varje utvärdering. Dessa uppgifter kommer att vara omsorgsfullt konstruerade och noga utprövade, och vara försedda med omfattande anvisningar för bedömning av elevsvar. Ett förslag är att Skolverkets provbank vidareutvecklas genom att uppgifterna och delar av det empiriska materialet läggs in i en databas som är fritt tillgänglig för lokal användning.

Ledningsformer och förankring

Det nationella kunskapsbedömningssystemet måste vara stabilt över lång tid, och det måste ha en bred och djup förankring i alla intressentgrupper. En referensgrupp bör därför skapas med representation av bland andra kommuner, lärarfackliga organisationer, och elev- och föräldraorganisationer.

Skolverkets överväganden

4 Skolverkets överväganden

Vid de ovan nämnda forskarsymposierna 2005-09-15 och 2006-01-26 betonades att det utförda utredningsarbetet är mycket ambitiöst och seriöst och saknar motstycke såväl i Norden som i den svenska skoladministrationen. Åtskilliga synpunkter lämnades på Gustafssons utredning och förslag, bl.a. på hur stark kursplaneanknytningen bör vara; vilken koppling som kan göras till internationella jämförande studier; vilka syften ett system av detta slag bör och kan tillgodose; vilken periodicitet systemet bör ha; vilka ämnen och ämneskombinationer som bör ingå; hur systemet ska styras och samordnas; hur databehandling och rapportering ska ske; samt risken för oönskade effekter.

Jan-Eric Gustafssons förslag, liksom synpunkterna från symposierna samt Skolverkets övriga erfarenheter från verkets nationella utvärderingar liksom deltagande i internationella jämförande studier av olika slag, har legat till grund för de överväganden och förslag som redovisas i detta avsnitt. Inledningsvis redovisas dock den uppskattning av kostnaderna för Gustafssons förslag som Skolverket har genomfört.

4.1 Kostnadsuppskattningar

Skolverket har genomfört uppskattningar av vad ett system med upprepade mätningar omfattande samtliga grundskolans ämnen kan komma att kosta i fullt utbyggd skala, i den omfattning och struktur som beskrivs i Jan-Eric Gustafssons rapport. De kostnadsposter som beaktas är kostnader för ramverk- och instrumentutveckling, bjudning av instrument, datainsamling, databasuppbyggnad samt en första avrapportering. Till grund för uppskattningar ligger bl.a. Skolverkets erfarenheter från tidigare och pågående deltagande från internationella studier; från arbetet med nationella prov; och från NU 03. Beräkningsunderlaget redovisas i bilaga 1 till föreliggande rapport.

Kostnadsuppskattningarna är grundade på bedömningar av kostnader för ett ämnesområde, i detta fall naturvetenskap (NO). De olika kostnadsposterna kan variera mellan olika ämnen, men då osäkerheten i uppskattningarna är relativt stor avstår från att försöka göra mer detaljerade kostnadsuppskattningar. Då det i det i avsnitt 2.2 föreslagna rullande nationella utvärderingsystemet framgår att flera ämnen utvärderas vid samma mättillfälle så finns det med säkerhet en del samordningsvinster som bidrar till att reducera de totala kostnaderna för det föreslagna systemet. Ej heller för sådana eventuella samordningsvinster görs några försök att skatta storleksordningen då de bedöms som marginella i sammanhanget. Sådana eventuella samordningsvinster berör främst kostnader som berör bjudning av instrument, datainsamling samt

databasuppbyggnad. De största kostnadsposterna, ramverksframtagning samt instrumentutveckling, berörs i mindre utsträckning av eventuella samordningsvinster.

De gjorda beräkningarna grundar sig på ett ämne/en ämnesgrupp som är skalbaserat med kunskapsinstrument bestående av framför allt flervalsfrågor och öppna kortsvarsfrågor – detta i jämförelse med exempelvis ämnesgruppen SO, där mer uppgiftsorienterade mätningar med komplexa uppgifter troligen kan vara aktuella. Skolverket har i nuvarande läge inte underlag att bedöma vilka skillnader i kostnader som kan finnas mellan ett mer uppgiftsbaserat kunskapsinstrument jämfört med ett skalbaserat kunskapsinstrument. Därför görs ej heller här några försök att bedöma skillnader i kostnader mellan olika ämnen. Frågan om olika ämnens skilda förutsättningar i bl.a. mätänseende behandlas mer utförligt i avsnitt 4.2.

Kostnaderna för det av Jan-Eric Gustafsson föreslagna rullande utvärderings-systemet vilket har redovisats i avsnitt 2.2, kan förväntas vara större då mätningarna inleds i respektive ämne, jämfört med kostnaderna för mätningar vid efterföljande mättillfällen. De av Skolverket uppskattade totala kostnaderna för ett ämnes första cykel⁴⁹ enligt Gustafssons förslag uppgår till 6,4 mkr för det enskilda ämnet, där 4,2 mkr avser årskurs 8/9 och 2,2 mkr årskurs 4/5. Av dessa är 4 mkr kostnader för ramverksframtagande samt instrumentutveckling. Vid påföljande cykler skattas det totala kostnaderna per ämne till 3,6 mkr, där 2,2 mkr avser årskurs 8/9 och 1,4 mkr årskurs 4/5. De lägre kostnaderna för efterföljande cykler beror främst på att kostnader för framtagande av ramverk kan ses som en initial kostnad för systemet. Även framtagning av instrument medför högre initialkostnader i det föreslagna systemet. Det är rimligt att utgå från att alla uppgifter i ett nytt instrument för kunskapsutvärderingens första cykel behöver nyutvecklas. Återanvändningen av delar av föregående cyklers instrument kan förväntas medföra att kostnaderna för instrumentutveckling sjunker efter den inledande mätomgången, och här antas dessa kostnader sjunka med hälften.

Under antagande om att samtliga ämnen i grundskolan ska ingå i systemet, samt att mätningar görs i samtliga ämnen vid ett tillfälle i årskurs 4/5 (ca 11 ämnen aktuella) och ett tillfälle i årskurs 8/9 (ca 18 ämnen), medför detta att kostnaderna för hela systemet initialt skattas till drygt 100 mkr. Vid efterföljande mättillfällen minskar kostnaderna till knappt 55 mkr, dvs. drygt tio miljoner kronor per år i genomsnitt under antagande om att mätningarna i varje ämne upprepas vart femte år.⁵⁰

⁴⁹ Begreppet ”cykel” avser de aktiviteter som är förknippade med ett mättillfälle.

⁵⁰ Mätningar i ett ämne vart femte år har också föreslagits av Jan-Eric Gustafsson (se vidare avsnitt 2.2).

Slutsatser

I Skolverkets rapport till regeringen 2003-10-30 gjordes, som inledningsvis nämnts, bedömningen att det givna förslaget till ett förändrat provsystem var kostnadsneutralt. Föreliggande utredning visar dock att så knappast är fallet, om samtliga grundskolans ämnen ska utvärderas enligt en rullande modell, med modern mätteknik och krav på att resultaten ska vara riksrepresentativa, tillförlitliga och möjliga att jämföra över tid. I den fortsatta framställningen kommer därför olika aspekter på ett rullande utvärderingssystem att diskuteras, i syfte att belysa möjligheterna att skapa och vidmakthålla ett system med mer rimlig ambitionsnivå och därmed mer rimliga kostnader.

4.2 Måluppfyllelsebedömning eller trendmätning – olika ämnens olika förutsättningar

Måluppfyllelsebedömning, trendmätning och behovet av ramverk

Flertalet kunskapsbedömningsstudier utnyttjar något slag av dokument som beskriver vilka delområden inom ett kunskapsområde som ska studeras samt även hur dessa kunskaper ska studeras. Ett sådant dokument – eller samling av dokument – betecknas, som framgår av Jan-Eric Gustafssons utredning⁵¹ som ett *ramverk*. En möjlig kategorisering av olika typer av ramverk är att benämna ramverk som är direkt baserade på läro- och kursplaner som läroplansbaserade ramverk, vilka ska ses i motsats till ”oberoende” ramverk.

Skolverkets nationella utvärderingar NU 92, UG 95 och NU 03 har i första hand varit inriktade mot mätningar av kunskaper och färdigheter i enskilda skolämnen, och då fokuserat graden av *måluppfyllelse* i respektive ämne. I fall som dessa är det naturligt att utnyttja befintliga svenska styrdokument som läro- och kursplaner vid utvecklingen av prov för att bedöma elevers kunskapsnivåer. Vid instrumentutvecklingen har utnyttjats mer eller mindre explicita ramverk baserade på de styrdokument, dvs. läro- och kursplaner, som existerade vid tidpunkten för i första hand NU 92 – eftersom det uttalade syftet med såväl UG 95 som NU 03 var att göra s.k. ”repeats” på NU 92. Begeppet ”repeat” för dock närmast tanken till trendmätningar, och som tidigare framgått i denna rapport⁵² visar erfarenheterna från NU 03 att det inte är helt oproblematiskt att över tid fortsätta att använda s.k. läroplansbaserade ramverk.

De förslag som kommer att lämnas av den av regeringen tillsatta utredningen om översyn av grundskolans mål- och uppföljningssystem,⁵³ vilken enligt direktiven ska lämna sina bedömningar och förslag senast 2007-05-01,

⁵¹ Se vidare avsnitt 2.2.4.

⁵² Se vidare avsnitt 1.2.2.

⁵³ Utbildningsdepartementet (2006).

kan komma att innebära förändringar av kursplanemålen i åtskilliga grundskoleämnen och därmed bedömningen av måluppfyllelse – vilket egentligen ytterligare understryker de tidigare redovisade svårigheterna i att tillämpa den gamla ”NU-modellen”.

I flera nationella och internationella studier har utvärderingen av elevers kunskande inte utvärderats mot måluppfyllelse, utan mot ett ”oberoende” ramverk. I ett sådant ramverk ersätter det oberoende ramverket läro- och kursplanernas roll i beskrivningen av kunskapsområden som ska studeras. I bl.a. NAEP har avsaknaden av en gemensam läroplan i USA inneburit att ramverken som skapats har definierats så brett att de inrymmer en hel mångfald av innehållsval och sätt att lägga upp undervisningen. Men även i de fall då det finns en läroplan finns det starka skäl till att inte binda ramverket så hårt vid läroplanen att ramverket måste ändras så snart läroplanen ändras. I Nya Zeeland distanserades ramverket för NEMP från den nya zeeländska läroplanens mer detaljrika uttalanden, då sådana detaljer tenderar att ändras över tid, medan nationell *trendmätning* kräver ett långsiktigt perspektiv om uppdraget ska lyckas. I de internationella studierna som PISA, PIRLS och TIMSS är ramverken utvecklade med delvis olika utgångspunkter: Ramverksutvecklingen i exempelvis TIMSS har beskrivits som en process som ska spegla minsta gemensamma nämnare mellan de i studien ingående ländernas utbildningssystem. I PISA däremot har man inte utgått från läroplaner över huvud taget, utan ramverket har snarare baserats på en föreställning – framtagen och överenskommen inom i första hand OECD – om vilka kunskaper och färdigheter som är viktiga i samhället nu och i framtiden. Samtidigt indikerar ett par aktuella mindre studier som Skolverket har initierat, att överensstämmelsen mellan TIMSS’ och PISA:s ramverk och instrument å ena sidan och de svenska kursplanerna och de svenska nationella proven å den andra, är tämligen god.⁵⁴

Internationella jämförande kunskapsmätningar som TIMSS, PIRLS och PISA kan alltså ses som exempel på undersökningar baserade på ”oberoende” ramverk. Även studier som NAEP och NEMP kan ses som hörande till denna senare kategori, då det i första fallet för NAEP inte finns någon gemensam nationell läroplan i USA som kan utnyttjas för ramverksutveckling, och i det andra fallet för NEMP undveks medvetet alltför starka bindningar till den i Nya Zeeland existerande läroplanen för att möjliggöra trendjämförelser även efter ändringar i läroplanen.

Den gemensamma nämnaren för de studier som baserar kunskapsmätningar på ”oberoende” ramverk, är en uttalad målsättning att ha ramverk som är

⁵⁴ Skolverket (2006b); Lindström (2006). Den senare studien planeras, efter sedvanlig beredning i Skolverket, redovisas senare under hösten 2006.

stabila över tid. Valet att grunda mätningarna på ”oberoende” ramverk har baserats på en önskan av att ha ramverk som är stabila över tid, för att ge en god grund för att kunna göra utsagor om förändringar i elevgruppers kunnande över tid. Ett av de mest grundläggande kraven på mätningar med syfte att göra sådana utsagor är att ramverken är oförändrade mellan de olika tidpunkter då elevers kunnande ska jämföras.

För att kunna genomföra trendmätningar krävs att de kunskapsinstrument som används är konstruerade för ändamålet. Identiska test är en sådan möjlighet, en annan är att använda ett stort antal s.k. ankaruppgifter som är identiska vid olika mättillfällen. I båda fallen krävs att instrumenten utvecklas med ett och samma ramverk i botten. I första fallet med identiska test blir det bakomliggande ramverket naturligtvis identiskt. I andra fallet med ankaruppgifter finns möjligheter att förändra det underliggande ramverket.

Kraven på att ramverken måste vara oförändrade även efter förändringar i läro- och kursplaner och samtidigt kunna spegla graden av måluppfyllelse i ett skolsystem kan ge upphov till ett spänningsförhållande. Vid de forskarseminarier som Skolverket anordnade i anslutning till Jan-Eric Gustafssons underlagsrapport framhölls att risken finns att ramverken kan komma att betraktas som en de facto kursplan. I det amerikanska NAEP utnyttjas i varje utvärderingscykel ett ramverk som ska beskriva vad som ska fokuseras på i ett eller flera brett definierade innehållsområden, vilka svarar mot läroplansdefinierade ämnen. Detta ramverk har utvecklats i brett sammansatta konsensusgrupper. Därefter har ramverket utnyttjats för att ta fram provuppgifter som svarar mot de mål som definierats i ramverket. Erfarenheter från NAEP har också visat att ett sådant förfarande kan ge nationellt och regionalt representativa data om utbildningssystemet utan att detta har inneburit en implicit läroplan. De tidigare nämnda, av Skolverket initierade, aktuella jämförelserna av ”oberoende” ramverk och läroplansbaserade ramverk som för närvarande genomförs indikerar också som sagt att överensstämmelsen mellan de studerade ramverkstyperna är tämligen god.

Framtagandet av ett av styrdokumentens oberoende ramverk ligger också som en viktig komponent i Jan-Eric Gustafssons förslag till ett svenskt system. Som visats i avsnitt 3.1 är ett sådant ramverksarbete i det enskilda ämnet (eller gruppen av ämnen) en tämligen omfattande, och därmed tids- och kostnadskrävande, process.

Olika ämnen har olika förutsättningar

Regeringens uppdrag till Skolverket omfattar uttalat grundskolans samtliga ämnen. De kartläggningar och analyser som gjorts och de erfarenheter som finns från såväl NU 03 som andra länders nationella utvärderingssystem, visar

dock entydigt att olika skolämnen har olika förutsättningar att bedömas i en modell vars primära syfte är att kunna uttala sig om förändringar i resultat över tid.

Vissa ämnen – svenska, engelska, moderna språk, matematik och de naturvetenskapliga ämnena (NO) – kännetecknas av ett jämförelsevis stabilt kunskapsinnehåll och kan därmed betecknas som ”rostfria”. Inom dessa ämnen finns etablerade mättraditioner med – på gott och ont – konventionella prov som grund. De lämpar sig väl för moderna skalbaserade mätningar av den typ som Jan-Eric Gustafsson har beskrivit i avsnitt 3.2.7. Det är också dessa ämnen som oftast är föremål för såväl nationella prov som internationella jämförande kunskapsmätningar, och som oftast betecknas som utbildningspolitiskt mest intressanta. Kunskapsläget om resultaten är därför också förhållandevis gott.

Andra ämnen – främst de samhällsorienterande ämnena (SO) – är däremot inte på samma sätt ”rostfria” utan kunskapsinnehållet är av naturliga skäl mer känsligt för samhälleliga och globala förändringar. Samtidigt speglar SO-ämnena i viktiga avseenden det nationella kulturarvet, kunskap om det egna samhället och det egna statsskicket osv. NU 03 visar också att framför allt samhällskunskap och historia av föräldrar och elever betraktas som viktiga skolämnen. Etablerade mättraditioner i form av konventionella prov finns och är vanliga i den vardagliga skolverksamheten. Dessa mätningar lutar dock betydligt mer mot de uppgiftsbaserade modeller som också redovisats i avsnitt 3.2.7. Kunskapsinnehållets förändring över tid liksom dess variation mellan olika länder är dock påtagligt varför mätningar över tid är svårare än för den första typen av ämnen. De relativt få internationella studier som finns på området⁵⁵ begränsar sig till sådant kunskapsinnehåll som förutsätts vara gemensamt för olika länders läroplaner eller motsvarande, i första hand kunskap om och attityder till demokrati.

Den tredje gruppen ämnen – främst de praktisk-estetiskt inriktade ämnena – är på många sätt traditionstyngda då de flesta tycks ha en klar och delvis historiskt betingad bild av vad respektive ämnes kärna är – dvs. en viss form av praktisk kunskap som att teckna, spela ett instrument, laga mat osv. NU 03 visar att dessa ”göra”-moment i stor utsträckning lever kvar trots återkommande försök att modernisera ämnenas kursplaner och knyta dem till viktiga övergripande kompetenser som skapande förmåga, samarbetsförmåga osv. Förändringar i måldokumentet är relativt vanliga i syfte att få ämnena att hänga med sin tid, och de kan således betecknas som både ”rostfria” men samtidigt myck-

⁵⁵ IEA-studierna ”Civic Education in ten Countries: An Empirical Study” (1970-74) – i vilken dock Sverige endast deltog på gymnasial nivå; CIVED (1999/2000); och det planerade ICCES 2009.

et känsliga för samhällsutvecklingen. Inom dessa ämnen finns få vedertagna mättraditioner av konventionell provtyp. De kunskapsprov som ingick i t.ex. musik och bild i NU 03 riskerar att fokusera på detaljer och också kunskaper som inte längre betonas särskilt starkt i nationella måldokument.⁵⁶ Liknande erfarenheter finns i det nederländska PPO-systemet, vars mätinstrument i motsvarande skolämnen ur ett svenskt perspektiv framstår som direkt gammalmodiga. Såväl svenska erfarenheter som erfarenheter från andra länder visar att nationella utvärderingar av dessa ämnen är sparsamt förekommande och ofta har fokus på kartläggningar av praktiken, dvs. lektionsinnehållet snarare än regelrätta kunskapsmätningar. Denna typ av bedömningar (av ”den realiserade kursplanen”) kan dock användas för att bedöma såväl förändringar över tid som i förhållande till de nationella mål som betonas vid en given tidpunkt. Av motsvarande skäl finns praktiskt taget inga internationella jämförande studier alls för dessa ämnen.

Ovanstående kategorisering av grundskolans ämnen är naturligtvis konventionell men speglar också ämnenas karaktär och mättekniska traditioner (eller avsaknaden av sådana). Härutöver finns ett antal ämnen som av olika skäl är mer svårkategoriserade, som sällan eller aldrig har utvärderats nationellt och som vart och ett har sin karaktär – teknik, svenska som andraspråk, modersmål och teckenspråk.

Sammanfattningsvis kan konstateras att renodlade trend- och måluppfyllelsebedömningar av elevers kunskapsresultat i olika ämnen kan och behöver kompletteras eller – om så av olika skäl befinner sig nödvändigt – ersättas med andra typer av utvärderingar, exempelvis kartläggningar av vilket kunskapsinnehåll som faktiskt förmedlas på lektionerna;⁵⁷ kartläggningar av arbetsformer och inflytande; elevernas motivation och intresse; ämneslärares kunskapsyn osv. Samtidigt behöver framhållas att dessa typer av studier inte ger information om elevernas faktiska kunskapsresultat, vare sig över tid eller i förhållande till de nationella målen.

Slutsatser

Ett nationellt utvärderingssystem omfattande grundskolans samtliga ämnen behöver ta hänsyn till ämnenas olika förutsättningar enligt ovan. Detta påverkar såväl möjligheten att utarbeta överenskomna nationella ramverk som möjligheten att använda vissa trendmätningmetoder. De synpunkter som lämnades vid de ovan nämnda forskarsymposierna tog fasta på detta faktum liksom att det föreslagna systemet är mycket ambitiöst. Synpunkter framkom att ett

⁵⁶ Som exempel kan anges perspektivlära i bild; eller kunskap om vissa klassiska kompositörer i musik.

⁵⁷ Detta kallas på engelska ”opportunity to learn” (OTL) är ett inte alldeles ovanligt inslag i vissa internationella komparativa studier som t.ex. TIMSS.

nationellt kunskapsutvärderingssystem bör fokusera på de politiskt mest intressanta ämnena/kunskaperna (svenska, matematik och NO) och att det finns skäl att utvärdera dessa oftare och mer regelbundet än andra ämnen. Samtidigt vill Skolverket i likhet med regeringen understryka att varje ämne bör ses som lika viktigt och bidragande till att uppnå olika aspekter på läroplanens gemensamma mål.

I avsnitt 4.1 har redovisats de uppskattade kostnaderna för ett svenskt system enligt det förslag som Jan-Eric Gustafsson på Skolverkets uppdrag har tagit fram. Dessa kostnader baseras bl.a. på instrumentutvecklingskostnaderna i de ämnen i vilka det finns nationella prov. Förslaget förutsätter vidare att nationella ramverk tas fram för samtliga ämnen i grundskolan, utgående från tanken på skalbaserade mätningar med full eller åtminstone god domäntäckning (över hela ämnet). Som ovan visats varierar dock förutsättningarna för att tillämpa detta tillvägagångssätt mellan grundskolans olika ämnen. Gustafsson har själv också angivit att en avvägning mellan skalbaserade och uppgiftsbaserade mätmetoder behövs i ett nationellt system för kunskapsutvärdering, och att detta bl.a. har att göra med de olika ämnenas karaktär och bedömnings-traditioner. I vilken grad en utvärdering av ett enskilt ämne vilken grundas på annan metodik – t.ex. uppgiftsbaserade mätningar – skulle leda till en avsevärt annorlunda kostnadsbild är dock som tidigare nämnts synnerligen svårt att bedöma. Likaså är det svårt att avgöra om kostnaden för att ta fram ett nationellt ramverk för ett enskilt ämne blir annorlunda än för något annat ämne, beroende på det enskilda ämnets karaktär. Skolverket har knappast underlag för att hävda att så skulle vara fallet. Verkets erfarenheter tyder snarare på att de kostnadsskillnader som finns mellan nationella prov i olika ämnen har att göra med kostnadsskillnader mellan olika universitetsinstitutioner, snarare än med de olika ämnenas karaktär, innehåll, ”betydelse” el. dyl. Erfarenheterna från NU 03 – där de flesta grundskoleämnena utvärderades med hjälp av en rad olika universitetsinstitutioner runt om i landet – styrker denna uppfattning.

Ett betydligt säkrare sätt att få ner kostnaderna för ett nationellt kunskapsutvärderingssystem är snarare, enligt Skolverkets uppfattning, att ett sådant system avgränsas till ett relativt begränsat antal ämnen, och att därvid också i högre grad än hittills se de internationella kunskapsmätningarna som led i en nationell strategi för kunskapsutvärderingar. Den sistnämnda aspekten utvecklas vidare i följande avsnitt.

4.3 Koppling till internationella studier – möjligheter och begränsningar

Ett förslag som förs fram i Jan-Eric Gustafssons underlagsrapport är att använda internationella kunskapsstudier som delkomponenter i ett nationellt rullande utvärderingssystem där så befins vara lämpligt. Ska de internationella studierna utnyttjas i en nationell rullande utvärdering finns flera frågeställningar som behöver beaktas. Några av de mer väsentliga frågorna är: Mäter de internationella studierna rätt saker? Är de periodiskt återkommande? Är de även utformade som trendstudier? Mäter de internationella studierna tillräckligt ofta för de nationella trendbehoven? Är de elevpopulationer med avseende på åldrar och/eller årskurser som undersöks i de internationella studierna överensstämmande med de elevgrupper som bör ingå i en nationell utvärdering? Är de studiedesigner som utnyttjas i de internationella studierna också lämpade för en nationell utvärdering? Är de internationella studierna och deras ramverk⁵⁸ och instrument någorlunda överens med läroplanens och kursplanernas kunskapsyn? Går de internationella studierna att påverka med nationella tillägg?

En grundläggande förutsättning är att det finns internationella jämförande kunskapsstudier inom de ämnesområden som ska ingå i ett nationellt rullande system för utvärdering av den svenska grundskolan. Vidare måste de internationella studier som är lämpliga/möjliga att utnyttja i en nationell utvärdering vara planerade att genomföras med regelbundet återkommande mätningar, om de ska kunna ingå i ett nationellt rullande utvärderingssystem. För att resultat från internationella studierna ska gå att utnyttja för att beskriva förändringar i elevers kunskande över tid i en rullande nationell utvärdering krävs också att berörda studierna är planerade att vara trendstudier, dvs. att de är tänkta att genomföras regelbundet återkommande över tiden och att kunskapsinstrument som utnyttjas gör det möjligt att uttala sig om trendmässiga förändringar i elevernas kunskapsnivåer.⁵⁹ Det krävs också att de internationella studierna har en för det nationella rullande utvärderingssystemet lämplig periodicitet, då cykellängd i de internationella studierna är svår att påverka. Det medför att de rullande nationella utvärderingarnas cykellängd i vissa ämnen kan bli styrd av de internationellas cykellängd. Tiden mellan mättillfällena får ej heller vara för lång – annars försvåras möjligheterna att analysera möjliga orsakerna till en observerad trendmässig förändring av elevers kunskaper och färdigheter.

⁵⁸ För internationella studier finns givetvis ingen läroplan att utnyttja vid instrumentutveckling och tolkning av data, utan ett ramverk utvecklats vilket har som syfte att definiera vilka kunskaps- och färdighetsområden som skall studeras samt beskriva instrumentet som skall utnyttjas i dessa studier.

⁵⁹ För att kunna göra jämförelser över tid krävs bland annat identiska test eller ett stort antal s.k. ankaruppgifter som är desamma vid olika mättillfällen.

De internationella studier som är möjliga att låta ingå i en nationell rullande utvärderingsmodell ska regelbundet ha återkommande mätningar, samt helst också vara utformade som trendstudier, vilket avgränsar den fortsatta diskussionen till i huvudsak PISA, TIMSS och PIRLS.⁶⁰ Dessa studier har alla en målsättning att kunna uttala sig om förändringar i elevers kunskaper och färdigheter över tid, dvs. trend. PISA mäter läsförståelse, matematik och naturvetenskap hos 15-åriga elever med tre års mellanrum.⁶¹ TIMSS mäter matematik och naturvetenskap hos elever i årskurs 4 samt 8 med fyra års mellanrum.⁶² PIRLS mäter elevers läsförståelse i årskurs 4 vart femte år.

Periodiciteten för TIMSS och PIRLS ligger vid eller nära den för det nationella utvärderingssystemet föreslagna periodiciteten på fem år. PISA, som upprepas vart tredje år, har en periodicitet som avviker från det planerade nationella utvärderingssystemet.⁶³

Den svenska grundskolan har inte någon formell stadiindelning och då det från statens sida endast formulerats mål för årskurs 5 och 9 i grundskolan, är det väsentligt att de internationella studier som eventuellt ska ingå som delar i ett nationellt kunskapsutvärderingssystem kan kopplas till de nationella målen. PISA:s 15-åriga elever överlappar i stor utsträckning elevpopulationen i årskurs 9.⁶⁴ För TIMSS del är eleverna för såväl den yngre som den äldre studiepopulationen något yngre än de elever för vilka det finns nationella mål angivna i kursplanerna. Val av vilka årskurser som kan vara lämpliga i ett nationellt kunskapsutvärderingssystem diskuteras närmare i avsnitt 4.4, samt även i Jan-Eric Gustafssons bilagda underlagsrapport.

En viktig aspekt att ta hänsyn till är de urvalsmodeller som används i de internationella studier som kan komma i fråga som delkomponenter i ett nationellt utvärderingssystem. Urvalsdesignen definierar också hur studiepopulation ska avgränsas. Urvalsdesignen anger även hur urvalet av elever från studiepopulationen ska genomföras. Den beskriver även de krav på minsta urvalsstorlekar som ställs i studien i fråga. Även hur olika undergrupper repre-

⁶⁰ En i sammanhanget intressant framtida studie på EU-nivå är den språkindikator som diskuteras. Det är oklart när denna språkindikator skall realiseras, och även med vilken regelbundenhet som mätningar inom ramen för en EU:s språkindikator kan komma att genomföras. Enligt nuvarande planering skall de mätningar av elevers språkfärdigheter som skall ingå i språkindikatorn genomföras i slutet av den obligatoriska skolan, vilket i praktiken förväntas innebära att studiepopulationen blir elever i 15 års ålder.

⁶¹ Detta avser rapportering av trend på generella övergripande skalor, och rapportering på detaljerad nivå uppdelat efter olika kunskapsdimensioner ges endast vart nionde år.

⁶² Mätningar har även gjorts av gymnasieelever, dock ej med ett regelbundet återkommande trendperspektiv.

⁶³ Det skall upprepas att rapportering av trend i PISA sker på generella övergripande skalor, och rapportering på detaljerad nivå uppdelat efter olika kunskapsdimensioner ges endast vart nionde år, även om mindre detaljerade trendresultat rapporteras vart tredje år.

⁶⁴ I genomsnitt går 95 % av alla 15-åringar i årskurs 9, 4 % i årskurs 8 och 1 % i första årskursen i gymnasieskolan.

senteras i urvalsdesignen och därmed i urvalet har betydelse för vilka analyser som kan genomföras. Genom att stratifiera urvalet är det möjligt att säkerställa att viktiga undergrupper förekommer i tillräcklig stor utsträckning i urvalet för att det ska gå att presentera resultat för dessa undergrupper med tillräcklig god precision. Urvalet i flertalet av de internationella studierna (bland annat PISA, PIRLS och TIMSS) utnyttjar en urvalsdesign med flerstegsurval. I första steget väljs skolor, eventuellt inom strata. Därefter görs urval av en, några eller alla klasser inom de utvalda skolorna (PIRLS och TIMSS) alternativt så väljs helt slumpmässigt ett antal elever på varje utvald skola (PISA). Grundläggande principer i urvalsdesignen är som regel inte möjliga att påverka. Till dessa principer hör som regel definitionen av studiepopulation, samt hur urvalet ska genomföras på generell nivå (flerstegsurval, elevurval, klassurval, minsta urvalsstorlek, etc.). Däremot finns det som regel möjligheter att påverka hur exempelvis stratifieringsprinciper ska tillämpas i respektive deltagarland. Det medför att det finns goda möjligheter att på nationell nivå påverka vilka undergruppsresultat som är möjliga att presentera – exempelvis elever med utländsk bakgrund; elever som går i fristående skolor; m.m.

Ska de kunskapsprov som används i de internationella studierna också passa för den nationella utvärderingens syften krävs, som tidigare nämnts, att graden av överensstämmelse mellan det nationella ramverket⁶⁵ i ett ämne och en internationell studies ramverk är tillräckligt stor för den nationella utvärderingens behov. Även uppgifternas fördelning över olika kognitiva och innehållsliga delområden (s.k. domäner) är en aspekt som måste beaktas. Uppgifter och uppgiftstyper som saknas eller är under- eller överrepresenterade i de internationella studierna relativt de svenska behoven i en nationell rullande utvärderingsmodell kan ge upphov till problem. För de områden som är underrepresenterade eller saknas i en internationell studie är en tänkbar lösning att skapa kompletterande uppgifter. De områden som är överrepresenterade är svårare att hantera, därför att det av praktiska skäl inte är möjligt att ta bort uppgifter. Om det inte skulle vara möjligt att tona ned de delar av den internationella studien som är överrepresenterade jämfört med det svenska ramverket är en möjlighet att försöka ta hänsyn till detta vid tolkningen av resultaten. Ytterligare en aspekt som är relevant att beakta är hur de internationella studierna och deras ramverk hanterar bredd i förhållande till djup i kunskaper. Internationella kunskapsstudier är i allmänhet skalorienterade, vilket medför att de instrument som utnyttjas för kunskapsmätning till stora delar innehåller flervalfrågor och

⁶⁵ Detta nationella ramverk kan, som närmare framgår av avsnitt 3.2, vara starkt kopplat till de svenska läro- och kursplanerna vilket var utgångspunkten för exempelvis NU 92, där utvärderingen framför allt gjordes mot relevanta mål i Lgr 80. Det nationella ramverket kan också vara konstruerat för att möjliggöra stabila trendmätningar, så som beskrivs i Jan-Eric Gustafssons bilagda underlagsrapport.

korta elevsvar. Mer komplexa uppgifter saknas många gånger i de internationella kunskapsinstrumenten. Sådana mer komplexa uppgiftstyper kan också vara svårare att utformas som nationella tillägg till de internationella kunskapsinstrumenten. Avvägningen mellan utnyttjandet av skalorienterade uppgifter och en uppgiftsorienterad ansats diskuteras på annan plats i denna rapport.

De två nämnda aktuella mindre studier som Skolverket har initierat indikerar som tidigare nämnts att överensstämmelsen mellan TIMSS' och PISA:s ramverk och instrument å ena sidan och de svenska kursplanerna och de svenska nationella proven å den andra, är tämligen god inom de avseenden som studerats – läsförståelse, matematik och naturvetenskap.⁶⁶ Samtidigt bör noteras att utgångspunkterna för TIMSS' och PIRLS' ramverk å ena sidan, och PISA:s å den andra, är helt olika. I PISA är utgångspunkten *inte* de medverkande ländernas nationella läroplaner, vilket är fallet för TIMSS och PIRLS. PISA:s ramverk baseras istället på kompetensbeskrivningar som expertgrupper inom OECD utformat utifrån förväntade allmänna krav på de elever som lämnar utbildningssystemen – kompetensbeskrivningar om vilka bred internationell konsensus kan anses råda. Kritik mot PISA:s ansats togs bl.a. upp vid de forskarsymposier som Skolverket anordnade i anslutning till Jan-Eric Gustafssons rapport, med innebörd att mätning av generella kompetenser kan riskera att mer fungera som ett intelligenstest.

I såväl PISA som PIRLS utvärderas enbart elevers läsförståelse, och för att en nationell utvärdering av svenskämnet även ska täcka de delar som ej täcks av PISA eller PIRLS krävs något slag av kompletterande studier. Det gäller exempelvis områden som skriftlig produktion, muntlig framställning och litteratur – vilka i varierande utsträckning har inkluderats i Skolverkets tidigare s.k. nationella utvärderingar av grundskolan. Man kan också överväga att på något sätt inkludera svenska som andraspråk, mot bakgrund av bl.a. den ökande andelen elever som studerar detta ämne⁶⁷ liksom dess betydelse ur integrationspolitisk synvinkel.⁶⁸ Hur sådana kompletteringar lämpligen ska utformas behöver dock utredas ytterligare. Svenskämnet bör mot denna bakgrund ses som ett högt prioriterat ämne i ett svenskt kunskapsutvärderingssystem.

För de studier som diskuteras i detta avsnitt finns vissa möjligheter till att komplettera de internationella kunskapsinstrumenten med nationella tillägg i syfte att förbättra deras överensstämmelse med en nationell kunskapsutvärderings behov. Sådana kompletteringar av internationella instrument får som

⁶⁶ Skolverket (2006b); Lindström (2006; Opublicerat manus).

⁶⁷ Andelen grundskoleelever som läste svenska som andraspråk var läsåret 2001/02 5,5 %; läsåret 2002/03 6,1 %; läsåret 2003/04 6,5 %; läsåret 2004/05 6,6 %; och läsåret 2005/06 6,8 %.

⁶⁸ Svenska som andraspråk ingick f.ö. som en tämligen blygsam del i utvärderingen av svenskämnet i NU 03.

regel ske under förutsättning av bjudningsförfarandet inte förändras.⁶⁹ Ska tillägg nyttjas förutsätter det bland annat att kompetens finns att tillgå nationellt i fråga om urvalsmetodik, instrumentutveckling och statistisk analys motsvarande den internationella. Nationella tillägg medför som regel även krav på större antal elever och skolor i urvalet, vilket får kostnadsmässiga konsekvenser. En annan möjlighet är att komplettera det internationella kunskapsinstrumentet med helt fristående kunskapstest. Detta alternativ kräver inte som regel att elev- och skolorval utökas, även om kraven på tillgång till nationell expertis för instrumentutveckling och analys kvarstår. Konsekvenser av bland annat att total testtid för eleverna förlängs måste dock beaktas, likväl som den större komplexiteten i bjudningsförfarandet. Långa testtider för elever kan medföra negativa konsekvenser i form av svarsbortfall och/eller bristande motivation hos både elever och lärare.

Vid utvärderingar, såväl internationella som nationella, finns som regel ett behov att bjuda enkäter till elever, lärare och skolledare, samt i vissa fall även till föräldrar. För alla aktuella internationella studier finns goda möjligheter att komplettera de internationella enkätinstrumenten med nationella tillägg, vilket kan bli aktuellt för att täcka aspekter som behöver belysas i en nationell utvärdering.

En positiv konsekvens av att basera periodiskt återkommande utvärderingar på internationella jämförande studier med ambitioner att redovisa trend är att en nationell trendlinje kan erhållas för vissa ämnen redan från start.

De internationella kunskapsstudiernas rapporteringstid är som regel runt 18 till 20 månader. Om den nationella kunskapsbedömningen ska utnyttja resultaten från de internationella kunskapsstudierna blir även den nationella rapporteringen beroende av tidpunkten för den internationella rapporteringen.

Slutsatser

I avsnitt 1.2.3 har nämnts deltagande i internationella jämförande kunskapsstudier som ett alternativ till en nationell utvärdering av motsvarande kunskapsområden i den svenska grundskolan. I de fall där en internationell studie står i överensstämmelse med intentioner och ambitioner i de nationella måldokumenterna, kan de internationella studierna med fördel ersätta de rent nationella mätningarna. De blir med andra ord en integrerad och mer eller mindre löpande del i det nationella systemet, snarare än – som hittills – en tämligen fristående verksamhet. Därmed ökar kravet på att även de internationella studierna genomförs med viss regelbundenhet och att de inte prövas förutsättningslöst vid varje nytt tillfälle.

⁶⁹ Exempelvis skulle möjligheten kunna prövas att i en första omgång inkludera samhällskunskapsämnet i form av en sådan nationell modul till ICCS.

Enligt de bedömningar som Skolverket har gjort angående deltagande i internationella kunskapsutvärderingar är ett deltagande i PISA oundvikligt. De kunskapsbehov som PISA fyller är något som krävs av politiker och andra beslutfattare, bl.a. mot bakgrund av behovet av referenspunkter i en alltmer globaliserad ekonomi. Det är dock mindre lämpligt att låta PISA utgöra den enda kunskapskällan för att täcka ett nationellt kunskapsutvärderingsbehov. Dels kan inte PISA för närvarande ge information om de lägre åldrarnas kunskapsnivåer och förändringar i kunskapsnivåer över åren dels ger PISA detaljerad information för ett ämne enbart vart nionde år, vilket är en alltför lång tid mellan mättidpunkterna. PISA:s ramverk i matematik och naturvetenskap täcker inte de svenska kursplanerna lika bra som TIMSS, även om det bättre täcker en del av de formuleringar som finns i läroplanen. Därför bör PISA och TIMSS årskurs 8 ses som komplement till varandra när det gäller ämnesområdena matematik och naturvetenskap. TIMSS för årskurs 4 täcker dessutom ett behov av information om kunskapsläget hos de yngre eleverna som inte täcks av någon annan internationell – eller nationell – studie. För svenskämnet del kan läsförståelse utvärderas med hjälp av PIRLS för de yngre eleverna (årskurs 4) och med PISA för de äldre eleverna (årskurs 9/15-åringar) – även om man härvid bör beakta att de olika mätinstrumentens konstruktion i PIRLS respektive PISA innebär svårigheter att uttala sig om progression i kunskaper mellan årskurserna 4 och 9. I båda dessa fall utvärderas enbart elevers läsförståelse, vilket enbart täcker delar av svenskämnet, och för att nå en acceptabel täckning vid en nationell utvärdering av svenskämnet finns det behov av olika kompletteringar. Hur sådana kompletteringar lämpligen ska utformas bör som nämnts utredas ytterligare, och svenskämnet bör mot denna bakgrund ses som ett högt prioriterat ämne i ett svenskt kunskapsutvärderingssystem.

Vad gäller EU:s språkindikator, vilken är planerad att samlas in från elever som är i slutet av sin obligatoriska skolgång, dvs. femtonåriga elever, finns för närvarande mycket begränsad information. Kommissionen har föreslagit att i den första omgången ska man av praktiska skäl testa språkkunskaperna i de fem språk som är de viktigaste undervisningsspråken i unionen som helhet (dvs. engelska, franska, tyska, spanska och italienska). Dessa språkfärdighetstest kan förväntas knytas till det för Europa gemensamma ramverket ”Common European Framework of Reference for Languages”(CEFR).⁷⁰ Då även den svenska undervisningen i engelska och moderna språk allt mer knyts till CEFR bedöms möjligheterna som principiellt goda att utnyttja dessa språkfärdighetstest i en nationell rullande utvärderingsmodell. Det finns dock osäkerhet vad avser dessa möjligheter. I avvaktan på hur arbetet med den s.k. språkindikatorn

⁷⁰ CEFR, ramverk för språk utvecklat av Europarådet.

i EU utvecklas, bör mot denna bakgrund planeringen av ett svenskt utvärderingssystem tills vidare inkludera engelska, ett ämne där resultatinformation annars kommer att saknas.

En samordning med de internationella studier kan förväntas ge upphov till flera besparingar, jämfört med en situation med fortsatt deltagande i de internationella studierna om samtidigt nationella instrument utvecklas inom samma ämnesområden. En ekonomisk besparing uppstår direkt från att utgifter för att framtagande av trendprov, bjudningen av dem samt analys- och rapportarbetet endast behöver utföras i den internationella studien. Utnyttjandet av den i Sverige begränsade tillgången till forskare och annan expertis kan ske på ett mer effektivt sätt, och risken för utträngning av andra önskvärda kunskapsstudier på grund av bristande tillgång till kvalificerad personal kan minimeras. Denna kompetens frigörs för andra forskningsinsatser, då resurser av detta slag ej binds upp i lika stor omfattning. Ytterligare en fördel är att kompetens på mät- och utvärderingsområdet kan byggas upp och vidmakthållas i det svenska forskarsamhället. Slutligen undviks att elever samt skolpersonal belastas i för stor utsträckning av en medverkan i alltför många provsammanhang.⁷¹

4.4 Periodicitet – hur ofta bör mätningar genomföras och i vilka årskurser?

Jan-Eric Gustafssons förslag innebär en femårig periodicitet, dvs. ett och samma ämne utvärderas vart femte år. Forskarsymposierna bekräftade i stort att denna bedömning är väl avvägd, både i förhållande den politiska nivåns behov av trendinformation; den resursmässiga belastningen på Skolverket och berörd nationell expertis; samt rimliga tidsperioder för avrapportering och analys av resultaten. I ett fullt utbyggt system, menade man, kan dock vissa ämnen som är mer svåra att mäta – exempelvis de s.k. praktiskt-estetiska ämnena – utvärderas med glesare intervall. Erfarenheterna från de amerikanska och nederländska systemen talar för att detta senare är en rimlig tanke.

PIRLS har f.n. en femårig periodicitet, TIMSS en fyraårig och PISA en treårig periodicitet. I realiteten innebär dock PISA:s skiftande fokus på olika ämnen vart tredje år att en fullständig trend för den enskilda kompetensen – t.ex. läsfärdighet – endast ges vart nionde år. Vid forskarsymposierna ansågs denna tidsutdräkt alltför lång, samtidigt som flera länder inom OECD av främst ekonomiska skäl har fört fram tanken på en något glesare periodicitet i PISA än tre år – hittills dock utan resultat. Även mot denna internationella

⁷¹ Kostnadsaspekter förknippade med ett utnyttjande av internationella studier för nationella utvärderingsbehov har även berörs i avsnitt 4.1.

bakgrund framstår alltså en femårig periodicitet i ett svenskt system som rimlig.

Gustafssons förslag innebär vidare att kunskapsmätningar bör göras vid två tidpunkter under elevernas skolgång – en vid slutet av den obligatoriska skolgången och en i tidigare åldrar. För svenskt vidkommande skulle detta lämpligen innebära årskurserna 4 eller 5, respektive 8 eller 9. Såväl i de internationella studierna (TIMSS, PIRLS) som i den nationella politiska debatten kan Skolverket konstatera att det finns brist på resultatinformation i elevers tidigare åldrar, och följdaktligen ett ökat intresse för sådan information – Skolverket har erfarit att frågan f.n. också diskuteras i OECD/PISA. Ett svenskt system skulle därför rimligen behöva inkludera även årskurs 4 eller 5. Mätningarna i dessa åldrar skulle dock inte behöva vara lika omfattande och kostnadskrävande, och skulle med fördel kunna vara mer tematiskt uppbyggda än inriktade på det specifika ämnet. Ramverkskonstruktionen skulle också vara gemensam för de två inkluderade årskurserna, vilket innebär rationaliseringsvinster och kostnadsbesparingar.

4.5 Ämneskombinationer och risker för önskade effekter

Ytterligare en fråga att beakta vid beslut om ett nationellt system för kunskapsutvärdering är vilka ämneskombinationer som ska ingå. Jan-Eric Gustafssons förslag, vilket återgivits i avsnitt 2.2, innebär en relativt konventionell gruppering av ämnen (praktisk-estetiska; främmande språk inkl. engelska; modersmål/svenska; matematik/NO; samt SO-ämnen) vilka utvärderas enligt ett rullande system. Gustafsson anger dock samtidigt själv att denna gruppering är tentativ och givetvis kan diskuteras.

Det finns dock samtidigt argument för att gruppera ämnen på ett betydligt mer okonventionellt sätt. Enligt detta synsätt skulle en otraditionell gruppering av ämnen bidra till att bryta upp konserverande uppfattningar, statuskillnader och önskade ämneshierarkier. Att sådana skillnader och hierarkier finns och genomsyrar såväl attityder som skolarbetets organisering bär NU 03:s resultat tydliga vittnesbörd om. Exempelvis har under utredningsarbetet framförts förslag om att NO-ämnen och SO-ämnen borde utvärderas tillsammans under den tentativa rubriken "Omvärldskunskap";⁷² likaså att bild och svenska bör ses som olika aspekter på kommunikativ kompetens; liksom att kursplanerna för matematik och slöjd innehåller gemensamma inslag av storleksuppfattning och förståelse för geometri, vilka kan utvärderas samlat genom gemensamma mätinstrument. Vid forskarsymposierna framfördes åtskilliga uppfattningar av denna typ med innebörden att ett system av föreslagen typ

⁷² Detta gjordes f.ö. i såväl NU 89 som US 98.

också bör vara utvecklingsdrivande. En traditionell gruppering av ämnen med traditionella, i huvudsak psykometriska, mätinstrument kan ha en både styrande och konserverande effekt på skolsystemet och undervisningen. I synnerhet i de tidigare skolåren är det som ovan nämnts fullt rimligt att tänka sig mätinstrument som är mer ämnesövergripande och tematiskt uppbyggda. Detta var f.ö. en av de bärande idéerna i Skolverkets nationella utvärdering av skolan 1998 (US 98), *Läroplanerna i praktiken*.⁷³

Skolverket instämmer i att det finns risk för att den traditionella synen på ämnen, ämnesinnehåll och ämnesgrupper kan ha en konserverande effekt – verket har bl.a. understrukit detta i slutsatserna i NU 03:s huvudrapport.⁷⁴ Samtidigt talar Skolverkets erfarenheter för att det finns klara rationaliseringsvinster i utvecklingen av gemensamma ramverk för ämnen med sinsemellan liknande ämnesinnehåll (t.ex. moderna språk). Verkets erfarenheter från de internationella studierna, exempelvis PISA, talar vidare för att de styrande effekterna på undervisningen i svenska skolor är begränsade i stickprovsundersökningar med inte alltför täta intervall. De styrande – eller, om man så vill, förebildliga – ekterna av det nationella provsystemet och i synnerhet de obligatoriska proven är i så fall betydligt större.

Det finns inte heller något hinder för att mer övergripande kompetensområden i tillämplig utsträckning kan beaktas i anslutning till ett nationellt utvärderingssystem med mer konventionell gruppering av ämnen, eller för den delen i anslutning till vissa av de internationella studierna. Exempelvis ingick problemlösningsförmåga som en valbar del av PISA 2003. Ett annat exempel är det aktuella norska s.k. PISA-PLUS-projektet där klassrumsobservationer kompletterar de mer traditionella skalbaserade mätningarna i PISA.⁷⁵

Särskild värd att ta fasta på i Gustafssons utredning vad gäller ämneskombinationer är Härnqvists redan 1975 framförda tanke att vid kunskapsmätningar i olika ämnen använda läsfärdighet/läsresultat som kontrollvariabel.⁷⁶ Inom PISA får eleverna vid varje undersökningstillfälle besvara uppgifter inom läsning, matematik och naturvetenskap, vilket gjort det möjligt att bland annat belysa den stora betydelse som läskompetens har för prestationer inom matematik och naturvetenskap. Liknande slutsatser har dragits i NU 03 vad gäller läsfärdighetens betydelse för prestationer i de samhällsorienterande ämnena.

⁷³ Skolverket (1999).

⁷⁴ Skolverket (2004d), s. 132.

⁷⁵ "The Learner's Perspective Research Study in Norway - A classroom study to enlighten problematic PISA findings in mathematics" under ledning av professor Kirsti Klette vid Pedagogisk Forskningsinstitut, Universitetet i Oslo.

⁷⁶ Härnqvist, (1975).

4.6 Hur ska ett nationellt system komma till lokal användning?

I de direktiv som gavs till Jan-Eric Gustafsson rörande ett rullande nationellt kunskapsbedömningssystem sades bland annat att resultaten från utvärderingen i första hand skulle presenteras på nationell systemnivå, men också att "det är dock önskvärt att skolor/skolhuvudmän ges möjlighet till att på eget initiativ genomföra lokala utvärderingar vilkas resultat kan relateras till de nationella".

Det sätt att ge skolor/skolhuvudmän möjlighet att utvärdera elevers kunskaper och färdigheter som föreslås i Gustafssons utredning är att ge stöd till utvärdering på lokal nivå genom att använda det uppgiftsmaterial som släpps fritt efter varje utvärdering. I Gustafssons utredning konstateras att även om det är enkelt och kostnadseffektivt att vid varje tillfälle använda samma uppgifter är detta inte hållbart i längden, och i synnerhet inte inom områden med en snabb kunskapsutveckling. Där görs också bedömningen att "även om det är möjligt att hemlighålla en del av uppgifterna från ett tillfälle till ett annat kräver ett svenskt nationellt kunskapsbedömningssystem en sådan öppenhet att det inte är rimligt att bygga trendstudier på att instrumenten i sin helhet ska hållas hemliga.". Det ska nämnas att i NEMP och NAEP (de tidigare nämnda kunskapsbedömningssystem i Nya Zeeland respektive USA) byts mellan 30 procent och 50 procent av uppgifterna ut mellan varje utvärderingsomgång.

Sådana uppgifter har fördelen av att de är omsorgsfullt konstruerade och nog utprovade, och kanske också försedda med omfattande anvisningar för bedömning av elevsvar. Därför lämpar sig sådana frisläppta uppgifter väl för att utnyttjas i lokala kunskapsutvärderingar. Det bör även vara möjligt att ange nationella referensnivåer på uppgiftsnivå, vilket medför att lokala resultat kan relateras till resultat på nationell nivå. Uppgifterna och åtminstone vissa delar av det empiriska materialet kan läggas in i en databas där det är fritt tillgängligt för lokal användning. Databasen kan omfatta såväl de uppgifter som ingår i den skalorienterade ansatsen, som de som ingår den uppgiftsorienterade ansatsen. De uppgifter som utvecklats för användning i den skalorienterade ansatsen kan även användas på ett uppgiftsorienterat sätt, och i många fall omfattar även dessa uppgifter elevproduktion som kan vara värdefull i formativa och diagnostiska sammanhang. Uppgifterna ska dock i huvudsak användas på ett uppgiftsorienterat sätt oberoende av till vilken ansats de utvecklats. Skolverkets provbank för grundskolan, vilken f.n. är under utveckling, bör lämpligen kunna utnyttjas för att göra uppgifterna och delar av det empiriska materialet

tillgänglig för lokal användning. Även andra möjligheter att tillgängliggöra uppgifter för lokal användning bör dock övervägas.⁷⁷

Om de internationella systemen för kunskapsutvärdering (PISA, TIMSS, PIRLS osv.) ska ingå som delkomponenter i nationellt utvärderingssystem finns vissa begränsningar i möjligheterna att tillhandahålla uppgifter för lokal utvärdering för de berörda ämnesområdena. De uppgifter som släpps fria efter varje omgång i respektive internationell studie publiceras och är fritt tillgängliga. Ett utnyttjande av frisläppta uppgifter från de internationella studierna i lokala utvärderingar kompliceras dock samtidigt av upphovsrättsliga frågeställningar.⁷⁸ Även om uppgifter är frisläppta för spridning och diskussion är de därmed av upphovsrättsliga skäl inte automatiskt fria att användas i ett helt nytt sammanhang, som i lokala utvärderingar.

Samtidigt anfördes vid de tidigare nämnda forskarseminarierna i anslutning till Jan-Eric Gustafssons underlagsrapport att det finns en potential för Skolverket att tydligare än vad som kanske har gjorts ”marknadsföra” de uppgifter som släpps fria från exempelvis TIMSS – något som t.ex. skedde i form av särskilda rapporter i anslutning till redovisningen av resultaten från TIMSS 1995.

Slutsatser

Med det förslag till kunskapsbedömningssystem som här föreslås, finns goda möjligheter att tillhandahålla frisläppta uppgifter från varje utvärderingsomgång, i syfte att utnyttjas i en lokal utvärdering. Ett lämpligt forum för detta är Skolverkets provbank för grundskolan, vilken f.n. är under utveckling. Den lokala utvärderingen begränsas dock till att utnyttja frisläppta uppgifter på ett uppgiftsorienterat sätt, oberoende om uppgifterna utvecklats för en skalbase-rad ansats eller en uppgiftsbaserad ansats. För de områden där resultat från internationella studier utnyttjas för nationella kunskapsutvärderingssyften kan möjligheten att tillhandahålla uppgifter för lokal utvärdering vara begränsad av upphovsrättsliga skäl.

⁷⁷ En intressant modell tillämpas i Finland, där skolor som inte ingår i Utbildningsstyrelsens nationella urval av skolor i de nationella utvärderingar som görs, har möjlighet att köpa tjänsten (proven; analys av resultaten; samt resultatrapport till den beställande skolan i vilken denna kan jämföra sina resultat med nationella data) av Utbildningsstyrelsen (närmast den finska motsvarigheten till Skolverket).

⁷⁸ Detta avser framför allt de fall då exempelvis upphovsrättskyddade texter eller bilder används. Detta är f.ö. inte inikt för de internationella studierna, utan liknande restriktioner finns bl.a. kring vissa instrument som användes i US 98 och NU 03.

Referenser

Referenser

- EU (2005). Kom (2005) 356 meddelande från kommissionen till europaparlamentet och rådet: Europeisk indikator för språkkunskaper
- Gustafsson, J.-E. (2006). Ramverk för ett system för uppföljning av kunskapsutvecklingen i grundskolan. Underlagsrapport till Skolverket.
- Härnqvist, K. (1975). The International Study of Educational Achievement. *Review of Research in Education*, 3, s. 84–09.
- Lindström, J.-O. (2006) *En jämförelse av NU-03, TIMSS 2003 och PISA 2003 i förhållande till varandra och de svenska kursplanerna i Matematik och NO*. Opublicerat manus på uppdrag av Skolverket (planeras, efter sedvanlig beredning i Skolverket, redovisas senare under hösten 2006).
- Selghed, B. (2004) *Ännu icke godkänt – Lärares sätt att erfarar betygssystemet och dess tillämpning i yrkesutövningen*. Malmö Studies in Educational Sciences no. 15. Malmö: Malmö högskola, Lärarutbildningen.
- Skolverket (1995). *Svensk läsundervisning i ett internationellt perspektiv*. Skolverkets rapport nr 80. Stockholm: Liber.
- Skolverket (1999a). *Läroplanerna i praktiken*. Skolverkets rapport nr 175. Stockholm: Liber.
- Skolverket (1999b). *Tema tillståndet i världen*. Underlagsrapport, beställningsnummer 99:485. Stockholm: Liber.
- Skolverket (1999c). *Kommunikativ förmåga i främmande språk*. Underlagsrapport, beställningsnummer 99:489. Stockholm: Liber.
- Skolverket (2000). *Nationella kvalitetsgranskningar 2000*. Rapport nr. 190. Stockholm: Liber.
- Skolverket (2003). *Det nationella provsystemet – vad, varför och varthän?* Bilaga till Skolverkets svar på regeringsuppdrag avseende resultatinformation (U2003/2060/S), deluppdrag E: Skolverkets bedömning av dagens system med nationella prov med avseende på kvalitet och kostnadseffektivitet. Skolverkets dnr 01-2003:2038.

Skolverket (2004a) Det nationella provsystemet i den målstyrda skolan – Omfattning, användning och dilemman. (beställningsnummer 04:870). Stockholm: Fritzes.

Skolverket (2004b) Policy och riktlinjer för deltagande i internationella jämförande studier, Intern PM 2004-03-03. Dnr 03-2004:631.

Skolverket (2004c). Internationella studier under 40 år – Svenska resultat och erfarenheter, Skolverkets aktuella analyser 2004. Stockholm: Fritzes.

Skolverket (2004d). *Nationella utvärderingen av grundskolan 2003. Sammanfattande huvudrapport*. Skolverkets rapport 250. Stockholm: Fritzes.

Skolverket (2005) *Övergripande frågeguide – förskoleklass, grundskola och obligatorisk särskola – kommunala och fristående huvudmän*. Intern PM 2005-09-12.

Skolverket (2006a). *Bedömningsunderlag – förskoleklass och grundskola, kommunal och fristående huvudman*. Intern PM 2006-08-12.

Skolverket (2006b) *Med fokus på läsförståelse – En analys av skillnader och likheter mellan internationella studier och nationella kursplaner*. Skolverkets aktuella analyser 2006. Stockholm: Fritzes.

Tholin, J. *Att kunna klara sig i okänd natur. En studie av betyg och betygsriterier – historiska betingelser och implementering av ett nytt system*. Borås: Högskolan i Borås.

Utbildningsdepartementet (2003). *Bedömning av dagens provsystem med avseende på kvalitet och kostnadseffektivitet*. Regeringsuppdrag, Dnr U2003/2060/S.

Utbildningsdepartementet (2004). Regeringsuppdrag avseende resultatinformation 2004-12-22.

Utbildningsdepartementet (2006). *Översyn av grundskolans mål- och uppföljningssystem m.m.* Direktiv 2006:19.

Bilaga

Bilaga

Av Skolverket uppskattade initiala samt efterföljande kostnader för ett ämne i ett rullande utvärderingssystem samt totala systemkostnader för grundskolan (i tusen kronor), enligt det förslag som redovisats av professor Jan-Eric Gustafsson i underlagsrapport till Skolverket 2006-05-16.

Poster	Cykel 1		Cykel 2 och efterföljande cykler...	
	Åk 8/9	Åk 4/5	Åk 8/9	Åk 4/5
Ramverksutveckling	1 000			
Instrumentutveckling	1 800	1 200	900	600
Tryck och distribution	200	200	200	200
Skanning	300	150	300	150
Bedömning	300	150	300	150
Databasuppbyggnad	150	150		
Databasdrift	100	100	100	100
Rapportering	400	200	400	200
Totalt per ämne	4 250	2 150	2 200	1 400
Totala systemkostnader per cykel (18 åk8/9+11 åk 4/5)		10 0149		54 999

Skolverket

www.skolverket.se